Cross-Domain Health Misinformation Detection on Indonesian Social Media

Divi Galih Prasetyo Putri[1], Savitri Citra Budi[2], Arida Ferti Syafiandini[3],
Ikhlasul Amal[4], Revandra Aryo Dwi Krisnandaru[5]
Department of Electronic and Informatics Engineering, Universitas Gadjah Mada, Indonesia[1,5]
Department of Health Information Management, Universitas Gadjah Mada, Indonesia[2]
Research Center for Computing, National Research and Innovation Agency (BRIN), Indonesia[3]
Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia[4]

***Abstract*—Indonesia is among the world's most prolific countries in terms of internet and social media usage. Social media serves as a primary platform for disseminating and accessing all types of information, including health-related data. However, much of the content generated on these platforms is unverified and often falls into the category of misinformation, which poses risks to public health. It is essential to ensure the credibility of the information available to social media users, thereby helping them make informed decisions and reducing the risks associated with health misinformation. Previous research on health misinformation detection has predominantly focused on English-language data or has been limited to specific health crises, such as COVID-19. Consequently, there is a need for a more comprehensive approach which not only focus on single issue or domain. This study proposes the development of a new corpus that encompasses various health topics from Indonesian social media. Each piece of content within this corpus will be manually annotated by expert to label a social media post as either misinformation or fact. Additionally, this research involves experimenting with machine learning models, including traditional and deep learning models. Our finding shows that the new cross-domain dataset is able to achieve better performance compared to those trained on the COVID dataset, highlighting the importance of diverse and representative training data for building robust health misinformation detection system.**

***Keywords*—Health misinformation; machine learning; social media**

I. Introduction

The proliferation of internet users in Indonesia has been steadily increasing. A 2023 survey by the Indonesian Internet Service Provider Association reported that internet penetration had reached 78%, marking a significant rise compared to the previous year[1]. The COVID-19 pandemic in early 2020 played a pivotal role in accelerating the adoption of internet-based applications for daily activities. With the shift to remote work, online education, e-commerce, and telemedicine, the internet became indispensable for modern life. Among these changes, access to health information via digital platforms has expanded significantly. Social media platforms, in particular, have enabled the rapid dissemination of health-related content through user-generated content (UGC), encompassing text, images, videos, and comments on a wide range of topics. Features such as likes, shares, and comments amplify the spread of information, which spans various subtopics, including general health, vaccines, diseases like Ebola and cancer, and public health crises such as the COVID-19 pandemic [1], [2].

However, the exponential growth of information on social media presents a dual challenge: while access to health-related content is democratized, it becomes increasingly difficult for users to assess the credibility and quality of this information. Studies have identified social media platforms like Facebook, Twitter, and Instagram as primary vectors for the spread of health misinformation, which often propagates faster than accurate information [3]. Health misinformation has been linked to severe consequences, particularly during the COVID-19 pandemic, where widespread falsehoods created confusion, reduced vaccine uptake, and undermined herd immunity efforts. The World Health Organization (WHO) reported that misinformation during the pandemic led to over 6,000 hospitalizations and 800 deaths globally [4]. Such incidents underscore the real-world consequences of misinformation, which can escalate from individual confusion to public health crises.

Misinformation surrounding vaccines exemplifies the long-standing impact of health-related falsehoods. For instance, between September 2018 and July 2019, 85% of the 649 reported measles cases in the U.S. involved unvaccinated individuals [5]. Furthermore, recent surveys indicate that misinformation about chronic diseases, such as diabetes and cancer, is among the most concerning categories of health-related content on social media. The persistent spread of misinformation highlights the critical need to address this issue comprehensively. Left unchecked, such falsehoods can lead to misinformed decisions, delays in seeking proper medical care, and ultimately, adverse health outcomes for individuals and communities.

The urgency of tackling health misinformation has prompted researchers to explore automated solutions for misinformation detection. Automated systems can enable users to instantly assess the credibility of content accessed on social media, empowering them to make informed health decisions and reducing the risks associated with misinformation. While significant progress has been made in health misinformation detection research, much of the existing work has concentrated on English-language data [6]. This focus leaves a critical gap in addressing misinformation in non-English contexts, such as Bahasa Indonesia, a language spoken by over 270 million people.

In Indonesia, 76% of users perceive social media as a trustworthy source of information [7]. Moreover, according to

[1]https://m.bisnis.com/amp/read/20230308/101/1635219/survei-apjii-pengguna-internet-di-indonesia-tembus-215-juta-orang.

data from the Indonesian Telecommunications Society, approximately 40% of the hoax news articles circulating in Indonesia in 2019 were related to health [8], making the detection of misinformation even more pressing. Recent efforts have begun creating datasets and machine learning models for detecting misinformation in Bahasa Indonesia. However, these efforts have primarily focused on domain-specific topics, such as COVID-19, limiting their applicability to other health-related misinformation. Moreover, existing studies often neglect the diverse and evolving nature of misinformation across different health domains, which can include varied topics like traditional medicine, vaccines, mental health, and non-communicable diseases.

