

Incorporating News Tags into Neural News Recommendation in Indonesian Language

Maxalmina Satria Kahfi, Evi Yulianti, Alfian Farizki Wicaksono
Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

Abstract—News recommendation system holds the potential to aid users in discovering articles that align with their interests, which is critical to alleviate user information overload. To generate effective news recommendations, one key capability is to accurately capture the contextual meaning of text in the news articles, since this is pivotal in acquiring useful representations for both news content and users. In this work, we examine the effectiveness of neural news recommendation with attentive multi-view learning (NAML) method to conduct a news recommendation task in the Indonesian language. We further propose to incorporate news tags, which at some levels may capture the important contextual meanings contained in the news articles, to improve the effectiveness of the NAML method in the Indonesian news recommendation system. Our results show that the NAML method leads to significant improvement (if not comparable) in the effectiveness of neural-based Indonesian news recommendations. Further incorporating news tags is shown to significantly increase the performance of the NAML method by 5.86% in terms of NDCG@5 metric.

Keywords—News recommendation; recommendation systems; news tags; user modeling

I. INTRODUCTION

A recommendation system plays a crucial role in helping users discover items that match their needs and preferences [1]. In business, it is a vital part of marketing strategies, particularly for boosting online sales by offering customers a curated selection of items tailored to their preferences [1, 2]. This process of recommendation can be automated using a wide range of techniques, such as content-based methods and collaborative filtering [3], and recently involves computing user and item embeddings which serves as a basis for predicting how much a user is likely to prefer a particular item [4, 5]. Recommendation systems have found application in various domains, including movie recommendation [6, 7, 8], news recommendation [9, 10, 11, 12, 13], and music recommendation [14, 15, 16, 17].

News recommendation stands as one of the most frequently employed use cases of recommendation systems within the digital landscape, notably embraced by online platforms and news publishers [2, 18, 10]. Given the overwhelming amount of news articles generated daily, it is impractical for users to manually sift through all of them to find content that aligns with their interests [19]. Every day, an enormous volume of news articles is generated and published online, posing a challenge for users to efficiently discover news that aligns with their interests [2, 20]. Therefore, the implementation of personalized news recommendation becomes vital, as it enables online news platforms to target user preferences effectively and alleviate the issue of information overload [10].

In the past few years, news recommendation models have benefited from the use of deep learning methods [9, 11, 10, 13, 12]. These models are often referred to as neural recommendation models. Various neural recommendation models have been introduced to tackle news recommendation challenges, including NAML (Neural news recommendation with Attentive Multi-view Learning) method [11]. The NAML method uses an attentive multi-view learning model that learns unified news representations from different kinds of news information, such as news titles, bodies, and categories. This method takes advantage of multiple useful information from news articles that could enrich the semantics captured in the news representation. As a result, this method has been shown to outperform a range of deep learning methods for news recommendation, such as Convolutional Neural Network (CNN) [21], Deep fusion model (DFM) [22], Deep news recommendation based on knowledge-aware CNN (DKN) [23], etc. The NAML method also offers the flexibility to incorporate other useful information that could be exploited from news articles to produce more accurate news representation, which may result in more effective news recommendations. With this reported effectiveness as well as the potential to utilize extra knowledge for improving news representation, we propose to use the NAML method to perform the Indonesian neural news recommendation task in this work.

As mentioned above, the NAML architecture enables us to easily incorporate extra information to produce news representation. Therefore in this work, we also want to investigate the effectiveness of integrating news tags as supplementary information into the news recommendation model, aimed at improving the news representation. We argue that news tags at some levels may capture the important points in the articles, which therefore may provide extra useful information to generate better news representation. Previous research has primarily relied on a limited set of news components, such as title, category, and subcategory [24]. To the best of our knowledge, the use of news tags to enhance recommendation performance has not been explored in previous work.

Finally, this research endeavors to address two primary research questions:

- 1) RQ1. How is the performance of the NAML neural recommendation method to perform news recommendation task in the Indonesian language?
- 2) RQ2. To what extent news encoder in the NAML model can benefit from the use of news tags in the Indonesian news recommendation task?

By providing answers to these research questions, this study contributes to the existing body of research by providing

insights into the effectiveness of a neural method, i.e., NAML, for Indonesian news recommendation systems, as well as the impact of incorporating news tags into the NAML architecture.

The rest of this paper is organized as follows. Section II is dedicated to an in-depth review of relevant literatures related to news recommendation models, news information, and text embedding. Section III outlines the research methodology adopted in this study. Moving on to Section IV, this is where the results of the data analysis are presented and discussed. Section V serves as the conclusion, encapsulating key findings and providing directions for future research based on the study's insights. Finally, Section VI provides recommendations for future work based on the findings of this study.

II. RELATED WORKS

A. Text Embedding

Text embedding is a technique that represents text as a vector of real numbers. This allows computers to process text in a more meaningful way, as the vector representations can capture the semantic meaning of the text. There are several text embedding methods that try to capture the semantic and syntactic meaning of the text input.

