EMITTER International Journal of Engineering Technology
Vol. 14, No. 1, June 2026, pp. 184~199
DOI : 10.24003/emitter.v14i1.1097

Consistency of News Titles and Contents With Multi Level Classification Using Bidirectional Encoder Representations from Transformer (BERT)

Afrida Helen, Akmal, Muhammad Razzaaq Fadillah

Department of Computer Science, Padjadjaran University, Bandung, Jawa Barat, Indonesia
Corresponding Author: helen@unpad.ac.id

Received January 18, 2026; Revised March 20, 2026; Accepted June 10, 2026

Abstract

News is written in a standard format that consist of a title followed by the content. The title must be present a content of news. Sometimes found that news title did not represent the content, causing readers to be disappointed when they finish reading the whole news. This research proposes multi-class and multi level classification using BERT algorithm to detect the suitability of news titles and their contents. The first level performs classification for complete data and the second level performs classification for data that has been separated into 2 parts, Title dan Conten. The dataset pertains to the realm of Covid-19 news because its substantial volume and diverse range of conversation topics. It is sufficient to be used as a dataset. We select several news topic classes that are often discussed, that are health, politic, and economy. In order to determine suitability, we categorise both the news Title and the news Content into predefined topic classes. The dataset exhibits an inbalance for each topic class. This research proposed a model to determine the appropriateness of both the news Title and the news Content. We use Bidirectional Encoder Representations from Transformers for building model. BERT is one of the state-of-the-art in the field of Natural Language Processing (NLP). Multi-layer classification that separates dataset (title and content) determine the consistency between them is very effective in increasing accuracy, and the BERT model is very supportive in solving linguistic problems well. This research succeeded in providing an accuracy 15% greater compared to previous research.

Keywords: Consistency, Multilevel Classification, Natural Language Processing, Title, Content.

1. INTRODUCTION

Various news topics have filled mass media and have grown rapidly with the advent of internet technology, including online news media in Indonesia. The delivery of news through digital platforms has proven to be highly

effective in reaching a wide audience quickly due to its easy accessibility across multiple devices. In practice, every news article has a specific topic that serves as its main focus. This topic can generally be identified through both the headline and the content of the article, meaning that there should ideally be consistency between the two [1]. However, in reality, inconsistencies between headlines and article content are often found. Not infrequently, writers or content providers deliberately use sensational or attention-grabbing headlines that do not accurately represent the actual content of the news. This phenomenon is commonly known as clickbait, which aims to increase the number of visits or clicks from readers.

The discrepancy between headlines and content can have negative impacts, such as reducing public trust in media and causing disappointment among readers who feel misled. Furthermore, this practice can obscure the true information intended to be conveyed, thereby affecting the overall quality of information dissemination in society. The primary motivation behind such actions is generally economic, as higher traffic to a news link often leads to greater revenue through advertisements or other forms of monetization. Therefore, readers are encouraged to be more critical and selective when consuming news and not be easily influenced by provocative or misleading headlines.

2. RELATED WORKS

Several studies in the field of social sciences have examined the alignment between news headlines and the corresponding content. Studies [2–4] for instance, the suitability between headlines and article bodies by comparing the topics represented in each. In these approaches, a news item is considered inappropriate or inconsistent when the topics extracted from the headline and the content differ significantly. In addition, other researchers have explored this issue using machine learning techniques [5,6] developing automated models to detect mismatches between headlines and news content more efficiently and at scale. These models typically rely on topic classification, semantic similarity, or text representation methods to evaluate whether the headline accurately reflects the substance of the article.

This study investigates the suitability between news and content by comparing the general topic of discussion with a predefined target class. The approach is motivated by the rapid growth of online news platforms, where discrepancies between and have become increasingly common [7]. Such inconsistencies often arise from the use of attention-grabbing that may not fully represent the underlying information. However, it is important to recognize that enforcing overly strict standards of suitability may lead to valid news articles being misclassified as inappropriate. News writing often involves nuanced relationships between and content, where a headline may highlight a particular angle or aspect of a broader discussion. Moreover, similarity in general topic classification does not necessarily guarantee alignment in the specific issues being addressed. For example, if both the and

the content fall under the same general topic, such as health, the system may classify them as consistent. Nevertheless, the headline might discuss a specific health concern, while the content elaborates on a different aspect within the same domain. This indicates that topic-level matching alone may not be sufficient to fully capture the semantic alignment between headlines and news content. Therefore, more refined approaches that consider contextual and issue-level similarities are needed to improve the accuracy of suitability assessments.

3. ORIGINALITY

In this study, we propose three news topic that we consider popular compared to other topics. If the topic is close to the news title, then the topic must also be close to the news content [8]. This study conducted a multilevel classification to investigate the suitability of the title and content topics of the news using the BERT model. Bidirectional Encoder Representations from Transformers (BERT) is one of the state-of-the-art in the field of Natural Language Processing. The dataset was crawled based on keywords from the top tree online media: cnnindonesia.com, kompas.com, detik.com. The keywors are Politics, Health, and Economy. The first level, trains the BERT model based on three classes. The second level, separates the title and content into two diferent datasets. Then trains the BERT model for each. This means, the BERT model is trained separately on the title and the news content. Each dataset has three classes or labels; Politics, Health, and Economy. The labels from the training model on the title and content are matched, if the title label and the news content label are the same. The accuracy of previous research was 0.78%, while this research achieved 0.93%.The results, multi-level classification provide an accuracy value much better than the single-level classification of 15%. However, obtaining a large dataset can be a challenging task. This study tries to take a case study of Covid-19 news because the dataset of this news is very large. We want to show the effectiveness and high accuracy of the model we developed.

