

# Enhancing the Takhrij Al-Hadith based on Contextual Similarity using BERT Embeddings

Emha Taufiq Luthfi<sup>1</sup>, Zeratul Izzah Mohd Yusoh<sup>2</sup>, Burhanuddin Mohd Aboobaidar<sup>3</sup>

Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia<sup>1</sup>

Faculty of Information, Communication and Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia<sup>2, 3</sup>

**Abstract**—Muslims are required to conduct Takhrij to validate the truth of Hadith text, especially when it is obtained from online media. Typically, the traditional Takhrij processes are conducted by experts and apply to Arabic Hadith text. This study introduces a contextual similarity model based on BERT Embedding to handle Takhrij on Indonesian Hadith Text. This study examines the effectiveness of BERT Fine-Tuning on the six pre-trained models to produce embedding models. The result shows that BERT Fine-Tuning improves the embedding model average accuracy by 47.67%, with a mean of 0.956845. The most high-grade accuracy was the BERT embedding built based on the indobenchmark/indobert-large-p2 pre-trained model on 1.00. In addition, the manual evaluation achieved 91.67% accuracy.

**Keywords**—Hadith text; Takhrij; natural language processing; text-similarity; word embedding; BERT fine-tuning

## I. INTRODUCTION

With the growth of information on the Internet, nowadays, most people, including Muslims, use online media as a primary source of information or knowledge. Most Muslims have adopted online media as a direct reference for exploring religious content, including those seeking verses of Qurán or Hadith. The problem is, not all information or knowledge on the Internet are verified for its correctness and authenticity [1].

The primary sources of Islamic law are the Qurán and Hadith [2][3]. Qurán is God's significantly trustworthy and unchanged Holy Book, which has been used as Islamic main reference for more than 14 eras since its revelation [4]. Hadith is an Islamic rule derived from the accumulation of the Prophet Muhammad PBUH's expressions, behaviors, judgments, or character [5][6]. Unlike the Holy Qurán, the Hadith distributed among Muslims are not all trustworthy [7]. Thus, Muslims need to authenticate the correctness of the Hadith, especially when it is from online media.

The approach in validating the correctness of the Hadith is referred as criticism of Hadith [8]. Three Hadith studies support Hadith's criticism: (1) Musthalah Hadith's study; (2) Study of Takhrij and Dirasah Sanad; and (3) Study of Thurud Fahmil Hadith. To recognize a Hadith's authenticity status, a Muslim must perform a Takhrij. Takhrij refers to the examination of the existence of the Hadith in the Hadith Books (its initial sources) such as Kutub al-Sittah, the Six Canonical Books of Hadith, Muwatta Imam Malik, and others. There is some prior research that describes and employs Takhrij al-Hadith to define the status of Hadith. According to [9], Takhrij is performed to meet several objectives as follows: (1) Origin

text (Masdar al-Hadith), (2) Authority of Hadith, (3) Narrators Bonds (Sanad), (4) Hadith Manuscript (Matn), (5) Hadith status in other references (specifying Shawahid or witnesses and Mutaba'at or follow-up), (6) Narrator's profile (al-Jarh wa Ta'dil), (7) Levels of Hadith according to the Sanad, (8) Levels of Hadith according to the Matn and (9) juristic ruling of Hadith (Hukm al-Hadith). [10][11] describes several methods of Takhrij based on the book written by Mahmud al-Thahhan entitled Usul al-Takhrij Dirasah wa al-Asânid:

- 1) Matn's first-word method.
- 2) The Word indexing method.
- 3) The Companion name index method.
- 4) The Hadith theme method.
- 5) The search method is based on Hadith status.
- 6) The search method is based on multiple Matn or Sanad conditions.
- 7) Digital searching via computer CDs or the Internet.

The Takhrij methods numbered 1-6 above are classical methods. In tune with digitalization growth, method number 7 has arisen [12].

Table I summarized the methods, and it can be formed that the Takhrij is a process for obtaining the original Hadith Text through numerous techniques. Table I also shows that traditional Takhrij requires human expertise in the process. Fig. 1 shows the illustration of the Takhrij process.

From the expert's point of view, Takhrij must be based on Arabic Hadith texts to avoid distortion on the Hadith translation. As such, performing Takhrij for other language, such as Indonesian, presents additional challenges.

For illustration, Table II shows two examples of Matn Hadith (highlighted in grey) that are different in sentence but have identical context. As we analyzed deeper; there are several different uses of the word that bring same meaning (bold in Table II), as manifested in Table III. These present a challenge in the Takhrij process to confirm the Hadith authenticity, where the translations are textually diverse though contextually identical.

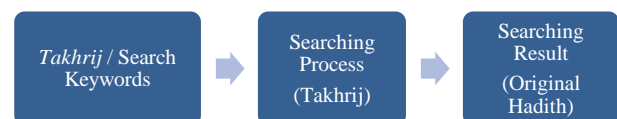


Fig. 1. Takhrij Process Illustration.