The Word2Vec model served as a fundamental baseline for word embedding [25]. Their work introduced two innovative model architectures specifically crafted for generating continuous vector representations of words using vast datasets. These representations were evaluated for their quality in a word similarity task. The Word2Vec model has demonstrated its capability to address various tasks, including text summarization [26, 27, 28], ranking for academic expert finding [29, 30], and text classification [31, 32, 33]. Several models similar to Word2Vec were subsequently introduced, including Glove (Global Vectors for Word Representation) [34] and FastText [35]. These models build upon Word2Vec's foundations, enhancing the learning of word representations to acquire deeper semantic insights from the corpus.

In 2017 BERT is introduced, a bidirectional model that can learn context-aware and informative representations of words and phrases [36]. BERT, as a bidirectional model, learns word representations considering both forward and backward contexts, distinguishing it from previous models limited to forward context understanding. BERT is pre-trained on a massive dataset of text and code, and it can be fine-tuned for a variety of natural language processing tasks. This model remains challenging for underrepresented languages, which face unique obstacles due to limited data availability and the substantial corpus required for effective training [37].

Notably, there have been notable works on BERT-based models created specifically for the Indonesian language. These models, known as IndoLEM [38] and IndoNLU [39], have been developed independently, each trained on a different corpus of Indonesian text. IndoLEM and IndoNLU represent valuable contributions to Indonesian NLP research. They serve as pre-trained language models that have learned to understand and encode the linguistic patterns, semantics, and contextual information present in Indonesian text data. In addition, the BERT-based model has been demonstrated to enhance performance in classification tasks [31]. In our research, we evaluate

the capabilities of these two distinct BERT-based models to obtain text embeddings. Furthermore, our part of the research is to conduct a comprehensive comparison of various text embedding methods and evaluate their performance within the context of a news recommendation system.

B. News Information

In the scope of news recommendation systems, the use of news information as news representation plays a pivotal role in constructing the entire system [2, 19]. The recommendation model aims to construct a user profile customized to individual preferences through an analysis of the articles that a user reads. To achieve this, news representation involves the transformation of some or all news information into vector form [24].

Typically, the components input into the model encompass essential elements such as the article's title, abstract, body, category, and subcategory. These components collectively serve as the news representation for generating a comprehensive vector representation of news articles. This representation is instrumental in facilitating the modeling of user preferences and enabling the recommendation system to provide personalized news suggestions that align with each user's unique interests and preferences [4, 5].

Another valuable component of news information is the news tags. Although the news tags show the topic information that is closely related to the news content, the existing personalized news recommendation methods usually ignore the value of tags. News tags are usually used by users to track certain topics in the news portal which can trace other articles based on that tags. This behaviour leads to our research to check whether we add news tags as additional information is necessary for news recommendation models.

One of the methods to process the tag information is by leveraging social bookmarking [40]. The method involves the utilization of tags, which are assigned by a community of users with shared preferences. It's important to note that while tags are still employed in this method, they now signify a collective understanding within the user group, aligning with their common interests and preferences. An alternative approach involves creating a probability relation graph among tags, exploring potential correlations among different tags [41]. Building on this foundation, they employ the term frequency-inverse document frequency (TF-IDF) method to determine tag weights. Additionally, they introduce a novel approach to calculate the correlation degree between tags, utilizing conditional probability as a key metric.

C. News Recommendation Methods

In recent years, the field of news recommendation systems has received significant attention due to the exponential growth of online news consumption and the need for personalized content delivery. Several studies have been conducted to explore various approaches for improving the accuracy and relevance of news recommendations. News recommendation systems can be broadly categorized based on how they model user behavior. Two primary categories exist: Candidate-Agnostic (C-AG) models and Candidate-Aware (C-AW) models [42]. Both of the categories are illustrated in Fig. 1.

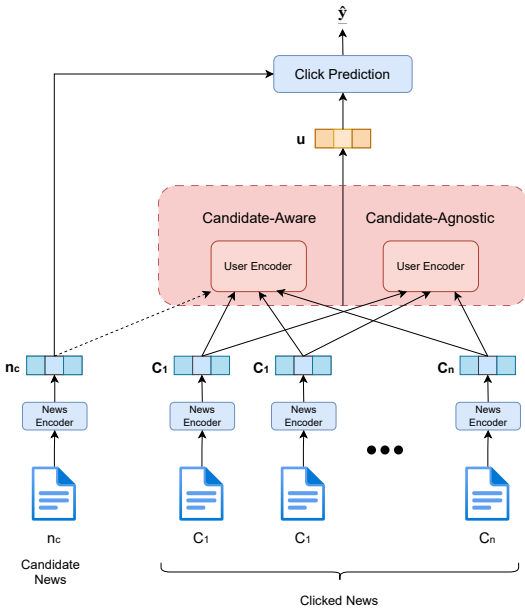


Fig. 1. News recommendation system general framework.

In C-AG Models, the User Encoder (UE) forms user embeddings exclusively from the embeddings of previously clicked news articles, without taking into account the candidate news articles [10, 9, 11, 12]. Essentially, the user embedding remains the same regardless of the specific candidate news being presented. One of the C-AG models is NAML that using a multi-view learning framework to learn unified news representations by incorporating several news information like title, body, and category. The NAML architecture is easily modified to add other news information. For the user encoder, NAML utilize a news attention network to identify crucial news articles, aiding in the acquisition of informative user representations [11]. The experimental results showcased significant enhancements in recommendation accuracy with the deep learning approach compared to traditional methods. Due to this performance, we have selected NAML as the primary model for our research focus.