4. SYSTEM DESIGN

The development of computational methods and algorithms greatly helps humans to do everything automatically. One of them is the problem of classification and prediction for the field of natural language processing. Especially for the field of natural language processing requires a large dataset, which makes the model learning process more sensitive when the input data changes.

4.1 News Articles

The standard format for writing a news article is to have a title and content. The title stays at the top and then followed by the content of the news. Each news has a topic of discussion. The topic can be known from the Title and from the content. The topic of the title must be the same as the topic of the

news Content [9]. The difference of the topic conveyed in the title and the content can cause a problem known as the Headline Incongruence Problem [10]. The Headline Incongruence Problem has the potential to mislead people who read the news, especially if they only read the title. News writers deliberately arouse the enthusiasm of readers when reading the title. They hope that readers will want to open the news link. In official news providers this may be very rare, but on non-formal news sites, the inconsistency of the title and the content is often used to increase the viewer rating of the site. In addition, the writer also tries to lead the reader's opinion to something else, such as black campaign.

In this case, the reader must be smart and careful. However, it is difficult to know the inconsistency of the title and content without reading both. By reading both, the topic discussed in each part of the news can be known and the consistency between the two can be analyzed. This process can actually be automated using machine learning models [11].

4.2 Deep Learning BERT for Text Task

The Bidirectional Encoder Representations from Transformers (BERT) model is one of the deep learning models. Currently, the BERT model is the best and most trusted in the field of Natural Language Processing (NLP)[12]. An important concept in BERT is the bidirectional training process in language representations. Language representations go through a pre-training process to reduce the need for custom architectures for specific tasks. BERT can represent both single sentences and paired sentences (for example pairs of questions and answers sentences in a question answering task) in one token sequence. This token is the result of tokenization using the WordPiece algorithm [13].

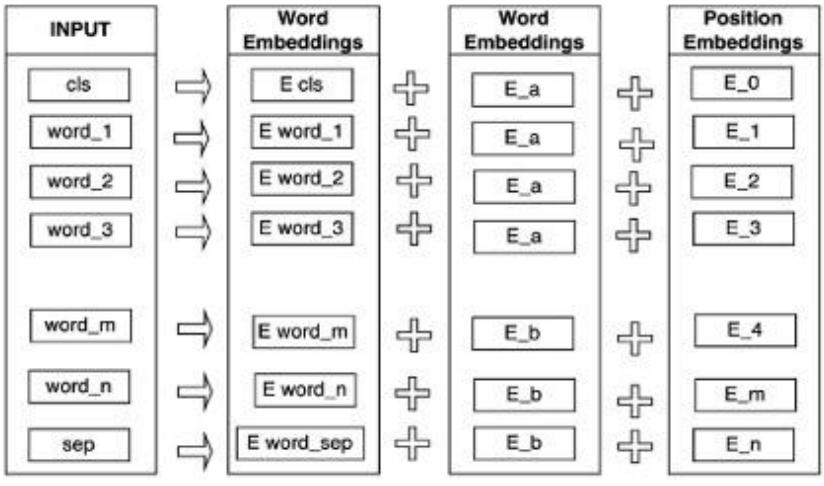

Figure 1. The representation of text input using BERT algorithm

To differentiate sentences, BERT will separate them in the text using special tokens ('[SEP]'). Then, a learned embedding is added for each token

indicating whether it belongs to sentence A or sentence B. In addition to the separator token, there is also a special classification token ('[CLS]') added to the beginning of each sentence as shown in Figure 1, the model input is denoted by E, the last vector in the hidden layer using the special sign '[CLS]' is denoted by C ∈ R^H, and the last for the i-th input token is denoted by T_i ∈ R^H. For a single token, the input representation is the sum of the token, segment, and position embeddings.

The BERT model has many layers of bidirectional transformer encoders. The encoder architecture in BERT is based on the original implementation of Transformer [14]. The number of encoder layers varies greatly depending on the model, as does the size of the hidden layers and the amount of attention. This research uses the BERT model. The BERT model has 12 encoder layers, 768 hidden layers, and 12 attention layers. So, the number of parameters in this model is 110 million. In general, BERT consists of two stages, namely the pre-training stage and the fine-tuning stage. The pre-training stage involves training the model with un-labelled data from various pre-training tasks. The BERT model first initializes with pre-training parameters for the fine-tuning stage. Next, all parameters undergo a fine-tuning process using labelled data. This data originates from tasks that the BERT model will solve, commonly known as downstream tasks in NLP.

4.3 Collect Dataset

The global outbreak of the Corona virus has become a topic of conversation in various parts of the world, including Indonesia. Since the pandemic began, for two years, news about this outbreak has filled online news in Indonesia. From observations, the news written discusses the development of cases, steps to overcome the outbreak, death rates, developments in vaccination programs, handling of socio-economic impacts, and etc. Online news portals usually categorize this news to make it easier for users to find the information they want. The dominant news categories are Politics, Health [15], and Economy [16,17].