TABLE I. TRADITIONAL TAKHRIJ METHODS

<b>1. Hadith Matn's first-word method.</b>	
	<b>Takhrij References:</b> (1) <i>Al-Jami' Al-Shaghir Min Hadis Al-Basyir Al-Nadzir</i> , (2) <i>Faydh Al-Qadir Bi Syarh Al-Jami' Al-Shaghir</i> , (3) <i>Al-Fathu Al-Kabir Fi Dhammi Al-Ziya'dah Ila Al-Jami' Al-Shaghir</i> , (4) <i>Jam'u Al-Jawami' or Al-Jam'i Al-Kabir</i>
	<b>Strength:</b> The intended Hadith will likely be discovered right away.
	<b>Weakness:</b> An inaccuracy in the pronunciations of the first word employed becomes an obstacle to obtaining the Hadith.
<b>2. The Word indexing method.</b>	
	<b>Takhrij References:</b> <i>Al-Mu'jam Al-Mufahras Li Al-Faadz Al-Hadis An-Nabawy</i>
	<b>Strengths:</b> (1) Speed up Hadith exploration; (2) Limit searchable Hadith to specific master books by specifying the name of the book, juz, and pages; and (3) Allow Hadith to search through any words in the index.
	<b>Weakness:</b> (1) Must be able to speak Arabic and have sufficient scientific knowledge; (2) Does not mention the narrators from among the Companions; and (3) On occasion, a Hadith cannot be reached with a single word and must be found using other words.
<b>3. The Companion name index method.</b>	
	<b>Takhrij References:</b> <b>A. The Books of <i>Al-Athr'af</i>:</b> (1) <i>Tuhfatu Al-Asyraf Bi Ma'rifati Al-Athr'af</i> , (2) <i>Al-Nukat Al Zhiraaf 'Ala Al-Athraaf</i> <b>B. The Books of <i>Al-Musnad</i>:</b> (1) <i>Musnad Al-Imam Ahmad Bin Hanbal</i>
	<b>Strength:</b> (1) Shorten Takhrij by introducing Hadith Scholars who narrated their books; (2) Allow Takhrij on all Sanads.
	<b>Weakness:</b> Less effective without prior knowledge of the Hadith's narrators.
<b>4. The Hadith theme method.</b>	
	<b>Takhrij References:</b> (1) <i>Kanzu al-'Ummaal</i> , (2) <i>Miftah Kunuz al-Sunnah</i> , (3) <i>Al-Mughny 'An Hamli al-Asfar</i> , (4) <i>Nashbu al-Rayah</i> , (5) <i>Al-Dirayah</i> , (6) <i>Muntaqaa al-Akhbar</i> , (7) <i>Al-Durr al-Mansur</i> , etc
	<b>Strength:</b> Knowledge of the Hadith's content is required.
	<b>Weakness:</b> Occasionally, the theme of Hadith cannot be defined, or the theme defined by researchers and book compilers is different.
<b>5. The search method is based on Hadith status.</b>	
	<b>Takhrij References:</b> <b>A. Mutawatir Hadith:</b> (1) <i>Al-Azhar Al-Mutanatsirah fi Al-Akhbar Al-Mutawatirah</i> , <b>B. Qudsi Hadith:</b> (1) <i>Al-Ittihafat al-Saniyyah fi al-Ahadis al-Qudsiyyah</i> , <b>C. Popular Hadith:</b> (1) <i>Al-Maqasid al-Hasanah</i> , (2) <i>Kasyful Khofa' Wa Muzilul Ilbas</i> , <b>D. Mursal Hadith:</b> (1) <i>Al-Marasil</i> , <b>E. Maudhu' Hadith:</b> (1) <i>Tanzih asy-Syari'ah al-Marfu'ah 'an al-Ahadis asy-Syani'ah al-Maudhu'ah</i> , (2) <i>Al-Masnu fi Al-Hadis Maudhu</i>
	<b>Strength:</b> Probably would facilitate the Takhrij since most of the Hadith contained in a paper based on its characteristics are very few, so it does not require more complicated reasoning.
	<b>Weakness:</b> Due to the small number of hadiths included, the scope is minimal.
<b>6. The search method is based on multiple Matn or Sanad conditions.</b>	
	<b>Takhrij References:</b> <i>Rijalul Hadith Book</i>

TABLE II. AN EXAMPLE OF TAKHRIJ HADITH

The Hadith of an-Nawawi	Takhrij References
<i>Dari Amirul Mukminin Abu Hafsh, Umar bin Khaththab, ia berkata bahwa dirinya pernah mendengar Rasulullah bersabda, "Sesungguhnya, amal itu bergantung pada niatnya. Dan sesungguhnya seseorang hanya akan mendapatkan sesuatu sesuai dengan yang diniatkannya itu. Siapa yang hijrahnya karena Allah dan Rasul-Nya maka ia akan mendapat Allah dan Rasul-Nya. Dan siapa yang hijrahnya karena dunia yang ingin didapatkannya atau wanita yang ingin dinikahnya maka ia akan mendapatkan sesuai dengan yang ditujunya itu."</i>	<i>Telah menceritakan kepada kami [Al Humaidi Abdullah bin Az Zubair] dia berkata, Telah menceritakan kepada kami [Sufyan] yang berkata, bahwa Telah menceritakan kepada kami [Yahya bin Sa'id Al Anshari] berkata, telah mengabarkan kepada kami [Muhammad bin Ibrahim At Taimi], bahwa dia pernah mendengar [Alqamah bin Waqash Al Laitsi] berkata; saya pernah mendengar [Umar bin Al Khaththab] diatas mimbar berkata; saya mendengar Rasulullah shallallahu 'alaihi wasallam bersabda: "Semua perbuatan tergantung niatnya, dan (balasan) bagi tiap-tiap orang (tergantung) apa yang diniatkan; Barangsiapa niat hijrahnya karena dunia yang ingin digapainya atau karena seorang perempuan yang ingin dinikahnya, maka hijrahnya adalah kepada apa dia diniatkan" (Shahih Bukhari 1)</i>