Conversely, in C-AW models, the User Encoders (UEs) produce user embeddings that are influenced by the content of the candidate news articles. This means that in C-AW Models, the user embeddings can vary depending on the specific candidate news article in consideration [23, 13]. While C-AG Models maintain a consistent user embedding, C-AW models adapt the user representation to the characteristics of the candidate news, thereby potentially improving recommendation accuracy.

Considering the importance of contextual information in news recommendations, researchers have explored the incorporation of additional factors such as temporal relevance and user context. With the advent of deep learning techniques, several studies have investigated the application of neural networks for news recommendation [10, 9, 11, 13].

With the evolution of news recommendation systems, several metrics are employed to compare the performance

among models. The most common ones include AUC (Area Under Curve), MRR (Mean Reciprocal Rank), and NDCG (Normalized Discounted Cumulative Gain) [4, 43]. For models treating the news recommendation task as a classification problem, the Area Under Curve (AUC) score is a frequently used metric to measure the accuracy of the model. MRR is calculated as the reciprocal of the position of the first relevant element in the ranking. NDCG considers graded relevance (rating values) along with positional information of the recommended items. Both MRR and NDCG assess the relevancy of recommendations from the news recommendation system (NRS) model.

In summary, news recommendation systems have been extensively studied, employing various approaches such as contextual information and deep learning approaches. These studies have contributed to the advancement of news recommendation systems, enhancing their accuracy, relevance, and personalization capabilities.

III. RESEARCH METHODOLOGY

A. Dataset

We conducted experiments on a real-world news recommendation dataset collected from one of the largest news portals in Indonesia. We randomly sampled users who had at least 5 news click records during 4 weeks from Feb 10 to March 16, 2023. The last 1 week was treated as a data test. We collected the behaviour logs of these users in this period, which are formatted into impression logs. An impression log is a record of the news articles that are presented to a user when they visit a news page, including information about the time of the visit and the user’s interactions such as clicks on these news articles. In Fig. 2, you can see the layout of the news article page and an example of news information presented within an article.

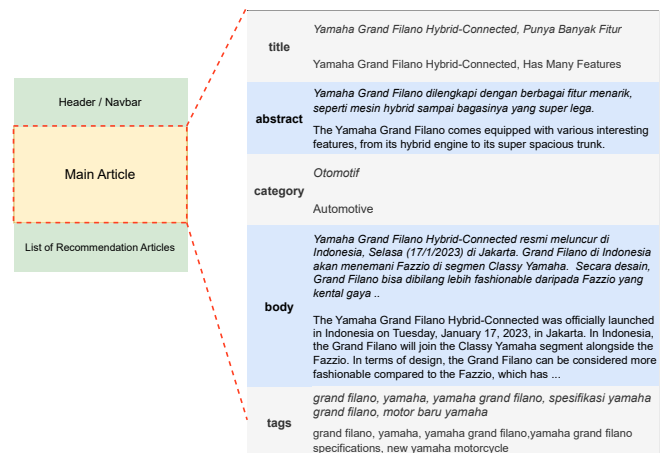


Fig. 2. Illustration of page layout and main article example.

Regarding the news dataset, we focus on several news components to figure out their influence on the performance of news recommendation systems. These components encompass the title, abstract (short description), news body, category, and subcategory. Furthermore, our primary research objective is to

scrutinize the impact of news tags on the model. By examining how each of these elements shapes the model’s performance, we aim to gain valuable insights into the factors that drive effective news recommendations. The statistical information about the dataset summarized in Table I.

TABLE I. STATISTICAL INFORMATION ABOUT THE DATASET

Component	Value
# User	34,053
# User Logs	181,875
# News	79,041
# Category	27
# Subcategory	148
Avg. Title Len.	10.40
Avg. Abstract Len.	16.80
Avg. Body Len.	371.22
Avg. Tags per Article	4.55

Fig. 3a, 3b, and 3c provide insights into the length distributions of news titles, abstracts, and bodies. Notably, news titles exhibit a distinct pattern of being quite short, averaging only 10.4 words. In contrast, both news abstracts and bodies present significantly longer text lengths, offering the potential for more in-depth coverage of news content. This discrepancy in text length underscores the value of incorporating diverse types of news information, including titles, abstracts, and bodies. Such integration enriches our comprehension of news articles by providing a more comprehensive and nuanced perspective on the content.