Web crawling is a strategy or way to collect news from the world wide web as a form of awareness of the rapid development of the internet [18]. The development of internet technology allows web crawling from various locations around the world. The purpose of this activity is to archive data from various sites. In general, the web crawling process follows the same procedure. The first step involves verifying the page for download. We perform a check to determine the ability of the page to be downloaded and its suitability for robot used. Only the accessed page initiates the download. The crawler will extract each page related to the accessed page. We save these pages for later download, then we extract all the words and store them in a database related to the current page. Web crawlers can collect other information, such as additional content and language on the page, and can store the outline of the current page and refresh the date information handled so that the framework knows when to check the page again. Depending on the purpose, the results

will be stored in a file system or database for later retrieval [19]. In addition, there are 2 options that can be done either manually or automatically. In this study, we define web crawling and web scraping as methods used in the web scraping process. This study integrates each stage for a smooth process and a comprehensive research design. Although we obtained over 24,000 news items, after preprocessing, only 2,309 remained that matched our three proposed classes. The failure occurred when the model separated the title and content. They should have been equal, but the results were different.

4.4 The Research Design

Figure 2 shows the entire research stage. Starting from data collection, dataset preprocessing, BERT model development, training the dataset using the proposed model, and evaluating the model. The BERT model was developed separately between the title and content. So that both datasets are in different storage. Then a combined dataset test was carried out to find out whether the title label and the news content label were the same or in the same topic. If there is a similarity in the labels of both, it means that the title topic is the same as the topic discussed in the news content.

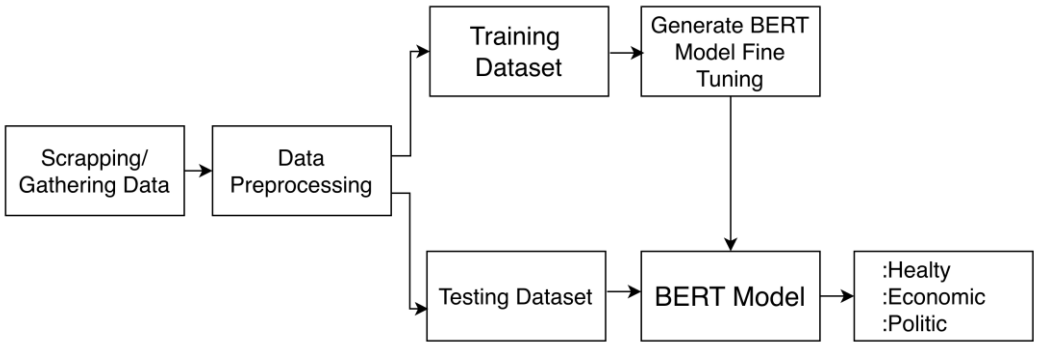

Figure 2. First Level; Building Cassification model for three classes, Politic, Ekonomi, and Health

4.4.1. Data Collection

This study used Indonesian-language news from online media as dataset. The dataset consists of news titles and their contents and is taken from the web scraping process. Web scraping was carried out on five national news sites, namely cnnindonesia.com, kompas.com, detik.com, merdeka6.com, and kumparan.com. These sites were chosen because they are popular and well-known by Indonesian people. In addition, these sites allow the use of robots for web crawling process. The web scraper was created using the library [20] and Selenium [21]. There are two types of web scrapers applied, namely web scrapers for static sites and web scrapers for dynamic sites.

In second level classification separate the title and content into two datasets. Then train the BERT model separately on the title dataset and the content dataset. BERT model classify dataset into three labels; Politics, Health, and Economy. The flowchart of this process is shown at Figure 3.

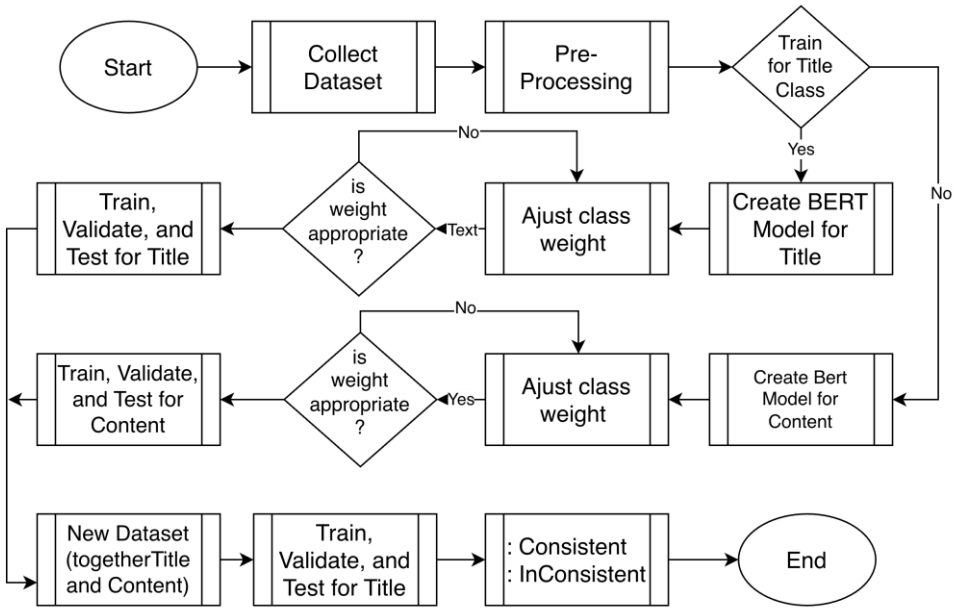

Figure 3. Second level classification; Start from collecting news titles and content, training for each data, combine them into one dataset, training, and testing to determine the consistency of title and content.

