**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 3, December 2025 Page: 2017−2024
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8861

# Collaboration between Convolutional Neural Network and Semantic Search for English Hadith Search Using Automatic Topic Classification, TF-IDF, and Sentence-BERT

**Akmal Sidki Razaka, Kemas Muslim Lhaksmana***

School of Computing, Telkom University, Bandung, Indonesia
Email: [1]akmalsidki@student.telkomuniversity.ac.id, [2,*]kemasmuslim@telkomuniversity.ac.id
Correspondence Author Email: kemasmuslim@telkomuniversity.ac.id
Submitted: **06/12/2025**; Accepted: **26/12/2025**; Published**: 26/12/2025**

**Abstract**−This research was conducted with the intention of developing an English-language hadith search system that is not only syntactically accurate, but also contextually appropriate. The system was developed using a combination of convolutional neural networks (CNN) and two text representation methods, namely Term Frequency–Inverse Document Frequency (TF-IDF) and Sentence-BERT (SBERT). CNN is used to classify hadiths into seven main categories based on chapter titles. In the semantic retrieval stage, TF-IDF and SBERT were utilized to represent the text of the hadith and user queries, then both were evaluated using cosine similarity. Testing was conducted using five queries commonly used in Islamic studies, then evaluated manually for semantic similarity. As a result, the tuned CNN achieved a classification accuracy of 94%. On the other hand, although the TF-IDF approach produced greater similarity results, SBERT proved to be superior in generating more relevant results in semantic searches. These results indicate that TF-IDF is superior in terms of speed, but SBERT is better at understanding sentence context in depth. This research contributes to the development of a meaning-based hadith search system and emphasizes the importance of a semantic approach in religious text search. Moving forward, system development can be directed toward multilingual support and evaluation on a larger scale.

**Keywords**: Hadith Retrieval; Convolutional Neural Network; TF-IDF; Sentence-BERT; Semantic Search

## 1. INTRODUCTION

With the advancement and development of digital technology today, access to a wealth of information and data, including access to hadith, is now readily available. Along with modern technology and the internet, Muslims are faced with two problems. First, the validity of the sources used in the compilation of books or rules and access to books or rules of hadith [1]. Although efforts have been made to provide an understanding of hadith, a literal or textual approach will not provide the deep, contextual, and multi-layered meaning required in hadith. Therefore, pure semantics must play a crucial role because the Arabic language, the language of hadith, has many possible meanings, especially in the context of hadith and its relationship with the Quran and other hadith [2]. Without alternative approaches, some fundamental relevance will be lost in hadith searches. Minor errors must also be minimized, such as user errors. Additionally, to address other challenges and difficulties, subtopics and hadith differentiation must be considered as incentives, prohibitions, or purely informational statements [3]. To address these issues, this study proposes the integration of the Convolutional Neural Network (CNN) method for automatic topic classification, Term Frequency-Inverse Document Frequency (TF-IDF) for initial feature representation, and Sentence-BERT for semantic matching in an English-translated hadiths search system.

TF-IDF is utilized for text representation because this method is considered more effective. This method measures weight based on its occurrence in the document and compares it with the entire corpus [3]. Meanwhile, Convolutional Neural Networks (CNNs) have been widely used to identify text because their convolutional layers are capable of identifying local features. This occurs because convolutional layers in CNNs are able to effectively classify specific patterns in sequential text data. This results in better accuracy in classification results [4]. On the other hand, Sentence-BERT, which is an extension of BERT, uses Siamese and Triplet networks to generate fixed sentence vectors. Its strength lies in understanding the importance of sentences with higher accuracy and efficiency. These vectors will achieve the best results in tasks related to semantic similarity, clustering, and semantic search. [5]. The integration of these three methods has great potential in building a hadith search system that is syntactically correct and semantically relevant. Therefore, a system must be developed to collaborate statistical methods such as TF-IDF, the topic modeling capabilities of CNNs, and the semantic richness provided by Sentence-BERT to produce more contextual, accurate, and responsive search results that take into account the variability of hadith content.