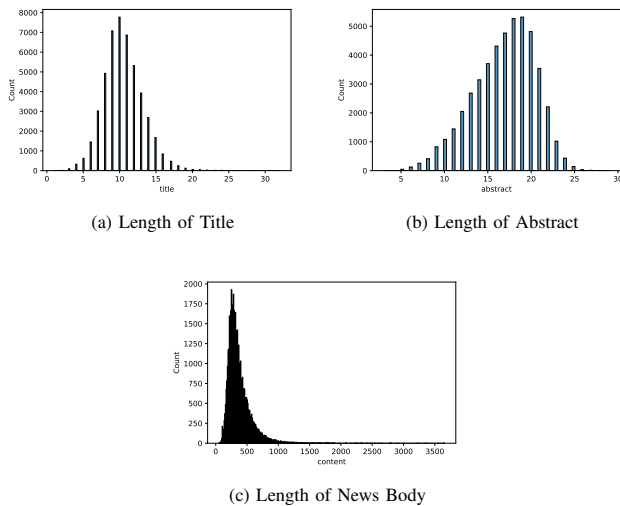


Fig. 3. Statistics of News Dataset.

In contrast to the approach outlined in [41], our methodology involves concatenating the news tags into a sentence or text format. We then treat this combined text in the same manner as other news information, such as the title and abstract. As a result, we transform the tags into a textual format, enabling them to be seamlessly integrated into the news content.

Our initial observations, as illustrated in Fig. 4, reveal an interesting connection between the amount of text data and the presence of words within news tags. With the expansion of text data, there is a higher probability of encountering

shared words between the tags and other textual content. This intriguing pattern suggests that even though news tags are found within other news information, they can contribute to a more comprehensive understanding of the article. This is because the proportion of news tags included in other news information remains relatively low. This highlights the potential significance of news tags in enriching news representation.

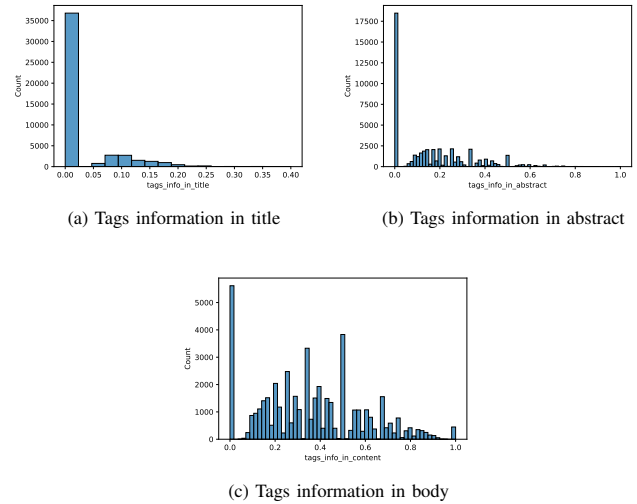


Fig. 4. Percentage of tags information included in other news information.

B. Experiment Settings

In our study, we employed two pre-trained models of word embedding, FastText and Glove. Additionally, we utilized the feature-based approach of Indonesian BERT architecture to generate more contextualized word embeddings. We reported average results in terms of AUC, MRR, NDCG@5 and NDCG@10. For the training objective, we primarily focus on parameter tuning through minimizing cross-entropy loss (with negative sampling) as the straightforward classification objective [44].

C. Text Embedding Models

As part of our benchmarking process, we trained three word embedding models—Word2Vec, Glove, and FastText—using the Indonesian Wikipedia corpus extracted from the Indo4B corpus [39]. Furthermore, we also trained a FastText model using the entire Indo4B corpus. For our BERT-based models, we leveraged models from two distinct sources, namely indolem [38] and indobert [39]. These pretrained BERT models played a crucial role in our research, serving as integral components in our investigations into text embeddings and their impact on news recommendation systems.

D. News Recommendation Models

In this paper, we assess several C-AG models, each distinguished by their News Embedding (NE) component, which essentially determines how they embed the clicked news articles: (1) The LSTUR [9], focuses on learning user representations using recurrent networks. In this model, a short-term

user embedding is generated from the clicked news articles using a Gated Recurrent Unit (GRU) [45]. This short-term embedding is then combined with a long-term embedding, which is constructed by initializing with random values and fine-tuning it during training. In this research, we use two types of LSTUR model which is LSTUR-con and LSTUR-uni. (2) NAML [11] employs additive attention [46] to encode users' preferences. This means that NAML utilizes an attention mechanism to capture and emphasize important aspects of a user's preferences when making recommendations. Additionally, our primary focus for this experiment will be on NAML, given its strong performance and adaptability, especially in accommodating adjustments to the included news information in the model. (3) NRMS [10] adopts a more intricate approach to learn user representations. It employs a two-layer encoder, consisting of multi-head self-attention [36] and additive attention. This design allows NRMS to capture and process user information from different perspectives, potentially leading to more comprehensive user representations for improved recommendations. (4) TANR [12] propose a new encoder that is trained to learn topic-aware news representations by simultaneously training it on an auxiliary topic classification task.

On the other hand, we examine one specific C-AW model to gain insights into their performance and effectiveness. We evaluate CAUM [13], there is a fusion of two key components: A candidate-aware self-attention network, which is used to model extensive connections among clicked news items while taking into account the specific candidate news; and a candidate-aware convolutional network (CNN) employed to capture immediate user interests from nearby clicks, also influenced by the content of the candidate news. Ultimately, the user's candidate-aware embedding is derived by considering both the long-range and short-term representations.