To address the gaps identified in previous studies, this research seeks to answer the following key research questions:

1) How can a cross-domain dataset for health misinformation detection in Bahasa Indonesia be effectively constructed to address diverse health topics?
2) Can machine learning models trained on the proposed cross-domain dataset generalize effectively across various health domains, reducing dependence on any single domain?
3) How does the performance of the proposed cross-domain dataset compare to existing domain-specific datasets in terms of robustness and quality?

Based on the proposed research questions, this study makes the following contributions to the field of health misinformation detection:

1) Dataset development: Creation of a cross-domain dataset for health misinformation detection in Bahasa Indonesia, encompassing diverse health topics to enable broader generalization.

2) Model evaluation: Preliminary experiments using multiple machine learning approaches to evaluate the effectiveness of the constructed dataset.

3) Benchmarking: Comprehensive comparison of the proposed dataset against existing domain-specific datasets to demonstrate its robustness and quality.

This paper is structured as follows: Section II reviews related work on health misinformation detection, highlighting gaps in existing research. Section III details the methodology for constructing the dataset and designing preliminary experiments. Section IV presents and discusses the experimental results, while Section V concludes the study and outlines directions for future work.

II. Related Works

Health misinformation detection has been extensively studied, primarily focusing on English. These studies leverage various approaches, such as linguistic and behavioral features, to identify and combat misinformation. For instance, Zhao et al. (2021) proposed a model combining central and peripheral features based on the Elaboration Likelihood Model (ELM), significantly improving detection accuracy by integrating user interaction patterns [6]. Similarly, Zhong et al. (2023) analyzed temporal and sentiment patterns in misinformation dissemination on Twitter [9], revealing that misinformation tends to persist longer and garner more engagement than credible information.

Despite these advancements, research on health misinformation detection in non-English contexts, particularly Bahasa Indonesia, remains limited. Faisal and Mahendra (2022) addressed this gap by developing a COVID-19-specific misinformation dataset and proposing a two-stage classifier leveraging IndoBERT for Indonesian tweets. Their approach demonstrated the efficacy of pre-trained language models but was constrained by its focus on the COVID-19 domain [7]. In fact, several studies on health misinformation focused on the topic of covid-19 [10], [11]. Another effort by Prasetyo et al. (2018) explored classification techniques for health-related hoax news in Bahasa Indonesia using the Modified K-Nearest Neighbor (MKNN) method. Their results showed an accuracy of 75%, with performance influenced by challenges such as unstructured text and diverse linguistic styles in Indonesian health news [12].

In the broader context of misinformation detection in Indonesia, Rohman et al. (2021) conducted a systematic literature review analyzing methods for fake news classification in Bahasa Indonesia. Their findings identified 19 commonly used algorithms, with Naïve Bayes [13] and Term Frequency-Inverse Document Frequency (TF-IDF) [14] being the most exploited approach. They highlighted the dominance of datasets sourced from platforms like turnbackhoax.id and Twitter and emphasized the need for exploring a wider range of methods beyond these mainstream approaches [15]. Additionally, a recent study by Arini et al. (2024) examined the sociocultural factors, such as varying levels of trust in information sources, that influence the spread and detection of misinformation during the pandemic [16].

The concept of cross-domain classification has emerged as a promising avenue for addressing the limitations of domain-specific models. Kansal et al. (2020) reviewed cross-domain sentiment analysis methods, emphasizing their potential to reduce reliance on single-domain datasets through domain adaptation and transfer learning [17]. These findings suggest that cross-domain approaches could enhance the generalizability of misinformation detection models across diverse health topics. Based on these insights, this research aims to address the gaps in existing studies by proposing a cross-domain dataset for health misinformation detection in Bahasa Indonesia. By leveraging state-of-the-art machine learning techniques and integrating cross-domain classification methods, this study seeks to overcome the challenges posed by linguistic and contextual diversity in Indonesian health misinformation.

III. Research Methodology

In this research, we develop a dataset for cross-domain health misinformation detection in Indonesian tweet. Moreover, we perform a preliminary study by exploring commonly used machine learning approach.

A. Data Collection and Annotation

The primary goal of this phase is to construct a cross-domain corpus capturing health misinformation phenomena in Indonesian social media. The dataset construction process is illustrated in Fig. 1.