TABLE III. AN EXAMPLE OF HOW WORDS ARE USED IN HADITH

The Hadith of an-Nawawi	Takhrij References
<i>amal</i>	<i>perbuatan</i>
<i>mendapatkan sesuatu</i>	<i>balasan</i>
<i>dan siapa</i>	<i>barangsiapa</i>
<i>didapatkannya</i>	<i>digapainya</i>
<i>ditujunya</i>	<i>diniatkan</i>

This paper employs semi-supervised BERT (Bidirectional Encoder Representations from Transformers) word embedding with a feed-forward neural network classifier to produce a Hadith text representation and determine its contextual similarity level. This work focuses on Hadith in the Indonesian language. The rest of the paper is structured as follows: Section II presents the previous work related to the contextual similarities. Section III about the theoretical definition of the text similarities, BERT, and the evaluation parameters used. Then Section IV describes the proposed model in this study. The results of this research examination are then discussed in Section V. Finally, Section VI explains the conclusion and directions for future research.

## II. RELATED WORK

Review on existing works shows that there are several work on word embedding techniques. Authors in [13] established a model that employs word2vec word embeddings, i.e., CBOW and Skrip-gram, innovated with SVM, for classifying the sentiment of social media tweets according to the context. This study shows that skip-gram 100-dimension achieved best classification performance with the values of precision, recall, f-score sequentially 64.4%, 58%, 61.1%.

Several other studies have applied BERT word embedding. The study by [14] utilized BERT sentence embedding for building automatic essay scoring. The outcome indicates that the BERT sentence embedding reaches an F1-score of 82.9%. Similarly, authors in [15] proposes a neural network with a pre-trained language model, M-BERT, that acts as an embedding layer to detect clickbait headlines. Evaluated with 5-fold cross-validation, it has an accuracy score of 91.4%, f1-score of 91.4%, a precision score of 91.6%, and ROC-AUC of 92%. Another study in [16] employs Latent Dirichlet Allocation and BERT embeddings to conduct topic modeling for graduate students' articles collected from the internet. The proposed model reached an average of 92.6% success rate for classifying the appropriate subject documents.

The study in [17] proposed INDOBERT, a novel Indonesian pre-trained language model, to evaluate and benchmark it across INDOLEM. The INDOLEM dataset covers several Indonesian language tasks, including word-forming and sentence-forming, word-meaning, and conversation. The results show that INDOBERT produces novel achievements in most of the tasks in INDOLEM. Authors in [18] introduces a novel model to recognize hate speech in Indonesian Twitter texts. The SVM and RFDI have been applied as machine learning approaches while BiGRU and pre-trained IndoBERT with BiGRU operate as deep learning strategies. The result shows that BiGRU and IndoBERT plus without stop word deletion obtained the most excellent accuracy of 84.77%.

Although several studies have successfully demonstrated the use of BERT word embedding, there is lack of studies on the contextual similarity between conducting Takhrij al-Hadith with Indonesian Hadith text.

## III. FUNDAMENTAL THEORY

### A. Text Similarities

Text similarity is an extensively applied method for obtaining relatedness between two texts [19]. A mechanism for representing text is required to measure the text-similarity of the natural language. Machines are incapable of understanding the notions of words. The usual technique for representing text is term vectors, in which terms or phrases are converted into vectors of real numbers [20].

Fig. 2 illustrates the variety of text representation forms. Work in [21] presents the weakness of traditional word embeddings, i.e., they carry no contextual representation of 'comparable' words. Furthermore, it raises a 'sparse' matrix problem on an extensive vocabulary. On the contrary, Word2Vec in Skip-Gram or CBOW can attract the semantic (contextual) representation. However, the 'sparse' matrix is still an obstacle to an extensive vocabulary.

### B. BERT

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained transformer form that can be fine-tuned with a single supplementary output layer. BERT fine-tuned the ability to generate various NLP tasks with new state-of-the-art outcomes for a broad range of assignments, including question answering, sentence classification, and sentence-pair regression, without significant task-particular architecture modification [22][23]. Fig. 3 shows the BERT architecture.

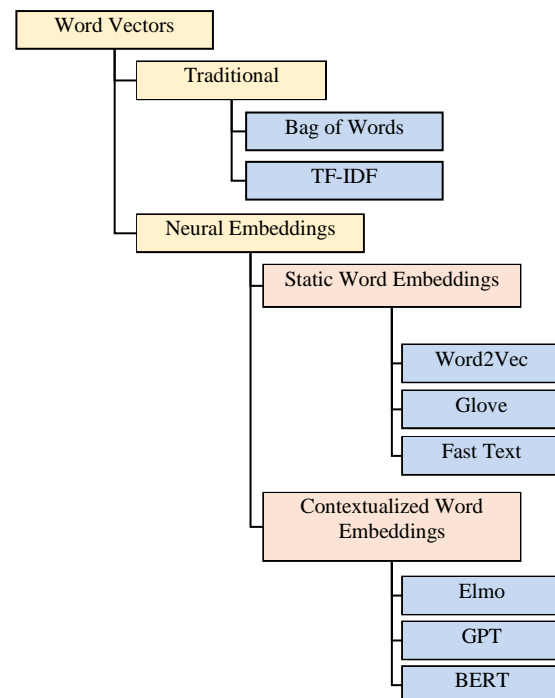


Fig. 2. Word Vectors Text Representation.