IV. RESULTS & ANALYSIS

Table II shows evaluation of the NRS has produced diverse performance results across various testing scenarios. These thorough assessments have uncovered valuable insights, particularly highlighting the competitive performance of two prominent models, NAML and CAUM, in terms of accuracy and relevancy. However, a notable distinction emerges as NAML demonstrates a significant advantage in terms of runtime efficiency, notably surpassing the CAUM model in processing speed. This unique combination of high performance and reduced computational overhead presents a compelling argument.

Using the NAML model as our main framework, we conduct a more in-depth assessment of how news information influences news representation and, consequently, its impact on news recommendation model performance. The results are shown in Table III. Our approach involves the independent processing of various news text information and categories. We merge these elements using an attention network, constructing a holistic representation that encompasses the collective impact of various news components. This method enables us to systematically investigate the synergistic effects of diverse news content on enhancing model performance.

Regarding news representation, the results in Table III reveal a key insight: relying on a single source of news infor-

mation may not suffice for optimal representation. The notable improvement lies in not depending solely on a single type of news information but integrating various types, such as title, abstract, and tags. This approach is suggested to potentially result in better performance, emphasizing the need to leverage the complementary aspects of diverse news components. By incorporating a variety of information sources, the model can achieve a more comprehensive and nuanced understanding of news articles, ultimately improving the quality of news representation.

title	tags
Launched Tomorrow, Free Homecoming by State-Owned Enterprises (BUMN) Can Be Joined by 65,603 People	Free Homecoming, Homecoming Together with BUMN
"The Glory" Director, Ahn Gil Ho, Denies Being a Bully Perpetrator	"The Glory," "The Glory" Director, Director of "The Glory" Bullying

(a) A User clicked news articles in a session

title	tags
"APA", Female "Whisperer" of Mario Dandy Who Claims AG Received Poor Treatment from D	Mario Dandy Satrio, Mario Dandy Satrio Abuses the Child of GP Anso's Official, Who is the Figure in the Mario Dandy Case, Who is the Figure, Anastasya Pretya Amanda
Head of Makassar Customs Admits Feeling Pressured	Head of Makassar Customs Andhi Pramono, Head of Makassar Customs Clarified by the Corruption Eradication Commission (KPK)
Not Present at the Scene, Former Girlfriend of Mario Objects to Being Linked to D's Assault	Mario Dandy Satrio, Mario Dandy Satrio Abuses the Child of GP Anso's Official, Mario Dandy Case, Mario Dandy Breakup, Who is Mario Dandy's Whisperer?

(b) Recommendation based on clicked news using tags

title	tags
"APA", Female "Whisperer" of Mario Dandy Who Claims AG Received Poor Treatment from D	Mario Dandy Satrio, Mario Dandy Satrio Abuses the Child of GP Anso's Official, Who is the Figure in the Mario Dandy Case, Who is the Figure, Anastasya Pretya Amanda
Korean Dramas and Indian Films Share Many Similarities, Here's the Explanation	India, Korea, Korean Drama, K-Drama, Indian Film
Ministry of Finance Holds 134 Tax Office Employees Have Shares Registered in Their Wives' Names	Ministry of Finance, Tax Office Employees, Tax Office Employees Have Shares

(c) Recommendation based on clicked news without using tags

Fig. 5. Recommendation results by using and without news tags for the same impression. The news clicked by the user in this impression is in blue and bold.

Table III provides further insights, notably indicating a significant improvement in model performance when news tags are incorporated alongside the abstract as part of the news information. This observation underscores the positive impact of integrating news tags into the model, particularly when combined with other textual elements like the abstract. Such integration significantly increases the performance of the NAML method by 0.88%, 3.11%, 5.86%, and 2.59% for AUC, MRR, NDCG@5 and NDCG@10, respectively.

The example of the NAML model using and without tags is illustrated in Fig. 5. Based on this example, it is evident

TABLE II. THE PERFORMANCE ON VARIOUS NEWS RECOMMENDATION MODELS. SIGNIFICANT DIFFERENCES WITH RESPECT TO BASELINES LSTUR/TANR/NRMS/CAUM ARE INDICATED USING †/ ‡ / * /◇ FOR $p < 0.05$. RUN TIME IS REPRESENTED IN THE FORMAT HH:MM:SS

Methods	AUC	MRR	nDCG@5	nDCG@10	Run Time
LSTUR- <i>ini</i> †	0.5064	0.1377	0.1131	0.1885	00:53:26
LSTUR- <i>con</i> †	0.5089	0.1426	0.1177	0.1939	00:53:05
TANR*	0.5038	0.1380	0.1136	0.1890	00:53:19
NRMS◇	0.5266†‡*	0.1473†*	0.1273†‡*	0.2029†‡*	01:21:17
CAUM◇	0.5378†‡*◇	0.1534†‡*◇	0.1345†‡*◇	0.2120†‡*◇	03:44:50
NAML	0.5375†‡*◇	0.1529†‡*◇	0.1354†‡*◇	0.2113†‡*◇	01:51:31°

TABLE III. PERFORMANCE METRICS BEFORE AND AFTER ADDING THE NEWS TAGS. SIGNIFICANT DIFFERENCES ARE INDICATED USING † FOR $p < 0.05$ OR ‡ FOR $p < 0.001$.