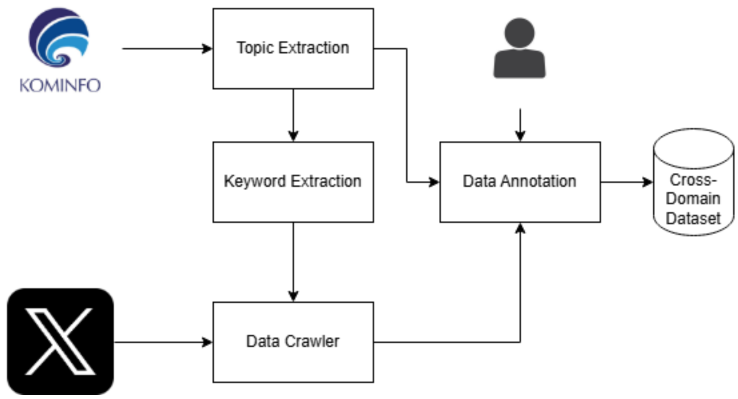

Fig. 1. Corpus building process.

1) Data collection: Data was collected from various health-related topics by utilizing weekly hoax news content published on the official website of the Ministry of Communication and Information Technology (KOMINFO) of Indonesia (Kominfo, 2023). A total of 27 hoax news articles covering topics such as COVID-19, HIV, polio, and other health concerns were analyzed. Table I summarizes the articles and the number of tweets crawled for each.

2) Keyword extraction: For each article, primary keywords were extracted and combined into search queries for crawling tweets. For example, for the article titled "[HOAKS] Buah Pala dan Gula Batu Mampu Mengatasi Jantung Berdebar" [18], keywords such as Pala, Gula Batu, and Jantung Berdebar were extracted and combined into search phrases like Pala AND Jantung Berdebar and Gula Batu AND Jantung Berdebar. These phrases were used in the crawling process, focusing on data from Twitter (now known as X).

3) Manual annotation: After collection, the tweets were manually annotated by two annotators based on predefined criteria. The annotation process involved verifying that the content was written in Indonesian, relevant to health information, and determining whether the content constituted misinformation or truth. The labels assigned were:

a) Misinformation: Content identified as false based on the KOMINFO news article.

b) True: Content verified as accurate.

To ensure the reliability of the annotations, inter-annotator agreement was measured using Cohen's Kappa. In cases of disagreement, a third annotator resolved conflicts. This process ensures high-quality annotations for downstream tasks.

B. Model Development

To assess the quality and applicability of the newly constructed dataset, a preliminary study was conducted on the task of health misinformation detection. This involved training and evaluating various machine learning models.

1) Traditional models: The following models were implemented using scikit-learn with Bag-of-Words as the feature representation:

- Naive Bayes
- Support Vector Machines (SVM)
- Logistic Regression

TABLE I. SUMMARY OF HOAX NEWS ARTICLES AND CRAWLING RESULTS

News Title	Topic	Numb. of Tweet
[HOAKS] Covid-19 bukan Virus, Sumber: Kementerian Kesehatan RI	covid	29
[HOAKS] Covid-19 adalah Senjata Biologis dari Cina	covid	716
[HOAKS] Vaksin Covid-19 Adalah Antena 5G dan Penyebab Kanker	covid	220
[HOAKS] Vaksin Covid-19 Sinovac Sebabkan Mpox	covid	29
[HOAKS] Obat Corona Bernama Pil-Kada	covid	64
[HOAKS] Tinta Tak Kasat Mata Dimasukkan ke Vaksin	covid	12
Vaksin Berbasis mRna Picu Kanker? Itu Disinformasi	covid	66
[HOAKS] Mengkudu Menyembuhkan Darah Tinggi Secara Total	high blood preasure	72
[HOAKS] Akar Kelapa dan Kuning Telur Dapat Meningkatkan Fertilitas	Fertility	39
[HOAKS] Video Tata Cara Pertolongan Pertama Penanganan Flu Burung	avian flu	23
[HOAKS] Terapi Minuman Rempah Bisa Menggantikan Cuci Darah pada Gagal Ginjal	kidney failure	7
[HOAKS] GERD Picu Kematian Mendadak	GERD	110
[HOAKS] Penularan HIV dan AIDS di Kolam Renang	HIV	294
[HOAKS] HIV Dapat Ditularkan oleh Gigitan Nyamuk	HIV	564
[HOAKS] Buah Pala dan Gula Batu Mampu Mengatasi Jantung Berdebar	Heart	10
[HOAKS] Parfum, Pengharum Ruangan, dan Wangi Dry Clean Sebagai Penyebab Kanker	Cancer	347
[HOAKS] Pisang Dempet Sebabkan Anak Lahir dengan Kembar Siam	cojoined twins	312
[HOAKS] Cara Mengecek Kadar Kolesterol melalui Warna Kuku	Cholestrol	34
[HOAKS] Tanaman Pacing Dapat Sembuhkan Mata Minus	near-sightedness	3
[HOAKS] Peringatan IDI Terkait Adanya Wabah Pengerasan Otak, Sumsum Tulang, dan Diabetes	brain, bone marrow, diabetes	381
[HOAKS] Dokumen Rahasia BPOM Sebut Vaksin Polio Tidak Aman	Polio	46
[HOAKS] Indonesia Sudah Lama Tidak Ada Wabah Polio	Polio	53
[HOAKS] KLB Polio Disebabkan Vaksin Polio Tipe-2	Polio	29
[HOAKS] Getah Bunga Mahkota Duri Sembuhkan Sakit Gigi Secara Instan	toothache	3
[HOAKS] Mengobati Sesak Nafas dengan Pijat Tangan	dyspnea	90
[HOAKS] Video Tata Cara Pertolongan Pertama Penanganan Stroke	stroke	70
Total		**3623**