There are many studies on machine learning that focus on classifying hadith texts. Ramadhani et al. (2025) research aims to perform multi-label classification on translations of Shahih al-Bukhari. They utilize TF-IDF and Support Vector Machine (SVM) as statistical approaches and margin-based or supervised learning approaches. One of the labels received an F-1 Score of 82.57%, indicating that data-based representation is effective in identifying text patterns [6]. Meanwhile, Darmalaksana et al. (2020) developed a hadith search engine using LSA and cosine similarity. This system was able to achieve a recall of 87.83%. The strong recall performance indicates that the system can effectively capture hadiths whose meanings are similar to the user's query [7]. Akbar et al. (2024) created a multi-label classification model for Indonesian translation of the Qur'an, classifying verses into four main topics: Tawhid, Worship, Morals, and History. The model uses a Bi-Directional Long Short-Term Memory (Bi-LSTM) algorithm combined with CBOW-based Word2Vec word embedding. This model produced an accuracy of 70.21%, precision of

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 3, December 2025 Page: 2017−2024
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8861

64.31%, and recall of 61.13%, demonstrating the suitability of this method in capturing the context of Islamic texts [8]. Luthfi et al. (2021) developed a hadith search system for Takhrij needs based on semantic similarity, using fine-tuned BERT to map the meaning of user hadiths to references such as Sahih Bukhari & Muslim. The evaluation showed an accuracy of up to 91.67% in determining contextual similarities between texts, demonstrating the effectiveness of the embedding-based approach. However, this study did not include topic classification in the system design [9]. Abu Bakar and Adiwijaya (2021) developed a classification system for Indonesian-translated hadiths. Grouping them into recommendations, prohibitions, and information using a Recurrent Convolutional Neural Network (CRNN) architecture, which combines CNN for feature extraction and RNN. This model achieved an accuracy of up to 80.79%, demonstrating better performance than other comparative techniques. However, this model has not yet integrated a semantic search system that considers the meaning between sentences [10]. Overall, these studies show that to create an efficient hadith search system, it is important to choose appropriate algorithms such as CNN for topic classification, apply effective text representation methods such as TF-IDF and Sentence-BERT, and recognize user query characteristics in order to achieve accurate, relevant, and contextually appropriate search results.

Based on previous studies, there has been no study that explicitly compares TF-IDF and SBERT for searching English-translated hadiths. Previous studies only relied on one method and none evaluated them head to head. On the other hand, the selection of text representation has a strong influence on search accuracy, especially for Islamic texts that have very complex linguistic nuances.

This study aims to create a search system for English-translated hadiths by comparing two text representations, namely TF-IDF and SBERT. Both text representations are entered into a consine similarity search to evaluate their ability to find hadiths with matching semantic meanings.

# 2. RESEARCH METHODOLOGY

## 2.1 Research Stages

This study combines CNN, TF-IDF, and SBERT, which are tested using an experimental quantitative approach for semantic search for English-translated hadiths. This study aims to develop a system that can classify topics and semantic searches that contextually aligned with user queries. The research stages are summarized in Figure 1 below:

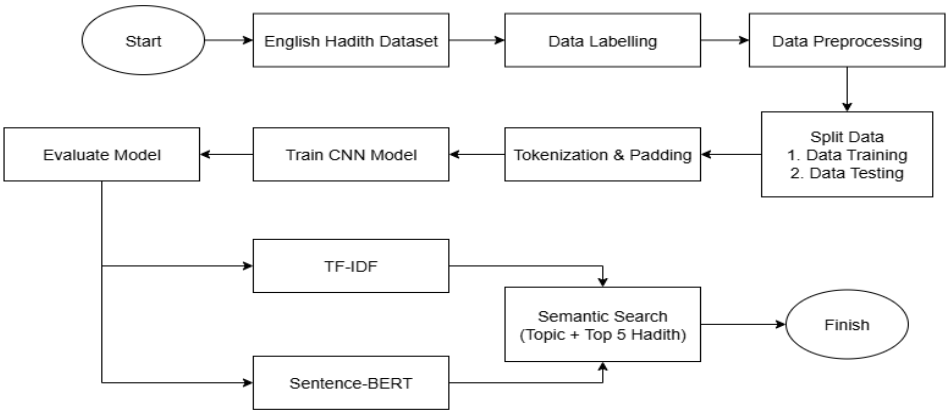


**Figure 1.** Research Stages

Figure 1 illustrates the workflow of the system. The process starts with an English Hadith dataset, followed by data labeling and preprocessing. The data is then split into training and testing sets, tokenized, padded, and used to train a CNN model, which is subsequently evaluated. In parallel, TF-IDF and Sentence-BERT are applied to support semantic search, producing topic-based retrieval with the top five relevant hadiths as the final output.