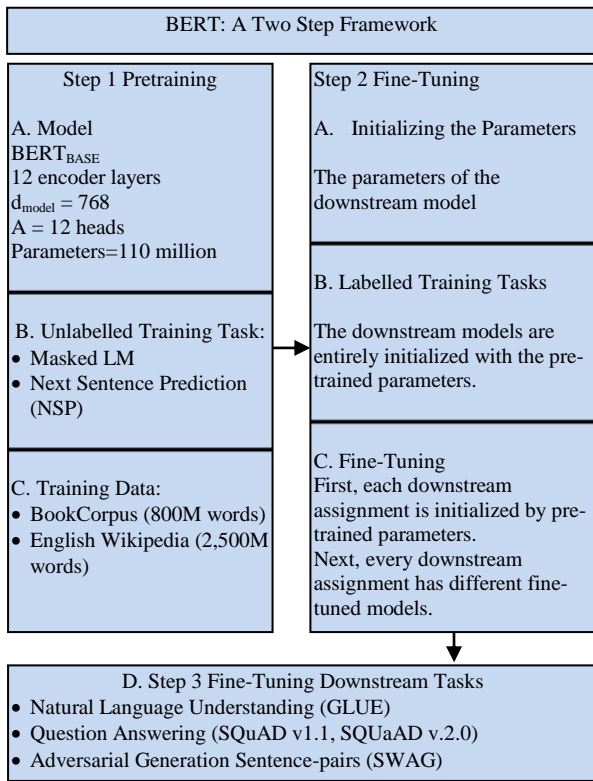


Fig. 3. BERT Architecture.

A feature-based or fine-tuning approach can assign pre-trained language representations for downstream assignments [22]. Fine-tuning is simple as the transformer self-attention mechanism provides BERT with the ability to perform various downstream assignments on a single text or text pair by exchanging suitable inputs and outputs. Every assignment only requires providing the assignment-particular inputs and outputs inside BERT also fine-tune entire parameters end-to-end.

### C. Evaluation Parameters

The evaluation of performances plays a crucial role in the development of classification models [24]. BERT Fine-Tuning employs Categorical Cross-Entropy Loss, also called Softmax Loss, as a loss function to calculate the distance within the current output of the algorithm and the expected output. It is a Softmax activation plus a Cross-Entropy loss, as shown in Fig. 4 [25].

Softmax output activation function:

$$f(s)_i = \frac{e^{s_i}}{\sum_j^c e^{s_j}} \quad (1)$$

Cross-Entropy Loss:

$$CE = -\sum_i^c t_i \log(f(s)_i) \quad (2)$$



Fig. 4. Categorical Cross-Entropy Loss Illustration.

Accuracy metrics operate to measure the number of accurate predictions to the total number of input specimens. The accuracy metric defined as [26]:

$$Acc = \frac{True\ Positive + True\ Negative}{Total\ Sample} \quad (3)$$

## IV. PROPOSED MODEL

### A. Methodology

The methodological approach practiced in this study is compiled in Fig. 5. The first step is gathering and preparing Hadith text as a dataset for training and testing the model. This study focuses on Hadith text that correlated to the Forty Hadith of al-Imam an-Nawawi. Ten unprocessed Hadith texts were gathered from [27] as Hadith texts to be tracked (Takhrij). Furthermore, fifty raw Hadith texts were gathered from <https://carihadis.com> as Takhrij references. The datasets are distributed as 70% training and 30% testing data. Table IV shows the specimens of raw Hadith text.

The second step is data pre-processing. Here are three sub-steps that are (a) text standardization, (b) eliminating undesirable items from Hadith text, and (c) data labeling. Fig. 6 shows the flow of the data pre-processing process in detail.

Table V shows the specimen of Hadith text resulting from pre-processing sub-step point a and point b.

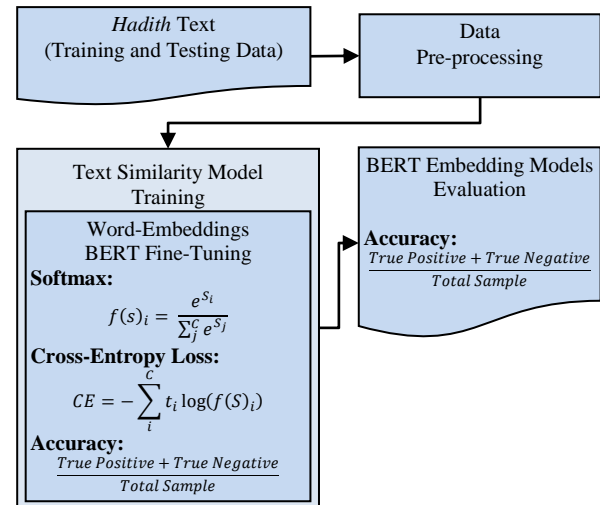


Fig. 5. Proposed Methodology.