4.4.2. Dataset pre-processing

First, it is necessary to prepare the raw data become a dataset that is ready to be trained. It should be noted that different deep learning models can cause differences in preprocessing techniques. In fact, there is a case where the BERT deep learning model (as used in this study) actually provides the best accuracy on a dataset that does not go through the preprocessing stage [22]. After studying the case, this study decided to apply the following dataset preprocessing techniques:

1) Remove special characters and URLs.
2) Remove stop words.
3) Fold cases.
4) Remove numbers.
5) Remove excess spaces.
6) Data labeling.
7) Tokenization

5. EXPERIMENT AND ANALYSIS

We use two steps experiment as mention before, experimen for first Level dan for second level.

5.1 Experiment Preparation

An illustration of the classification model architecture of this study can be seen in Figure 1. Model learning using pre-trained BERT. The pre-training model used is the same as the Hugging Face study (23). The classification model architecture is based on the BERT linear classification model for

sentiment analysis tasks [24]. The linear classification model processes the last hidden state of the '[CLS]' token using a dropout layer followed by a linear layer.

The BERT classification model is created separately, each to detect the news title label and then its content. The label are 'health', 'politics', 'economy', and 'others'. The 'others' label work for accommodating other data. The meaning is if the data does not member of the three labels provided. It should be noted that the token used by the classification model is the '[CLS]' token in the output layer only and that the embedding in the output layer of the BERT model is a contextual embedding. Then the BERT model maker assumes that the contextual embedding is sufficient to be embedded in only one token, namely the '[CLS]' token, where this embedding is sufficient to be used in the classification task. The hidden state of the '[CLS]' token in the last layer will be processed first by the linear layer and the Tanh activation function. The vector resulting from this processing is called the combined output [25]. This collected output vector will be used by the classification model to predict the data class.

Furthermore, the BERT classification model is trained for the loss function using the optimizer and scheduler. In this study, the authors used the cross-entropy loss function [26], AdamW optimizer [27]. During training, variations were made to certain hyperparameters to find the combination that produces the model with the best accuracy. The hyperparameters varied were the dropout probability in the classification layer, the learning rate in the optimizer, and the number of warm-up steps in the scheduler [28]. The variations in values used are as follows:

Probability of dropping out: 0.10, 0.15, 0.20, 0.25, 0.30
Learning rate: 9×10^(-6), 1×10^(-5), 2×10^(-5), 3×10^(-5), 4×10^(-5), 5× 10^(-5)
Heating step: 0.100

Apart from hyperparameters, there are other hyperparameters whose values remain constant during the training process, as follows:

Maximum token length: 30 (title model), 256 (content model)
Batch size: 16 (title model), 4 (content model)
Number of epoch: 4 (title model), 4 (content model)

Then, builds a model to analyse the consistency titles and content. The model will compare the label witch each others. Figure 4 shows the consistency detection flow between the news title and its content. The special case that needs to be discussed here is if both classes both fall into the 'other' category. The 'other' class is a class that contains news titles or content that are not included in the 'health', 'politics' or 'economic'. It is possible that the news title or content include into this class other than the three classes mentioned, for example automotive class, sports class, or other classes.

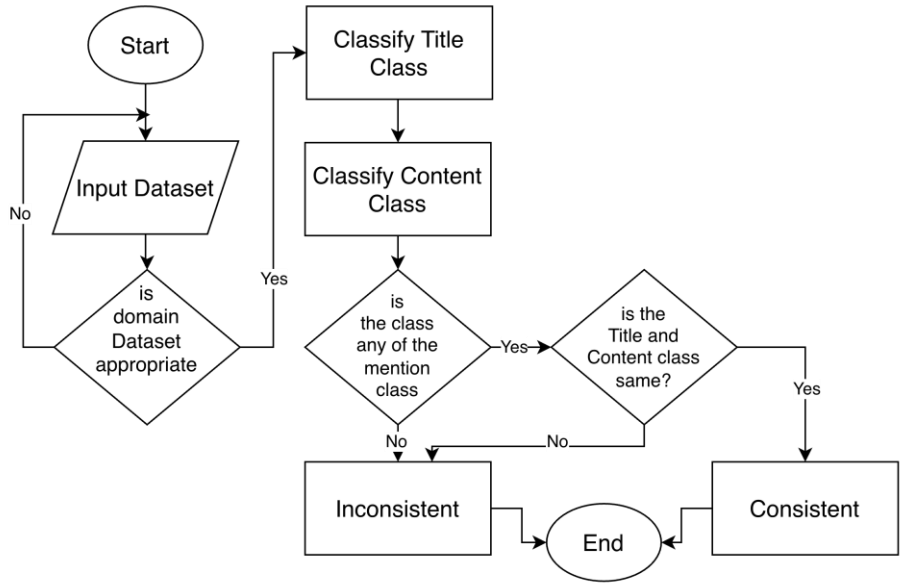

Figure 4. Flowchart for detecting consistency news title and content

To evaluate learning outcomes, testing was carried out using a separate dataset. Testing was carried out with three scenarios: testing the news title model, testing the news content model, and testing a combined model to detect consistency. Test results are assessed in the form of classification metrics. The classification metrics used to assess test results are precision, recall, F1 score, macro average [29], weighted average [30], and accuracy. Apart from that, the confusion matrix is also used to visualize testing on combined datasets.