- Decision Tree
- Random Forest

2) Deep learning models: Advanced deep learning models were also explored, including:

- Bi-LSTM (Bidirectional Long Short-Term Memory)
- BERT
- IndoBERT [19], a transformer-based language model specifically designed for Indonesian.

We use Optuna for automated hyperparameter optimization [20] for traditional models and Bi-LSTM. For the transformer-based model, we use the default configuration of the models in HuggingFace with learning rate of 1e-5.

3) Evaluation: Models were evaluated on two datasets:

- The newly constructed cross-domain corpus.
- A COVID-specific dataset from previous research [7].

The evaluation metrics included:

- Accuracy: Proportion of correctly classified instances.
- Macro Precision: Average precision across all classes.
- Macro Recall: Average recall across all classes.
- Macro F1-Score: Harmonic mean of precision and recall.

These metrics provide a comprehensive assessment of model performance, particularly in imbalanced datasets.

The results will be compared with classifications using a corpus from previous research (COVID Data) [7] as the training data. This comparison aims to assess the stability and reliability of the newly constructed corpus. The evaluation will be conducted using standard performance metrics, including accuracy, macro-precision, macro-recall, and macro F1-score.

IV. Result and Discussion

In this section, we will discussed the dataset developed from the data collection and annotation process as explained in Section III-A and the preliminary experimental result as explained in Section III-B

A. Dataset Results

TABLE II. Dataset Label Distribution

Label	No.of Tweet	Percentage
Misinformation	1590	44
True Information	762	21
Not Sure	82	2
deleted	1189	33
Total	3623	100

The tweet crawled in the data collection process yielded a total of 3,623 tweets. The distribution of tweets for each topic is presented in Table III. From the table it is evident that the topic with the highest number of tweet is COVID followed by HIV, Cancer and Cojoined Twins. The prominance of COVID data is likely because there are much more news article related to COVID. Moreover, COVID is a phenomenon with global impact and widespread discussion. The topic of cojoined twins also tweeted by many people because the content is related to myths that are common in Indonesian society. In contrast, other topics has limited number of tweets likely because the content is not widely known and discussed.

From the collected data, 1189 tweets were deleted. These tweets were excluded because they are not written in Indonesian, unrelated to health topic or duplicate. The remaining tweets were annotated into two labels: Misinformation and *True*, as shown in Table II. Among these, 1,590 tweets (44%) were labeled as *misinformation*, representing the largest portion of the dataset. The *True Information* label includes 762 tweets (21%) that were verified as accurate and reliable. A small subset of the data, 82 tweets (2%), was labeled as *Not Sure*, indicating cases where annotators found it difficult to determine whether the tweet is misinformation or true information. We reach an inter-annotator agreement with a Cohen's Kappa value of 0.91, which indicates almost perfect agreement.

B. Preliminary Results

In the preliminary experiment, we utilized the dataset that labeled as *misinformation* or *True*, in total of 2,352 data. In the experiment, 80% of the data was allocated for training, while the remaining 20% was used for data testing. The distribution of the data training and data testing is shown in Table III.