News Information	AUC	MRR	nDCG@5	nDCG@10
Title	0.5247	0.1450	0.1236	0.2001
Abstract	0.5073	0.1400	0.1165	0.1909
Content	0.5311	0.1495	0.1296	0.2062
Title + Category/Subcategory	0.5399	0.1522	0.1338	0.2109
Title + Abstract	0.5253	0.1489	0.1295	0.2041
Title + Content	0.5345	0.1500	0.1302	0.2081
Title + Category/Subcategory + Abstract	0.5359	0.1502	0.1316	0.2081
Title + Category/Subcategory + Content	0.5394	0.1542	0.1371	0.2129
Title + Abstract + Content	0.5357	0.1518	0.1322	0.2095
Title + Category/Subcategory + Abstract + Content	0.5366	0.1509	0.1314	0.2088
Title + Tags	0.5216	0.1420	0.1195	0.1963
Abstract + Tags	0.5273‡	0.1503‡	0.1323‡	0.2059‡
Content + Tags	0.5317	0.1503	0.1314	0.2069
Title + Category/Subcategory + Tags	0.5389	0.1520	0.1337	0.2119
Title + Abstract + Tags	0.5248	0.1482	0.1288	0.2045
Title + Content + Tags	0.5273	0.1472	0.1253	0.2024
Title + Category/Subcategory + Abstract + Tags	0.5325	0.1506	0.1321	0.2085
Title + Category/Subcategory + Content + Tags	0.5415	0.1550	0.1379	0.2154
Title + Abstract + Content + Tags	0.5313	0.1499	0.1308	0.2066
Title + Category/Subcategory + Abstract + Content + Tags	0.5413†	0.1556†	0.1391†	0.2142†

that the model that incorporates news tags performs better in terms of relevancy than the one that does not use tags. In detail, when evaluating the top recommendations generated by the model without incorporating tags, none of the suggested articles align with user-clicked preferences. However, with the inclusion of tags, the model’s recommendations exhibit a notable improvement. Two out of the top three suggestions are now articles that users have previously engaged with, showcasing the effectiveness of integrating tags in refining the recommendation outcomes.

TABLE IV. RESULTS ON DIFFERENT NUMBER OF CLICKED NEWS BY USERS

N Clicked News	AUC	MRR	nDCG@5	nDCG@10
1	0.5418	0.1561	0.1400	0.2160
5	0.5383	0.1516	0.1344	0.2104
10	0.5385	0.1544	0.1356	0.2130
20	0.5411	0.1533	0.1358	0.2132
50	0.5413	0.1556	0.1391	0.2142

The analysis of user interactions with news articles has yielded consistent findings across varying levels of user engagement with news content. These findings have been summarized and are presented in Table IV. Surprisingly, users who interact with only a few news articles demonstrate outcomes comparable to those who engage with a more extensive collection of articles. This intriguing observation suggests that the specific selection of the last N news items may not exert a substantial influence on the observed results. This insight prompts further exploration into the dynamics of user interactions and their impact on the recommendation process, potentially leading to more refined and nuanced approaches to

personalizing news recommendations.

TABLE V. RESULTS ON DIFFERENT TEXT EMBEDDING

Embedding Model	AUC	MRR	nDCG@5	nDCG@10
word2vec-wiki-id	0.5327	0.1515	0.1323	0.2086
glove-wiki-id	0.5378	0.1534	0.1345	0.2120
fastText-wiki-id	0.5355	0.1533	0.1344	0.2124
fastText-indo4b	0.5365	0.1551	0.1378	0.2120
indolem/indobert-base-uncased	0.5264	0.1499	0.1300	0.2053
indobenchmark/indobert-base	0.5354	0.1497	0.1302	0.2083
indobenchmark/indobert-large	0.5383	0.1530	0.1338	0.2123

In addition to this research, we undertook an in-depth examination of different text embedding methods and their performance when integrated into the NAML model. The comprehensive results of these evaluations can be found in Table V. Our objective was to ascertain whether utilizing BERT-based models, which have demonstrated remarkable capabilities in various natural language processing tasks, would yield a significant performance boost within the NAML method.

However, upon thorough analysis, our findings suggest that the differences in performance between word embeddings and BERT-based models are not statistically significant when applied to the NAML model. In other words, the NAML model does not exhibit a substantial improvement in recommendation performance when integrated with BERT-based text embeddings compared to more conventional word embeddings. This emphasizes carefully choosing the right models and text embedding methods when optimizing recommendation systems.

V. CONCLUSION

This research investigates the effectiveness of neural news recommendation with attentive multi-view learning (NAML) method to perform news recommendation in the Indonesian language. We further propose to incorporate news tags information into the NAML architecture in order to improve the semantics of news representation, which is aimed at enhancing the Indonesian news recommendation system. According to our experimental results, the NAML method can significantly outperform some state-of-the-art neural models, such as LSTUR, TANR, and NRMS, in Indonesian news recommendation task. Although NAML demonstrates effectiveness similar to the CAUM method, it maintains superiority in terms of efficiency.