In order to assess the results of learning, a distinct dataset was utilised for testing purposes. Three scenarios were utilised for testing: testing the news title model, testing the news content model, and testing a combined model to identify consistency. Evaluation of test outcomes is conducted using classification metrics. The classification metrics used to assess test results are precision, recall, F1 score [31], macro average [32], weighted average, and accuracy. Apart from that, the confusion matrix is also used to visualize testing on combined datasets

5.2 Experiment

The experiment used 24,465 training data consisting of news titles and content. Training data was downloaded using two types of web scrapers: static and dynamic. Static web scrapers retrieve news from cnnindonesia.com, kompas.com, detik.com, and merdeka6.com. While dynamic web scrapers retrieve news from kumparan.com. The web scraping results are stored in JSON format files. We perform text pre-processing using available tools, except for labeling. Preprocessing works sequentially, starting from Remove special characters and URLs, remove stop words, fold cases, remove numbers, remove excess spaces, and tokenization.

Next, label the dataset manually. Because the three labels are rather sensitive, we involve journalists to embed labels in the dataset. The dataset at the first level is 24,465 consisting of titles and content. Then, train the BERT model to classify dataset into three labels class. As a result, we get 2,309 data with political, health, and economic labels. The rest are included in the 'other' class. Unfortunately, at the second level, when the dataset is split into two, the number of titles and content is not the same. The title dataset is reduced to 1,699,000 as seen in Table 1.

Table 1. Original amount of data for each level

Level	Amount
Raw Data	24.465
Complete Dataset	2.309
Title dataset only	1.699
Content dataset only	2.309

The distribution class for title and content as seen in Figure 5 and 6. It can be seen that the data distribution is uneven for each type of class. The number of the 'health' class is much greater than the other classes. and the number of 'other' classes is much less than the other classes. Thus, the model will tend to classify data as 'health', but it is difficult to predict data that has the 'other' class. Since the data distribution is uneven in both datasets, the loss value calculation will apply a balance for each class. We use 10 folds cross validation for splitting dataset. This method is better than 80:20 splitting to reduce over fitting. The goal is to prevent the model from failing to predict fewer classes. Both models were trained. Table 2 show the best hyperparameters tuning that produce a best accuracy.