TABLE III. Distribution of Training and Testing Data on Each Topic

Topic	Testing	Training	Total
Covid	336	655	991
High Blood Pressure	0	24	24
Fertility	0	18	18
Avian Flu	0	17	17
Kidney Failure	0	7	7
GERD	0	67	67
HIV	134	511	645
Heart Problem	0	7	7
Cancer	0	167	167
Cojoined Twins	0	165	165
Cholestrol	0	17	17
Near-sightedness	0	3	3
Brain, bone marrow, diabetes	0	30	30
Polio	0	91	91
Toothache	0	3	3
Dyspnea	0	46	46
Stroke	0	54	54
Total	**470**	**1882**	**2352**

The outcomes of our experiments are presented in Table IV. We employed traditional machine learning approaches, including Naive Bayes, SVM, Logistic Regression, Decision Tree, and Random Forest as well as deep learning approaches, including Bi-LSTM, BERT, and Indo-BERT. We conducted two experiments for each model, utilizing the COVID-specific dataset and the cross-domain dataset as the training data. The performance of the trained models was evaluated using test data derived from the cross-domain dataset.

The results clearly demonstrate that models trained on the cross-domain dataset consistently outperformed those trained on the COVID-specific dataset. This trend was observed across

TABLE IV. RESULTS OF MISINFORMATION DETECTION EXPERIMENT USING COVID DATASET AND CROSS-DOMAIN DATASET

	Covid Dataset				Cross-Domain Dataset			
Model	macro-P	macro-R	macro-F	Acc	macro-P	macro-R	macro-F	Acc
Naive Bayes	0.601	0.581	0.583	0.668	0.734	0.735	0.735	0.768
SVM	0.623	0.580	0.583	0.668	0.844	0.800	0.816	0.849
Logistic Regression	0.591	0.557	0.550	0.670	0.833	0.807	0.818	0.847
Decision Tree	0.546	0.542	0.542	0.617	0.818	0.806	0.811	0.838
Random Forest	0.589	0.592	0.590	0.636	**0.885**	0.762	0.791	0.842
Bi-LSTM	0.599	0.611	0.598	0.623	0.828	0.773	0.792	0.832
BERT	0.595	0.597	0.596	0.643	0.766	0.776	0.770	0.796
IndoBERT	0.597	0.599	0.598	0.647	0.844	**0.856**	**0.849**	**0.866**

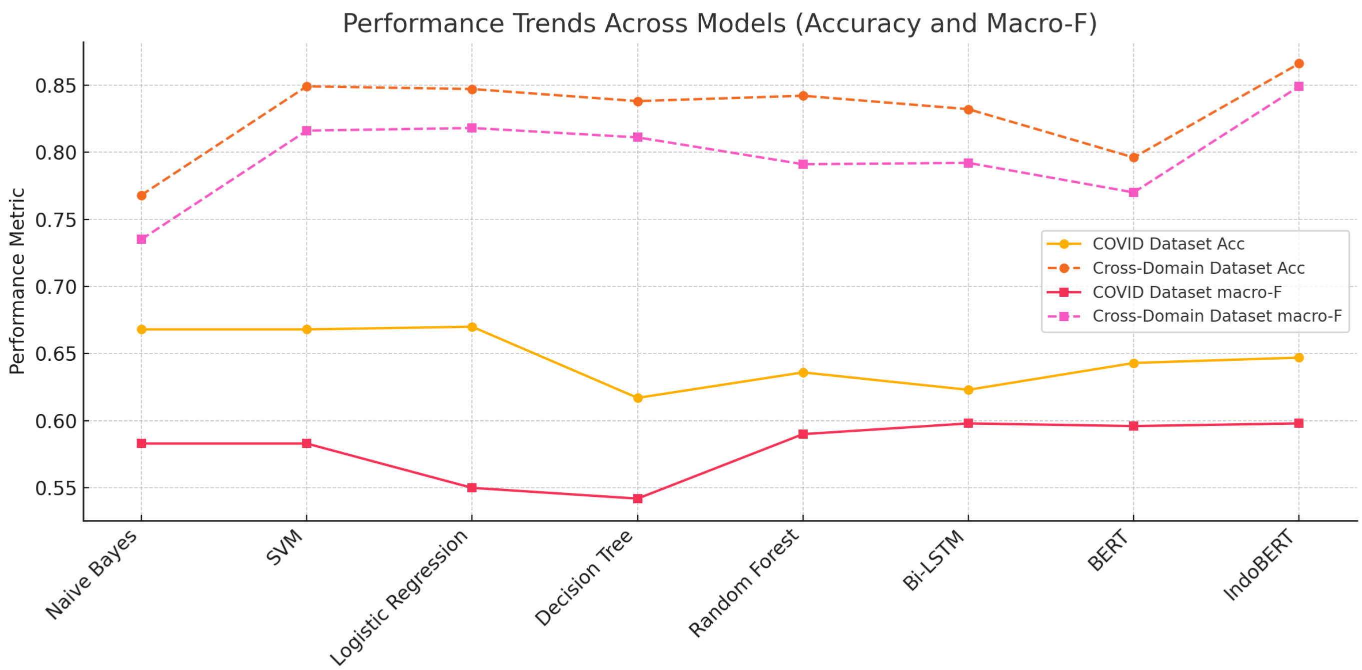

Fig. 2. Performance trends across models.

all models, both traditional and deep learning techniques. Notably, the use of cross-domain data appears to enhance the ability of the models to generalize across diverse contexts and topics, leading to improved predictive accuracy and robustness. The observed improvements in performance suggest that incorporating data from a variety of domains can mitigate the limitations of topic-specific datasets, which may lack diversity.