TABLE IV. AN EXAMPLE OF RAW HADITH TEXT

The <i>Hadith</i> of <i>an-Nawawi</i>	<i>Takhrij</i> References
<p>Dari Amirul Mukminin Abu Hafsh, Umar bin Khaththab, ia berkata bahwa dirinya pernah mendengar Rasulullah bersabda, "Sesungguhnya, amal itu bergantung pada niatnya. Dan sesungguhnya seseorang hanya akan mendapatkan sesuatu sesuai dengan yang diniatkannya itu. Siapa yang hijrahnya karena Allah dan Rasul-Nya maka ia akan mendapat Allah dan Rasul-Nya. Dan siapa yang hijrahnya karena dunia yang ingin didupakannya atau wanita yang ingin dinikahinya maka ia akan mendapatkan sesuai dengan yang ditujunya itu."</p>	<p>Telah menceritakan kepada kami [Al Humaidi Abdullah bin Az Zubair] dia berkata, Telah menceritakan kepada kami [Sufyan] yang berkata, bahwa Telah menceritakan kepada kami [Yahya bin Sa'id Al Anshari] berkata, telah mengabarkan kepada kami [Muhammad bin Ibrahim At Taimi], bahwa dia pernah mendengar [Alqamah bin Waqash Al Laitsi] berkata; saya pernah mendengar [Umar bin Al Khaththab] diatas mimbar berkata; saya mendengar Rasulullah shallallahu 'alaihi wasallam bersabda: "Semua perbuatan tergantung niatnya, dan (balasan) bagi tiap-tiap orang (tergantung) apa yang diniatkan; Barangsiapa niat hijrahnya karena dunia yang ingin digapainya atau karena seorang perempuan yang ingin dinikahinya, maka hijrahnya adalah kepada apa dia diniatkan" (<i>Shahih Bukhari 1</i>)</p>
	<p>Telah menceritakan kepada kami [Abdullah bin Maslamah] berkata, telah mengabarkan kepada kami [Malik] dari [Yahya bin Sa'id] dari [Muhammad bin Ibrahim] dari [Alqamah bin Waqash] dari [Umar], bahwa Rasulullah shallallahu 'alaihi wasallam bersabda: "Semua perbuatan tergantung niatnya, dan (balasan) bagi tiap-tiap orang (tergantung) apa yang diniatkan; barangsiapa niat hijrahnya karena Allah dan Rasul-Nya, maka hijrahnya adalah kepada Allah dan Rasul-Nya. Barangsiapa niat hijrahnya karena dunia yang ingin digapainya atau karena seorang perempuan yang ingin dinikahinya, maka hijrahnya adalah kepada apa dia diniatkan.". (<i>Shahih Bukhari 53</i>)</p>
	<p>Telah menceritakan kepada kami [Muhammad bin Katsir], telah mengabarkan kepada kami [Sufyan], telah menceritakan kepadaku [Yahya bin Sa'id] dari [Muhammad bin Ibrahim At Taimi] dari [Alqamah bin Waqqash Al Laitsi], ia berkata; aku mendengar [Umar bin Al Khaththab] berkata; Rasulullah shallallahu 'alaihi wasallam bersabda: "Sesungguhnya amalan itu tergantung kepada niatnya, dan bagi setiap orang akan mendapatkan sesuai apa yang telah ia niatkan. Barangsiapa yang hijrahnya kepada Allah dan RasulNya, maka hijrahnya adalah kepada Allah dan RasulullahNya, dan barangsiapa yang hijrahnya untuk dunia yang hendak ia dapatkan atau karena seorang wanita yang akan ia nikahi, maka hijrahnya akan mendapatkan sesuai apa yang ia maksudkan." (<i>Sunan Abu Daud 1882</i>)</p>

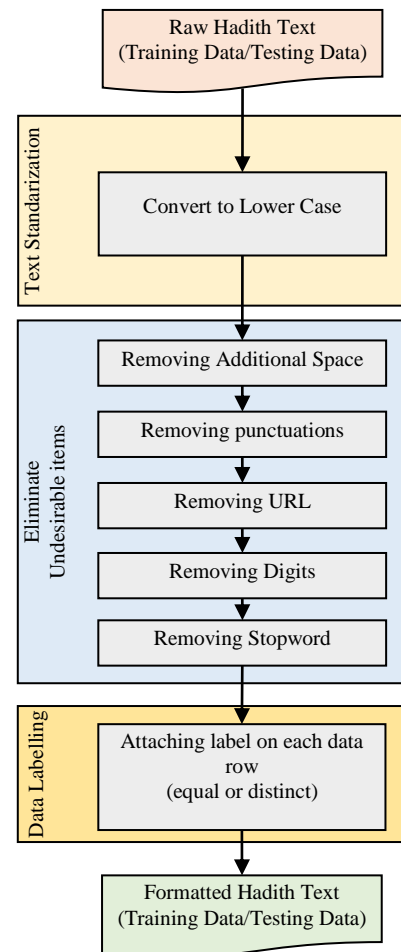


Fig. 6. Data Pre-Processing Flow.