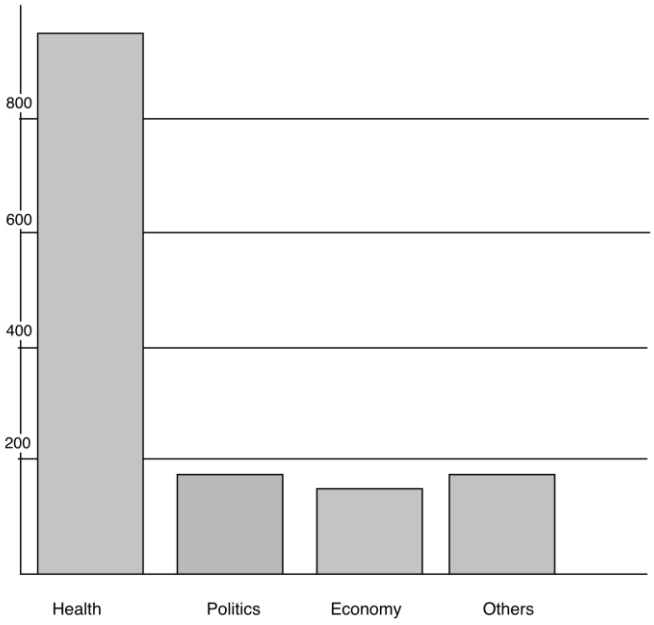

Figure 5. Data distribution for Politic, Health, and Economy class on Title

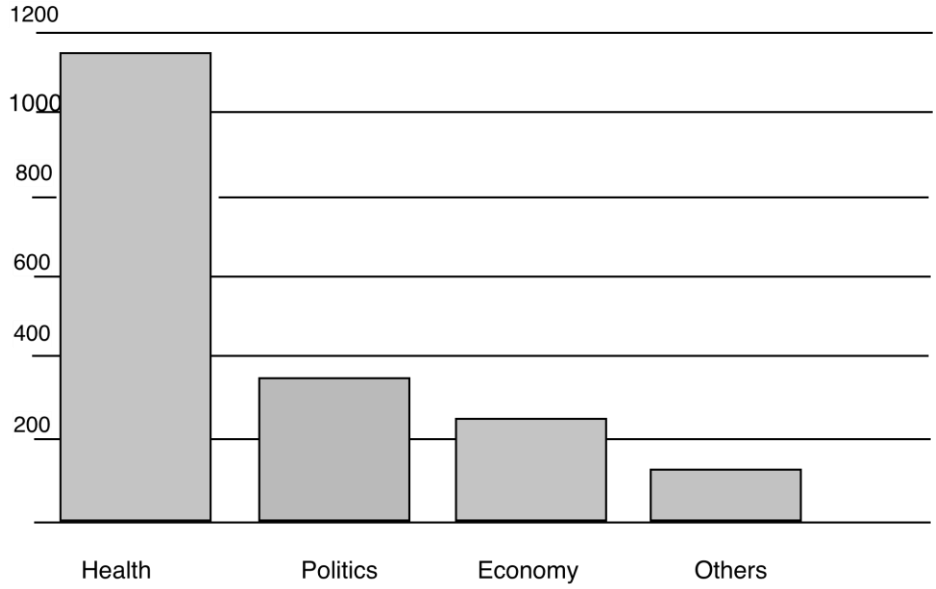

Figure 6. Data distribution for Politic, Health, and Economy class on Content

Table 2. The best hyperparameter for our models

Dataset	Dropout	LearningR	WarmupS	Acc
Title	0.15	$2x10^{-5}$	0	0.95
Content	0.1	$1x10^{-5}$	0	0.95

The fine-tuning for hyperparameter used the same dropout value for all tasks; 0.1 with recommended learning rate 2×10^(-5), 3×10^(-5), and 5×10^(-5). We found a different dropout value to the original BERT gave better results for news title. Apart from that, different learning rates also provide better results on the news content.

Table 3. The best results for the Title class using our proposed model

Classes	Precision	Recall	F1 Score
Health	0.95	0.99	0.97
Politic	0.94	0.83	0.88
Economy	0.94	0.94	0.94
Other	100	0.83	0.91
Macro average	0.96	0.90	0.92
Weighted average	0.95	0.95	0.95

Hyperparameters will differ depending on the dataset used. Studies believe this statement because it is proven based on experiments. Pre-training of both models gives the best results at a value of zero. Variations in the pretrain step are not mentioned as the original hyperparameter of BERT. However, the original source of the architecture uses a value of zero for the pretrain step so there is no difference. The classification metrics for the news title model test can be seen in Table 3.

Table 4. The best results for the Content class using our proposed model

Classes	Precision	Recall	F1 Score
Health	0.93	0.97	0.97
Politic	0.88	0.85	0.86
Economy	0.90	0.79	0.84
Other	1.00	0.88	0.93
Macro average	0.92	0.87	0.90
Weighted average	0.92	0.92	0.92

The metric of news title classification shows quite high accuracy, that is 0.95. However, the distribution of F1 scores for the news title is uneven. The class with the highest F1 score is the health class (0.97), while the class with the lowest F1 score is the politics class (0.88). The F1 score for the health class is high as it has a lot of training data and it is easy to differentiate from other classes. The political class has the lowest F1 score for quite a lot of data in this class. It is still misclassified as other classes. For the moment, the classification metrics for testing the news content model shown in Table 4.

We match the similarity of the title label and the news content, based on the trained model. At Figure 7, it can be seen that a lot of data with the label 'inconsistent' is classified as 'consistent'. This problem can also be seen through the following classification metrics as seen at Table 5. Overall F1_score is relatively high at 0.86. However, it should be noted that the precision value for the consistent class is the lowest compared to other classes (0.77). This is caused by the large number of false positive data for the 'consistent' class which comes from data in the 'inconsistent' class.

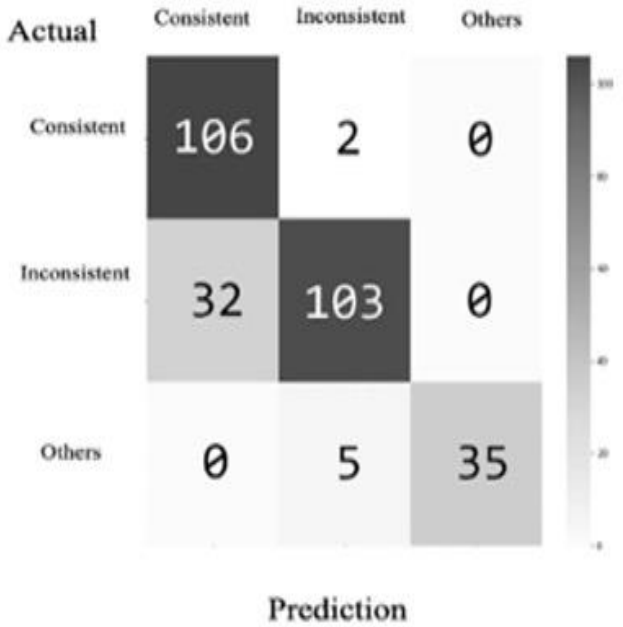

Figure 7. The confusion matrix represents a comparison of the actual and predicted values of consistency between the title and its content. Confusion matric to determine consistency title and content.