The results, as visualized in the Fig. 2, demonstrate significant differences in model performance when trained on the COVID dataset and the Cross-Domain dataset. Models trained on the Cross-Domain dataset consistently outperform those trained on the COVID dataset in both accuracy and macro-F scores. This highlights the superior generalization capability of the Cross-Domain dataset, which encompasses a broader range of health misinformation topics. Among the evaluated models, IndoBERT achieved the highest performance, with an accuracy of 0.866 and a macro-F score of 0.849 when trained on the Cross-Domain dataset. This underscores the effectiveness of transformer-based language models in handling complex linguistic features and diverse misinformation scenarios.

Simpler models such as Naive Bayes and Decision Tree also benefit from the Cross-Domain dataset, showing improved metrics compared to training on the COVID dataset. However, their performance is limited relative to more advanced models, reflecting their inability to capture nuanced patterns in the data. Models like SVM and Logistic Regression exhibit significant improvements with the Cross-Domain dataset, suggesting their adaptability to diverse training data while remaining computationally efficient. Meanwhile, deep learning models such as Bi-LSTM show stable improvements, but they are outperformed by transformer-based models like BERT and IndoBERT, indicating the latter's superior contextual understanding.

The comparison between datasets underscores the importance of dataset diversity in training robust misinformation detection models. The Cross-Domain dataset enables models to generalize across various health misinformation topics, in contrast to the COVID dataset, which limits models to a single domain. These results demonstrate the Cross-Domain dataset's ability to enhance training effectiveness, making it a valuable resource for developing generalizable health misinformation detection systems.

The heatmap shown in Fig. 3 offers a detailed view of the interplay between models, datasets, and evaluation metrics, providing unique insights beyond the trends observed