TABLE V. EXAMPLE OF OUTPUT OF TEXT STANDARDIZATION AND ELIMINATE UNDESIRABLE ITEMS PROCESS

Status, Hadith1, Hadith2
<p>equal, dari ummul mukminin ummu abdillah aisyah ia berkata bahwa rasulullah bersabda siapa saja yang mengadaada dalam perkara agama kami ini sesuatu yang bukan bagian darinya maka ia tertolak dalam riwayat muslim yang lain disebutkan siapa saja yang mengerjakan suatu amalan yang tidak ada keterangannya dari kami maka ia ditolak, telah menceritakan kepada kami abu marwan muhammad bin utsman al utsmeni berkata telah menceritakan kepada kami ibrahim bin sad bin ibrahim bin abdurrahman bin auf dari bapaknya dari al qasim bin muhammad dari aisyah berkata rasulullah shallallahu alaihi wasallam bersabda barangsiapa membuat perkara baru dalam urusan kami yang tidak termasuk darinya maka dia tertolak</p>
<p>distinct, dari abu hurairah abdurrahman bin shakhr berkata bahwa ia pernah mendengar rasulullah bersabda apa yang kularang jauhilah dan apa yang kuperintah lakukanlah semampu kalian sesungguhnya yang membinasakan umatumat sebelum kalian adalah mereka banyak bertanya dan berselisih dengan nabi, telah menceritakan kepada kami abu marwan muhammad bin utsman al utsmeni berkata telah menceritakan kepada kami ibrahim bin sad bin ibrahim bin abdurrahman bin auf dari bapaknya dari al qasim bin muhammad dari aisyah berkata rasulullah shallallahu alaihi wasallam bersabda barangsiapa membuat perkara baru dalam urusan kami yang tidak termasuk darinya maka dia tertolak</p>

The third step is to build the BERT Embedding for contextual similarities model. The model is built with the exercises on top of some semi-supervised BERT models that employ BERT Fine-Tuning. A precise fine-tuning approach is needed to fit the BERT to NLP tasks in contextual similarities (Takhrij).

This study investigates the fine-tuning of five different BERT pre-trained models. The target model is a single label classification model and trained with several parameters, which are:

- Maximum length: None
- Batch Size: 32
- Epochs: 100
- Classification labels: Equal and Distinct
- Dense Layer: 2 Layers
- Dense Activation: SoftMax
- Loss function: Categorical Crossentropy
- Loss metrics: Accuracy

Fig. 7 illustrates the built model.

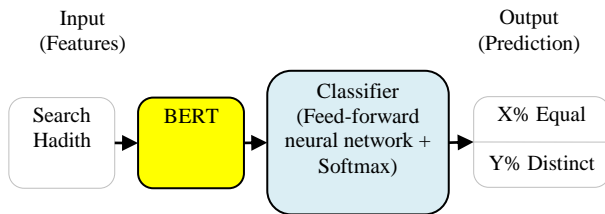


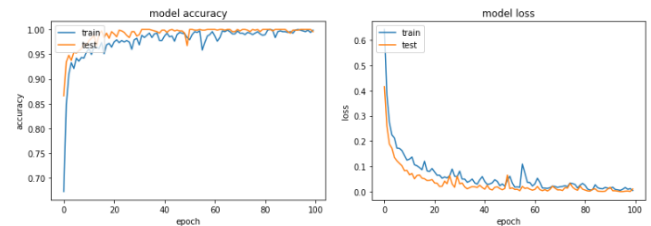
Fig. 7. BERT Model Illustration.

### V. RESULT

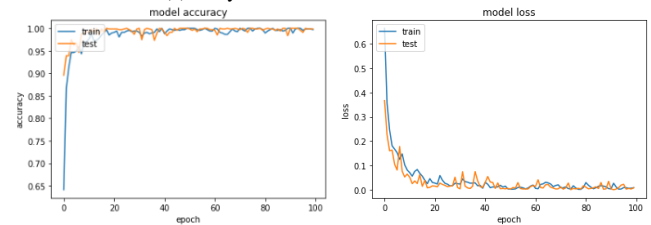
The BERT embedding model fine-tuning process is reported in Fig. 8, optimizing loss and accuracy values on each training and testing process iteration.

Table VI shows the accuracy result of the model built on the top of five different BERT pre-trained models. The BERT pre-trained model has choices that support the Indonesian language. The outcomes indicate that the BERT embedding model for contextual similarities gained average accuracy of 0.480060 and 0.956845 after fine-tuning, increasing by 0.4767886. The BERT pre-trained model indobenchmark/indobert-large-p2 achieved the best accuracy.

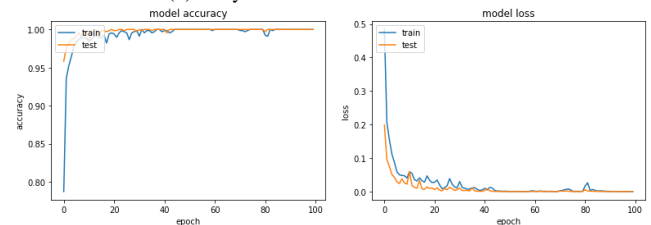
The fine-tune BERT embedding built on indobenchmark/indobert-large-p2 was then evaluated by manually comparing two texts of The Hadith of an-Nawawi with the Hadith text from the original books of Hadith. The result indicates an accuracy of 91.67%, as seen in Table VII.