Furthermore, our investigation into news representation underscores the significance of diversifying news information sources. Combining multiple types of news data, rather than relying solely on a single source, has shown promising potential for enhancing model performance. Subsequently, the incorporation of news tags into the NAML architecture has demonstrated notable effectiveness, resulting in an enhancement of 3.11% for MRR and 5.86% for NDCG@5 in recommendation system performance.

In addition, we conducted performance assessments of the NAML model using various text embeddings. Upon initial examination, it becomes evident that there's no significant improvement observed when utilizing different types of word embedding as news representations, i.e., Word2Vec, FastText, Glove, and IndoBERT. Lastly, our analysis of user interactions with news articles, has unveiled intriguing findings regarding user engagement levels and their influence on recommendation outcomes. We found that users who interact with only a few news articles demonstrate outcomes comparable to those who engage with a more extensive collection of articles. This insight encourages further exploration into personalized news recommendation strategies, with the potential to refine and tailor recommendations based on a deeper understanding of user behaviors.

VI. FUTURE WORK

This study's findings offer insights and several promising directions for future research aimed at enhancing news recommendation systems. First, we can fine-tune the model parameters using a contrastive learning objective, specifically supervised contrastive loss. This approach can be employed to improve the separation between clicked and not-clicked news articles within the representation space. Second, it is potential to explore the automatic extraction of entities from news articles using NER (Named Entity Recognition) method as an integral part of the news recommendation model process. This approach can help uncover the relatedness of entities between relevant and non-relevant news articles, providing valuable context for improving recommendation robustness. Finally, research in news recommendation systems should diversify beyond contextual user history, adopting different methodologies for broader insights and improved system performance.

ACKNOWLEDGMENT

This research was funded by the Directorate of Research and Development, Universitas Indonesia, under Hibah PUTI Pascasarjana 2023 (Grant No. NKB-020/UN2.RST/HKP.05.00/2023).

REFERENCES

- [1] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, and D. Sampath, "The youtube video recommendation system," in *Proceedings of the Fourth ACM Conference on Recommender Systems*, ser. RecSys '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 293–296.
- [2] A. S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: scalable online collaborative filtering," in *WWW*, 2007, pp. 271–280.
- [3] J. Freyne, M. Jacovi, I. Guy, and W. Geyer, "Increasing engagement through early recommender intervention," in *Proceedings of the Third ACM Conference on Recommender Systems*, ser. RecSys '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 85–92.
- [4] C. Wu, F. Wu, Y. Huang, and X. Xie, "Personalized news recommendation: Methods and challenges," *ACM Trans. Inf. Syst.*, vol. 41, no. 1, jan 2023.
- [5] X. Meng, H. Huo, X. Zhang, W. Wang, and J. Zhu, "A survey of personalized news recommendation," *Data Science and Engineering*, Sep. 2023.
- [6] J. B. Baskoro and E. Yulianti, "Sgcf: Inductive movie recommendation system with strongly connected neighborhood sampling," *Jurnal Ilmu Komputer dan Informasi*, vol. 15, no. 1, pp. 55–67, Feb. 2022.
- [7] Y. Liu, J. Miyazaki, and Q. Chang, "Jointly learning propagating features on the knowledge graph for movie recommendation," in *Database and Expert Systems Applications*. Cham: Springer International Publishing, 2022, pp. 3–16.
- [8] Y. Liu and J. Miyazaki, "Knowledge-aware attentional neural network for review-based movie recommendation with explanations," *Neural Computing and Applications*, vol. 35, no. 3, pp. 2717–2735, Jan. 2023.
- [9] M. An, F. Wu, C. Wu, K. Zhang, Z. Liu, and X. Xie, "Neural news recommendation with long- and short-term user representations," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 336–345.
- [10] C. Wu, F. Wu, S. Ge, T. Qi, Y. Huang, and X. Xie, "Neural news recommendation with multi-head self-attention," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 6389–6394.
- [11] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, and X. Xie, "Neural news recommendation with attentive multi-view learning," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, ser. IJCAI'19. AAAI Press, 2019, p. 3863–3869.