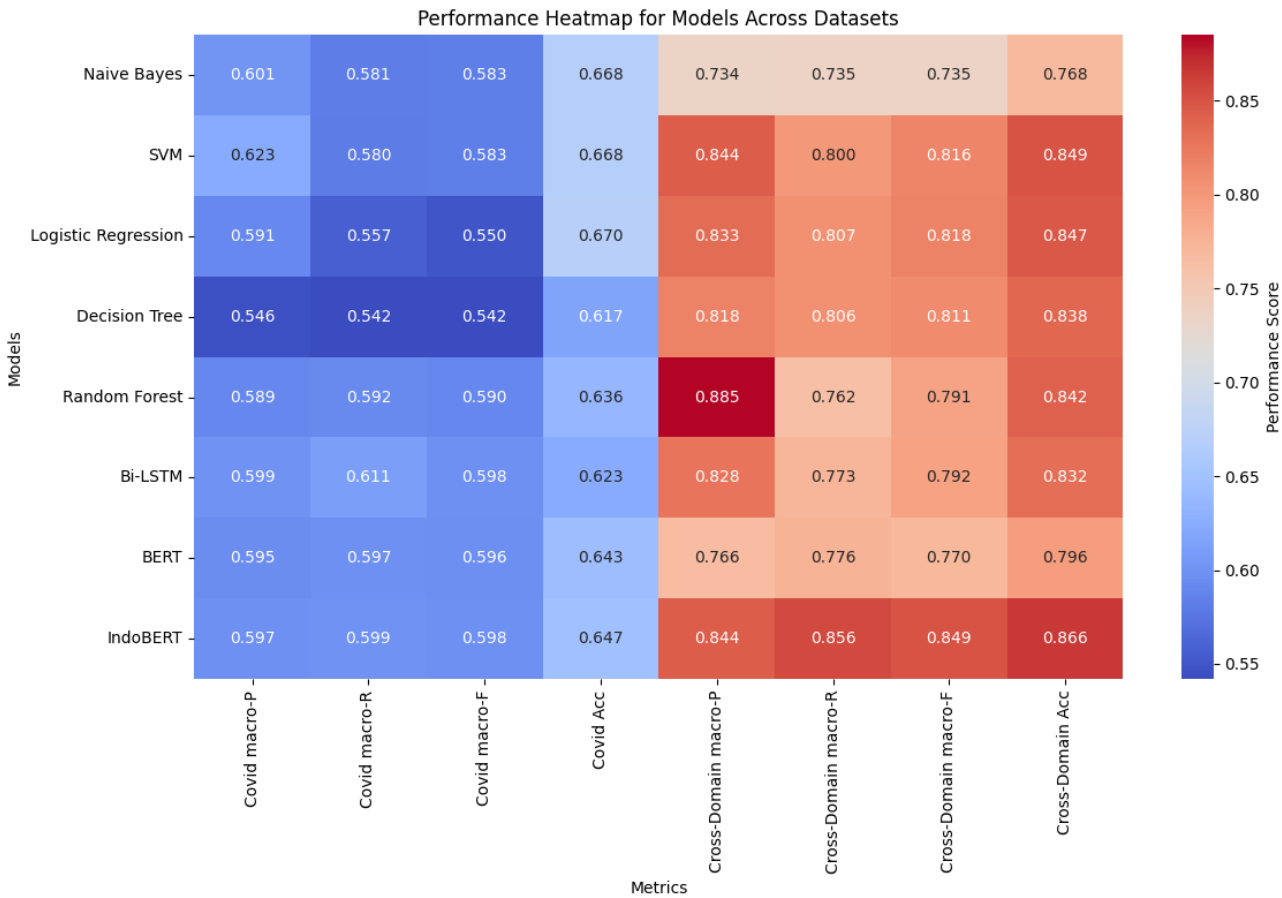

Fig. 3. Performance heatmap for models across datasets.

in the line charts. One key observation is the variability across metrics, where certain models demonstrate significant strengths in specific areas. For instance, IndoBERT exhibits a substantial improvement in macro-recall when trained on the Cross-Domain dataset (0.856) compared to the COVID dataset (0.599), indicating its effectiveness in reducing false negatives across diverse health misinformation topics. Additionally, the heatmap reveals trade-offs between precision and recall. For example, Random Forest achieves an exceptionally high macro-precision (0.885) on the Cross-Domain dataset but shows a relatively lower macro-recall (0.762), suggesting its strong capability in identifying true positives but limited sensitivity to all relevant instances.

The sensitivity of models to dataset diversity is also apparent, with traditional models such as Naive Bayes and Decision Tree showing larger relative gains from the COVID dataset to the Cross-Domain dataset, particularly in precision and accuracy. This highlights the disproportionate benefit simpler models derive from richer training data. Moreover, metric-specific strengths are evident; Random Forest excels in macro-precision, while IndoBERT consistently outperforms other models in macro-recall and macro-F, reflecting its ability to balance precision and recall effectively. Some models, such as BERT, display only modest improvements across datasets, suggesting the need for additional fine-tuning or data augmentation to fully exploit the diversity of the Cross-Domain dataset.

Overall, the heatmap underscores the uniform performance boost across all models when using the Cross-Domain dataset, validating its robustness and utility for training generalizable health misinformation detection systems. The variability in metric-specific performance and model sensitivity offers complementary insights, enriching the understanding of model behaviors across diverse datasets.

V. Conclusion

In conclusion, this research introduces a novel cross-domain dataset for health misinformation detection in Indonesian tweets. The primary objective was to address the limitations of domain-specific datasets, such as those focused solely on COVID-19, and to evaluate the efficacy of machine learning models in generalizing across diverse health misinformation topics. Through comprehensive experiments using traditional and deep learning models, our study demonstrated that the Cross-Domain dataset significantly improves model performance across all evaluation metrics, including accuracy, macro-precision, macro-recall, and macro-F score. Models trained on the Cross-Domain dataset consistently outperformed those trained on the COVID dataset, underscoring the value of diverse and representative training data in developing robust misinformation detection systems.

Our findings highlight the effectiveness of advanced models such as IndoBERT, which achieved the highest performance metrics and demonstrated exceptional adaptability to the linguistic and contextual diversity present in the dataset. Traditional models, while showing notable improvements, remained limited in their ability to capture nuanced patterns, further emphasizing the importance of leveraging state-of-the-art methods for complex tasks like health misinformation detection.

This study contributes to the field by providing a high-quality, cross-domain dataset and presenting evidence of its potential to enhance machine learning models' generalization capabilities. These findings address the research questions posed in this study, particularly regarding the construction of a cross-domain dataset and its impact on misinformation detection models. Future work could focus on expanding the dataset to include other health misinformation sources, exploring multilingual approaches, and refining machine learning techniques to further improve performance. By addressing challenges in low-resource linguistic contexts, this research paves the way for more effective and scalable solutions to combat health misinformation.

Acknowledgment

This work has been funded by University of Gadjah Mada under Grant Number 6541/UN1.P1/PT.01.03/2024 with title "Sistem Deteksi Misinformasi Kesehatan Lintas Domain Berbasis Artificial Intelligence pada Media Sosial di Indonesia".