Table 5. The best F1-score value of our proposed model for predicting consistency between title and content

Classes	Precision	Recall	F1 Score
Consistent	0.77	0.77	0.86
Inconsistent	0.94	0.94	0.84
Others	1	0.88	0.93
Macro averege	0.90	0.87	0.88
Weighted avareage	0.88	0.86	0.86

6. CONCLUSION

Several events greatly influenced the results of this study. First, when collecting initial data, only the top three could be taken because other topics were biased. When separating the title and content, the number of both should be the same, but it did not happen. The title is less than the content. The discrepancy between the number of titles and the number of news content can be caused by several possibilities. First cause the news title is too short, so that the keywords are not stated literally in the title. The model assumes that the title does not match the keywords. Second, highlighting the political class in the title and content that does not match. The content discusses more about politics. It is suspected that there is a black campaign by the local government that will be elected. Beside that, the other class at level 2 is much less, because it was previously selected at the first level and there is a deliberate keyword not mentioned in the title so that the news content can be directed according to desire. Based on the experimental results, it can be concluded that the built model is feasible to use because it has good accuracy. If we consider other topics, the health topic has the highest precision level. It is known that Covid-19 is a viral infection. As a result, most articles mainly focus on health-related topics, while politics and economics are secondary discussion areas. The relationship between health issues and a country's policies and economy is undeniable. Both have a mutually beneficial relationship. Political policies should prioritize and promote public health, while economic policies should align and strengthen it. Based on our research findings, it is evident that certain individuals adopt names for themselves or others based on Covid-related news. The subject of health undergoes a transformation and becomes a political issue. Certain news articles have relevance to two different subjects. The most prominent difference lies in the realm of the health subject and the political subject. However, we need to pay attention to the data set so that the data distribution is balanced and there is more variation in the news class.

Acknowledgments:

We would like thank to DRHPM Padjadjaran University supporting this research under contract No. 1763/UN6.3.1/PT.00/2024.

REFERENCES

[1] Yoon S, Park K, Lee M, Kim T, Cha M, Jung K. **Learning to Detect Incongruence in News Headline and Body Text via a Graph Neural Network.** *IEEE Access.* 2021;9:36195–206. https://doi.org/10.1109/ACCESS.2021.3062029

[2] Goebel JT, Susmann MW, Parthasarathy S, El Gamal H, Garrett RK, Wegener DT. **Belief-consistent information is most shared despite being the least surprising.** *Sci Rep.*, Vol 14, No.1, 2024. https://doi.org/10.1038/s41598-024-56086-2

[3] Chandra A. **Konstruksi Berita Olahraga Pada Media Massa Internet (Analisis Framing Berita Konflik Marquez dan Rossi pada Sindonews.com).** *Jurusan Ilmu Komunikasi – Konsentrasi Jurnalistik*, 4(2), pp.1–15, 2017.

[4] Andersen K, Shehata A, Skovsgaard M, Strömbäck J. S**elective News Avoidance: Consistency and Temporality.** *Communication Research*. Vol. 53, No. 3, pp. 291–318 2026. https://doi.org/ 10.1177/00936502231221689

[5] Yoon S, Park K, Shin J, Lim H, Won S, Cha M, et al. **Detecting Incongruity between News Headline and Body Text via a Deep Hierarchical Encoder**. *AAAI.* Vol 33, No.01, pp. 791–800 Jul 2019. https://doi.org/10.1609/aaai.v33i01.3301791

[6] Zheng HT, Chen JY, Yao X, Sangaiah AK, Jiang Y, Zhao CZ. **Clickbait Convolutional Neural Network**. *Symmetry*. Vol. 10, No. 5, 2018 May 1, pp.138. https://doi.org/10.3390/sym10050138

[7] Dijk TA van. **News as discourse**. *Hillsdale (NJ): L. Erlbaum associates*; 1988. (Communication).

[8] Wirawan C. **Indonesian BERT base model (uncased).** *Hugging Face* [Internet]. 2021 [cited 2025 Apr 24]. Available from: https://huggingface.co/cahya/bert-base-indonesian-522M

[9] Musman A, Mulyadi N. **Jurnalisme Dasar: Panduan Praktis Para Jurnalis.** Yogyakarta: *Citra Media*; 2022.

[10] Chesney S, Liakata M, Poesio M, Purver M. I**ncongruent Headlines: Yet Another Way to Mislead Your Readers**. I*n: Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism* [Internet]. Copenhagen, Denmark: Association for Computational Linguistics; 2017 [cited 2026 Apr 24]. p. 56–61. https://doi.org/10.18653/v1/W17-4210

[11] Jang J, Cho YS, Kim M, Kim M. **Detecting incongruent news headlines with auxiliary textual information.** *Expert Systems with Applications*. 2022 Aug;199:116866. https://doi.org/10.1016/j.eswa.2022.116866

[12] Devlin J, Chang MW, Lee K, Toutanova K. BERT: **Pre-training of Deep Bidirectional Transformers for Language Understanding.** *In: Proceedings of the 2019 Conference of the North [Internet]. Minneapolis, Minnesota: Association for Computational Linguistics;* 2019 [cited 2026 Apr 24]. p. 4171–86. https://doi.org/10.18653/v1/N19-1423

[13] Guo M, Ainslie J, Uthus D, Ontanon S, Ni J, Sung YH, et al. **LongT5: Efficient Text-To-Text Transformer for Long Sequences.** I*n: Findings of the Association for Computational Linguistics:* NAACL 2022 [Internet]. Seattle, United States: Association for Computational Linguistics; 2022 [cited 2026 Apr 24]. pp. 724–36. https://doi.org/10.18653/v1/2022.findings-naacl.55

[14] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. **Attention Is All You Need.** *In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17).* Red Hook, NY, USA: Curran Associates Inc.; 2017. pp. 6000–6010.