- [12] C. Wu, F. Wu, M. An, Y. Huang, and X. Xie, "Neural news recommendation with topic-aware news representation," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 1154–1159.
- [13] T. Qi, F. Wu, C. Wu, and Y. Huang, "News recommendation with candidate-aware user modeling," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 1917–1921.
- [14] Adiyansjah, A. A. S. Gunawan, and D. Suhartono, "Music recommender system based on genre using convolutional recurrent neural networks," *Procedia Computer Science*, vol. 157, pp. 99–109, 2019, the 4th International Conference on Computer Science and Computational Intelligence (ICCSICI 2019) : Enabling Collaboration to Escalate Impact of Research Results for Society.
- [15] M. Martijn, C. Conati, and K. Verbert, "'knowing me, knowing you': personalized explanations for a music recommender system," *User Modeling and User-Adapted Interaction*, vol. 32, no. 1, pp. 215–252, Apr. 2022.
- [16] W. G. Assuncao, L. S. G. Piccolo, and L. A. M. Zaina, "Considering emotions and contextual factors in music recommendation: a systematic literature review," *Multimedia Tools and Applications*, vol. 81, no. 6, pp. 8367–8407, Mar. 2022.
- [17] K. Dinnissen and C. Bauer, "Fairness in music recommender systems: A stakeholder-centered mini review," *Frontiers in Big Data*, vol. 5, 2022.
- [18] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, p. 5–53, Jan 2004.
- [19] O. Phelan, K. McCarthy, M. Bennett, and B. Smyth, "Terms of a feather: Content based news recommendation and discovery using twitter." in *ECIR*, 2011, pp. 448–459.
- [20] S. Okura, Y. Tagami, S. Ono, and A. Tajima, "Embedding-based news recommendation for millions of users," in *KDD*, 2017, pp. 1933–1942.
- [21] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2014.
- [22] J. Lian, F. Zhang, X. Xie, and G. Sun, "Towards better representation learning for personalized news recommendation: a multi-channel deep fusion approach." in *IJCAI*, 2018, pp. 3805–3811.
- [23] H. Wang, F. Zhang, X. Xie, and M. Guo, "Dkn: Deep knowledge-aware network for news recommendation," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018, p. 1835–1844.
- [24] F. Wu, Y. Qiao, J.-H. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, and M. Zhou, "MIND: A large-scale dataset for news recommendation," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, Jul. 2020, pp. 3597–3606.
- [25] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *ArXiv*, 2013.
- [26] E. Yulianti, N. Pangestu, and M. Jiwanggi, "Enhanced textrank using weighted word embedding for text summarization," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 5, pp. 5472–5482, Oct. 2023.
- [27] S. Abdulateef, N. A. Khan, B. Chen, and X. Shang, "Multidocument arabic text summarization based on clustering and word2vec to reduce redundancy," *Information*, vol. 11, no. 2, p. 59, Jan 2020.
- [28] M. Gambhir and V. Gupta, "Deep learning-based extractive text summarization with word-level attention mechanism," *Multimedia Tools and Applications*, vol. 81, no. 15, pp. 20 829–20 852, Jun. 2022.
- [29] T. V. Rampisela and E. Yulianti, "Academic expert finding in indonesia using word embedding and document embedding: A case study of fasilkom ui," in *2020 8th International Conference on Information and Communication Technology (ICOICT)*, 2020, pp. 1–6.
- [30] R. C. Lima and R. L. T. Santos, "On extractive summarization for profile-centric neural expert search in academia," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 2331–2335.
- [31] N. Nissa and E. Yulianti, "Multi-label text classification of indonesian customer reviews using bidirectional encoder representations from transformers language model," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 5, pp. 5641–5652, Oct. 2023.
- [32] B. Jang, M. Kim, G. Harerimana, S.-u. Kang, and J. W. Kim, "Bi- lstm model to increase accuracy in text classification: Combining word2vec cnn and attention mechanism," *Applied Sciences*, vol. 10, no. 17, p. 5841, Aug 2020.
- [33] N. E. Aoumeur, Z. Li, and E. M. Alshari, "Improving the polarity of text through word2vec embedding for primary classical arabic sentiment analysis," *Neural Processing Letters*, vol. 55, no. 3, pp. 2249–2264, Jun. 2023.
- [34] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543.
- [35] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.
- [36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.
- [37] A. F. Aji, G. I. Winata, F. Koto, S. Cahyawijaya, A. Romadhony, R. Mahendra, K. Kurniawan, D. Moeljadi, R. E. Prasojo, T. Baldwin, J. H. Lau, and S. Ruder,

- “One country, 700+ languages: NLP challenges for underrepresented languages and dialects in Indonesia,” in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 7226–7249.
- [38] 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*. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 757–770.
- [39] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, and A. Purwarianti, “IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding,” in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*. Suzhou, China: Association for Computational Linguistics, Dec. 2020, pp. 843–857.
- [40] Y. Takeda, I. Ohmukai, S. ADACHI, and T. Kato, “News recommendation method using group tags and thesauruses,” *Transactions of Japan Society of Kansei Engineering*, vol. 8, no. 3, pp. 813–818, 2009.
- [41] Y. Shen, P. Ai, Y. Xiao, W. Zheng, and W. Zhu, “A tag-based personalized news recommendation method,” in *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2018, pp. 964–970.
- [42] A. Iana, G. Glavas, and H. Paulheim, “Simplifying content-based neural news recommendation: On user modeling and training objectives,” in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 2384–2388.
- [43] D. Valcarce, A. Bellogín, J. Parapar, and P. Castells, “Assessing ranking metrics in top-n recommendation,” *Information Retrieval Journal*, vol. 23, no. 4, pp. 411–448, Aug. 2020.
- [44] C. Wu, F. Wu, and Y. Huang, “Rethinking infonce: How many negative samples do you need?” in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, L. D. Raedt, Ed. International Joint Conferences on Artificial Intelligence Organization, 7 2022, pp. 2509–2515, main Track.
- [45] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734.
- [46] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” *ArXiv*, vol. 1409, 09 2014.