References

[1] V. Suarez-Lledo and J. Alvarez-Galvez, "Prevalence of health misinformation on social media: systematic review," *Journal of medical Internet research*, vol. 23, no. 1, p. e17187, 2021.

[2] S.-F. Tsao, H. Chen, T. Tisseverasinghe, Y. Yang, L. Li, and Z. A. Butt, "What social media told us in the time of COVID-19: a scoping review," *The Lancet Digital Health*, vol. 3, no. 3, pp. e175–e194, 2021.

[3] J. R. Bautista, Y. Zhang, and J. Gwizdka, "Healthcare professionals' acts of correcting health misinformation on social media," *International Journal of Medical Informatics*, vol. 148, p. 104375, 2021.

[4] World Health Organization. (2020) Fighting misinformation in the time of covid-19: one click at a time. Accessed: 2024-12-05. [Online]. Available: https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time

[5] J. R. Zucker, J. B. Rosen, M. Iwamoto, R. J. Arciuolo, M. Langdon-Embry, N. M. Vora, J. L. Rakeman, B. M. Isaac, A. Jean, M. Asfaw *et al.*, "Consequences of undervaccination—measles outbreak, new york city, 2018–2019," *New England Journal of Medicine*, vol. 382, no. 11, pp. 1009–1017, 2020.

[6] Y. Zhao, J. Da, and J. Yan, "Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches," *Information Processing & Management*, vol. 58, no. 1, p. 102390, 2021.

[7] D. R. Faisal and R. Mahendra, "Two-stage classifier for covid-19 misinformation detection using bert: a study on indonesian tweets," *arXiv preprint arXiv:2206.15359*, 2022.

[8] I. Puspitasari and A. Firdauzy, "Characterizing consumer behavior in leveraging social media for e-patient and health-related activities," *International journal of environmental research and public health*, vol. 16, no. 18, p. 3348, 2019.

[9] B. Zhong, "Going beyond fact-checking to fight health misinformation: A multi-level analysis of the twitter response to health news stories," *International Journal of Information Management*, vol. 70, p. 102626, 2023.

[10] L. H. Suadaa, I. Santoso, and A. T. B. Panjaitan, "Transfer learning of pre-trained transformers for covid-19 hoax detection in indonesian language," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 3, pp. 317–326, 2021.

[11] A. P. Rifai, Y. P. Mulyani, R. Febrianto, H. M. Arini, T. Wijayanto, N. Lathifah, X. Liu, J. Li, H. Yin, Y. Wu *et al.*, "Detection model for fake news on covid-19 in indonesia," *ASEAN Engineering Journal*, vol. 13, no. 4, pp. 119–126, 2023.

[12] A. R. Prasetyo, I. Indriati, and P. P. Adikara, "Klasifikasi hoax pada berita kesehatan berbahasa indonesia dengan menggunakan metode modified k-nearest neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 12, pp. 7466–7473, 2018.

[13] H. Mustofa and A. A. Mahfudh, "Klasifikasi berita hoax dengan menggunakan metode naive bayes," *Walisongo Journal of Information Technology*, vol. 1, no. 1, pp. 1–12, 2019.

[14] D. Maulina and R. Sagara, "Klasifikasi artikel hoax menggunakan support vector machine linear dengan pembobotan term frequency–inverse document frequency," *Jurnal Mantik Penusa*, vol. 2, no. 1, 2018.

[15] M. A. Rohman, D. Khairani, K. Hulliyah, P. Riswandi, I. Lakoni *et al.*, "Systematic literature review on methods used in classification and fake news detection in indonesian," in *2021 9th International Conference on Cyber and IT Service Management (CITSM)*. IEEE, 2021, pp. 1–4.

[16] H. M. Arini, T. Wijayanto, N. Lathifah, Y. P. Mulyani, A. P. Rifai, X. Liu, J. Li, and H. Yin, "Detecting fake news during covid-19 in indonesia: the role of trust level," *Journal of Communication in Healthcare*, vol. 17, no. 2, pp. 180–190, 2024.

[17] N. Kansal, L. Goel, and S. Gupta, "A literature review on cross domain sentiment analysis using machine learning," *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines*, pp. 1871–1886, 2022.

[18] K. K. dan Informatika Republik Indonesia. (2024) Hoaks: Buah Pala dan Gula Batu Mampu Mengatasi Jantung Berdebar. Accessed on December 4, 2024. [Online]. Available: https://www.komdigi.go.id/berita/klarifikasi-hoaks/detail/hoaks-buah-pala-dan-gula-batu-mampu-mengatasi-jantung-berdebar

[19] 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 *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 757–770.

[20] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.