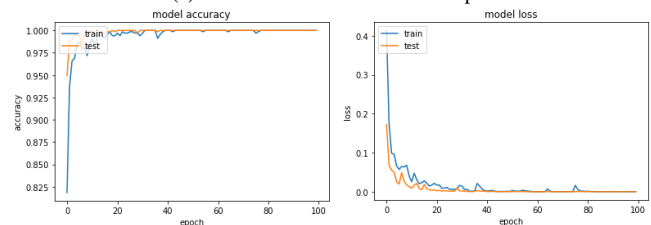
(a) Cahya/Bert-base-Indonesian-522M.



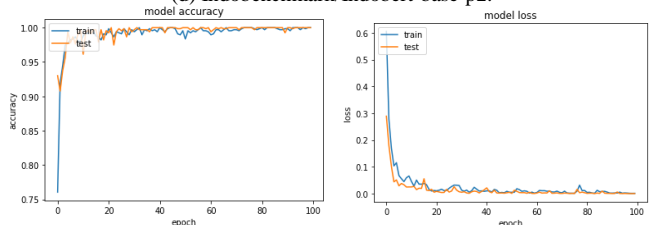
(b) Cahya/Bert-base-Indonesian-1.5G.



(c) Indobenchmark/Indobert-base-p1.



(d) Indobenchmark/Indobert-base-p2.



(e) Indobenchmark/Indobert-Large-p2.

Fig. 8. BERT Model Accuracy of Train and Test Process.

TABLE VI. MODEL ACCURACY

BERT pre-trained	Accuracy without Fine-Tuning	Accuracy with Fine-Tuning	Fine-Tuning Improvement
cahya/bert-base-indonesian-522M	0.626488	0.825893	0.199405
cahya/bert-base-indonesian-1.5G	0.630952	0.962798	0.331845
indobenchmark/indobert-base-p1	0.391369	0.997024	0.605655
indobenchmark/indobert-base-p2	0.373512	0.998512	0.625000
indobenchmark/indobert-large-p2	0.377976	1.000000	0.622024
Average	0.480060	0.956845	0.476786

TABLE VII. MANUAL EVALUATION

Takhrij Hadith Text	Sources Hadith Text	Results	Status
The Hadith of an-Nawawi 1	Shahih Bukhari 1	Equal 1.00	✓
	Shahih Bukhari 53	Equal 1.00	✓
	Shahih Muslim 3530	Equal 1.00	✓
	Sunan Abu Daud 1882	Equal 1.00	✓
	Sunan Ibnu Majah 4217	Equal 1.00	✓
	Sunan Nasai 74	Equal 1.00	✓
	Sunan Daruquthni 128	Equal 1.00	✓
	Shahih Ibnu Hibban 388	Equal 1.00	✓
	Shahih Muslim 10	Distinct 1.00	✓
	Shahih Muslim 11	Equal 0.58	✗
	Sunan Nasai 4904	Equal: 0.54	✗
	Sunan Nasai 4905	Distinct 0.99	✓
The Hadith of an-Nawawi 2	Shahih Bukhari 1	Distinct 1.00	✓
	Shahih Bukhari 53	Distinct 1.00	✓
	Shahih Muslim 3530	Distinct 1.00	✓
	Sunan Abu Daud 1882	Distinct 1.00	✓
	Sunan Ibnu Majah 4217	Distinct 1.00	✓
	Sunan Nasai 74	Distinct 1.00	✓
	Sunan Daruquthni 128	Distinct 0.99	✓
	Shahih Ibnu Hibban 388	Distinct 0.99	✓
	Shahih Muslim 10	Equal 0.91	✓
	Shahih Muslim 11	Equal 0.97	✓
	Sunan Nasai 4904	Equal 0.95	✓
	Sunan Nasai 4905	Equal 0.93	✓

VI. CONCLUSION AND FUTURE WORK

In this paper, a semi-supervised BERT word embedding with a feed-forward neural network classifier is proposed and implemented to produce a Hadith text representation and determine its contextual similarity level. This work focuses mainly for Hadith in the Indonesian text. The BERT fine-tuning raised average accuracy by 47.67%, with a 0.956845 mean accuracy in the training process. The pre-trained model Indobenchmark/indobert-large-p2 achieved the highest accuracy with training 1.00. The final manual evaluation achieved 91.67% accuracy on the Hadith contextual similarity identification. It means the proposed model in this study reaches the highest performance when used to conduct Hadith Takhrij (searching) for Indonesia Hadith text. As a future development of this experiment, there are some directions to be studied. The first direction is to extend the number of Hadith texts as an experiment dataset. That is needed because Hadith text is known to have some different Sanad and Matn structures. Another direction is in the order of automatic recognition of parts of the Hadith text. Identify its Sanad or Matn and then classify the Hadith text based on its structure.