[15] Velavan TP, Meyer CG. **The COVID-19 epidemic.** *Tropical Med Int Health*. vol. 25, no. 3, pp.278–280, 2020. https://doi.org/10.1111/tmi.13383

[16] Ahmad T, Khan M, Haroon, Musa TH, Nasir S, Hui J, et al. C**OVID-19: Zoonotic aspects.** *Travel Medicine and Infectious Disease.* Vol. 36. 101607 2020. https://doi.org/10.1016/j.tmaid.2020.101607

[17] Department of Biological Sciences, Federal University Gusau Zamfara, Nigeria., Mohammed M, Umar AY, Department of Biological Sciences, Nigerian Defence Academy, Kaduna State, Nigeria. **COVID-19 prevention and health promotion for the whole community**. *J Public Health Dis*. vol. 6, no. 2, pp. 31–38, 2023. https://doi.org/10.31248/JPHD2023.131

[18] Shamrat FMJM, Tasnim Z, Rahman AKMS, Nobel NI, Hossain SA. **An effective implementation of web crawling technology to retrieve data from the world wide web (www).** *International Journal of Scientific & Technology Research*. Vol. 9, No. 01, pp.1252–1256, 2020

[19] Zhao B. **Web Scraping**. *In: Schintler LA, McNeely CL, editors. Encyclopedia of Big Data [Internet]. Cham: Springer International Publishing;* 2017 [cited 2026 Apr 24]. p. 1–3. https://doi.org/10.1007/978-3-319-32001-4_483-1

[20] GitHub [Internet]. 2022. **Scrapy at a glance**. Available from: https://docs.scrapy.org/en/latest/intro/overview.html#scrapy-at-a-glance

[21] Selenium [Internet]. 2021. **The Selenium Browser Automation Project**. Available from: https://www.selenium.dev/documentation/

[22] Alzahrani E, Jololian L. **How Different Text-Preprocessing Techniques Using the BERT Model Affect the Gender Profiling of Authors.** *In: Proceedings of the 3rd International Conference on Machine Learning & Applications (CMLA).* Toronto, Canada; September 2021. https://doi.org/10.5121/csit.2021.111501

[23] Santos G. **What's Hugging Face?** *Towards Data Science* [Internet]. [cited 2022 Jun 19; updated 2025 Dec 28]. Available from: https://towardsdatascience.com/whats-hugging-face-122f4e7eb11a

[24] Gardazi NM, Daud A, Malik MK, et al. **BERT Applications in Natural Language Processing: A Review.** *Artificial Intelligence Review.* 2025;58:166. https://doi.org/10.1007/s10462-025-11162-5

[25] Jain SM. **Introduction to Transformers.** *In: Introduction to Transformers for NLP* [Internet]. Berkeley, CA: Apress; 2022 [cited 2026 Apr 24]. p. 19–36. Available from: https://link.springer.com/10.1007/978-1-4842-8844-3_2. https://doi.org/10.1007/978-1-4842-8844-3_2

[26] Valkov V. **Sentiment Analysis with BERT and Transformers by Hugging Face using PyTorch and Python. Curiousily** [Internet]. 2020 Apr 20 [cited 2022 Jul 4]. Available from: https://curiousily.com/posts/sentiment-analysis-with-bert-and-hugging-face-using-pytorch-and-python/

[27] Guan L. Weight **Prediction Boosts the Convergence of AdamW** [Internet]. arXiv; 2023 [cited 2026 Apr 24]. Available from: https://arxiv.org/abs/2302.00195. https://doi.org/10.48550/ARXIV.2302.00195

[28] Smith LN. **Cyclical Learning Rates for Training Neural Networks.** *In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV).* 2017. pp. 464–472.

[29] Zeng G. **Invariance Properties and Evaluation Metrics Derived from the Confusion Matrix in Multiclass Classification.** *Mathematics.* 2025;13(16):2609. https://doi.org/10.3390/math13162609

[30] Aguilar-Ruiz JS, Michalak M. **Multiclass Classification Performance Curve**. *IEEE Access.* vol.10, pp. 68915–68921, 2022. https://doi.org/10.1109/ACCESS.2022.3186444

[31] Alaros E, Marjani M, Shafiq DA, Asirvatham D. **Predicting Consumption Intention of Consumer Relationship Management Users Using Deep Learning Techniques: A Review**. *Indonesian J Sci Technol*. Vol. 8, no. 2, pp. 307–328, 2022. https://doi.org/10.17509/ijost.v8i2.55814

[32] Brahmi AE, Abderafi S, Ellaia R. **Artificial Neural Network Analysis of Sulfide Production in A Moroccan Sewerage Network**. *Indonesian J Sci Technol*. vol. 6, no.1, pp:193–204, 2021. https://doi.org/10.17509/ijost.v6i1.32322.