REFERENCES

- [1] R. Baru, S. Hadzrullathfi, S. Omar, and B. Ibrahim, "Identifying False Hadith Guidelines," Malaysian J. Islam. Stud., vol. 1, pp. 62–73, 2017.
- [2] S. R. Mohammad Najib, N. Abd Rahman, N. Kamal Ismail, N. Alias, Z. Mohamed Nor, and M. N. Alias, "Comparative Study of Machine Learning Approach on Malay Translated Hadith Text Classification based on Sanad," MATEC Web Conf., vol. 135, p. 00066, 2017.
- [3] N. A. P. Rostam and N. H. A. H. Malim, "Text categorisation in Quran and Hadith: Overcoming the interrelation challenges using machine learning and term weighting," J. King Saud Univ. - Comput. Inf. Sci., Mar. 2019.
- [4] Muahammad Khurram Khan and M.Alginahi Yaser, "The Holy Quran Digitization: Challenges and Concerns," Life Sci. J., vol. 10, no. 2, pp. 156–164, 2013.
- [5] Ayub, "Matn Criticism and Its Role in The Evaluation of Hadith Authenticity," (International J. Islam. Stud. Humanit., no. March, pp. 1–4, 2018.
- [6] A. Mahmood, H. Ullah, F. K., M. Ramzan, and M. Ilyas, "A Multilingual Datasets Repository of the Hadith Content," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 2, 2018.
- [7] E. T. Luthfi, N. Suryana, and A. S. H. Basari, "Digital Hadith Authentication: a Literature Review and Analysis," J. Theor. Appl. Inf. Technol., vol. 96, no. 15, 2018.
- [8] T. Rohman and U. Huda, "Methodology of Hadith Research : The Study of Hadith Criticism Metode Penelitian Hadis: Studi tentang Kritik Hadis," vol. 2, no. 1, pp. 73–84, 2019.
- [9] I. Suliaman et al., "The convenience of takhrij al-hadith through ICT apps: An exploratory analysis on selected hadith website and mobile apps," Int. J. Civ. Eng. Technol., vol. 9, no. 11, pp. 2649–2660, 2018.
- [10] A. Rahman, "Pengenalan Atas Takhrij Hadis," Riwayat J. Stud. Hadis, vol. 2, no. 1, p. 146, 2017.
- [11] A. Izzan, Studi Takhrij Hadis: Kajian Tentang Metodologi Takhrij dan Kegiatan Penelitian Hadis, Pertama. Bandung: Tafakur, 2012.
- [12] Istianah and S. Wahyuningsih, "The hadith digitization in millennial era: A study at center for hadith studies, Indonesia," Qudus Int. J. Islam. Stud., vol. 7, no. 1, pp. 25–44, 2019.
- [13] F. W. Kurniawan and W. Maharani, "Indonesian Twitter Sentiment Analysis Using Word2Vec," 2020 Int. Conf. Data Sci. Its Appl. ICoDSA 2020, pp. 9–14, 2020.
- [14] R. A. Rajagede, "Improving Automatic Essay Scoring for Indonesian Language using Simpler Model and Richer Feature," Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control, vol. 4, pp. 11–18, 2021.
- [15] M. N. Fakhruzzaman, S. Z. Jannah, R. A. Ningrum, and I. Fahmiyah, "Clickbait Headline Detection in Indonesian News Sites using Multilingual Bidirectional Encoder Representations from Transformers (M-BERT)," 2021.
- [16] S. M. Ozdemirci and M. Turan, "Case Study on well-known Topic Modeling Methods for Document Classification," Proc. 6th Int. Conf. Inven. Comput. Technol. ICICT 2021, pp. 1304–1309, 2021.
- [17] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," pp. 757–770, 2021.
- [18] A. Marpaung, R. Rismala, and H. Nurrahmi, "Hate Speech Detection in Indonesian Twitter Texts using Bidirectional Gated Recurrent Unit," KST 2021 - 2021 13th Int. Conf. Knowl. Smart Technol., pp. 186–190, 2021.
- [19] S. Viji, D. Tayal, and A. Jain, "A Machine Learning Approach for Automated Evaluation of Short Answers Using Text Similarity Based on WordNet Graphs," Wirel. Pers. Commun., no. 0123456789, 2019.
- [20] A. Bornstein, "Beyond Word Embeddings Part 2." [Online]. Available: <https://towardsdatascience.com/beyond-word-embeddings-part-2-word-vectors-nlp-modeling-from-bow-to-bert-4ebd4711d0ec>. [Accessed: 19-May-2021].

- [21] R. K. Gupta, "Journey to BERT : Part 1," medium.com, 2020. [Online]. Available: <https://medium.com/swlh/journey-to-bert-part-1-a89413855a10>. [Accessed: 08-Aug-2021].
- [22] M. C. Kenton, L. Kristina, and J. Devlin, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," no. Mlm, 1953.
- [23] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf., pp. 3982–3992, 2020.
- [24] Y. Liu, Y. Zhou, S. Wen, and C. Tang, "A Strategy on Selecting Performance Metrics for Classifier Evaluation," *Int. J. Mob. Comput. Multimed. Commun.*, vol. 6, no. 4, pp. 20–35, 2014.
- [25] R. Gómez Bruballa, "Understanding Categorical Cross-Entropy Loss, Binary Cross-Entropy Loss, Softmax Loss, Logistic Loss, Focal Loss and all those confusing names," 2018. [Online]. Available: [https://gomburu.github.io/2018/05/23/cross\\_entropy\\_loss/](https://gomburu.github.io/2018/05/23/cross_entropy_loss/). [Accessed: 29-Aug-2021].
- [26] Ž. Vujović, "Classification Model Evaluation Metrics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 599–606, 2021.
- [27] M. D. Al-Bugha and M. Mistu, *Al-Wafi: Syarah Hadis Arba'in Imam An-Nawawi (Edisi Indonesia)*. Jakarta: Qisthi Press, 2014.