## 2.2 English Hadist Dataset

This research utilizes materials provided by sunnah.com, an open platform that provides hadith in English from various major hadith books. This dataset consists of 42,694 hadiths, containing important information such as the name of the book, chapter title, hadith text, and hadith quality. To maintain reliability, the hadiths used are the most credible ones, namely hadiths with the predicate sahih narrated by Bukhari and Muslim. These two books are generally considered the most authentic hadith books in the field of hadith science because they apply strict criteria in selecting sanad and matan, and have been recognized by the majority of scholars regarding the authenticity of their contents [11] [12]. After the filtering process, the amount of data for training and testing the system became 12,266 hadiths, which were then used in the topic labeling and text processing processes.

## 2.3 Data Labeling

The labeling in this study is based on the EnglishBabName column, which represents the thematic titles of the hadith chapters. However, these titles have a wide variety of topics and numbers, making them inefficient to use directly as

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 3, December 2025 Page: 2017−2024
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8861

labels. Therefore, a topic mapping process was conducted using a semantic approach focused on keywords. The approach included extracting keywords from chapter headings that deal with prayer, fasting, and fellowship that capture the main themes from a given hadith. Then, the researchers systematically grouped the titles of these chapters into seven main groups according to their thematic content. Each chapter is categorized into Worship, Biography & History, Knowledge & Da'wah, Family, Faith & Morals, Muamalah, and Law & Governance. This categorization is based on a combination of knowledge about Islam and keyword annotation. These annotations serve to provide a semantic structure suitable for search systems based on thematic context. They also serve to group chapters from the dataset for more accurate classification by CNN.

### 2.4 Data Preprocessing

Preprocessing is carried out with the aim of preparing hadith content so that it can be effectively utilized by semantic classification and search processes. Data preprocessing plays a crucially important role in natural language processing because it ensures that the data remains consistent and coherent before being used to train machine learning models. The accuracy of classification results is largely dependent on the quality of preprocessed data, since errors that occur at this stage can directly impact the model's performance in analyzing text [13]. This approach is designed to eliminate information that is unnecessary for classification and to remove formatting codes and symbols while retaining the meaning of each hadith.

The first step in data preprocessing is to standardize column names or column headers. Then, the cleansing process is carried out with the aim of removing parts that do not have significant meaning in the hadith text. This process is done by removing punctuation marks, numbers, and special characters. Then, a case-folding method is applied that changes all characters to a lowercase state so that it remains uniform across the text. In the next stage, stopwords are ignored by removing highly occurring words that don't hold inherent meaning, like "and," "the," and "is" as defined by the English corpus given by NLTK. Next, tokenization is performed by breaking sentences into words using the word tokenize function [14]. All results from this process are stored in two new columns: hadith_display, which contains the clean text for display, and preprocess_text, which is used as input for the model. Thus, the hadith text is now in a standardized format and ready for topic classification and semantic search processes. Table 1 outlines the preprocessing steps of this study.

**Table 1.** Preprocessing Stages

| Stages | Result |
|---|---|
| Original Text | Narrated 'Umar bin Al-Khattab: <p> I heard Allah's Apostle ... his emigration was for what he |
| Cleansing | narrated umar bin alkhattab i heard allahs apostle saying the reward of deeds ... |
| Case Folding | narrated umar bin alkhattab i heard allahs apostle ... |
| Stopword Removal | narrated umar bin alkhattab heard allahs apostle saying reward deeds depends intentions ... |
| Tokenization | ['narrated', 'umar', 'bin', 'alkhattab', 'heard', 'allahs', 'apostle', 'saying', ...] |

### 2.5 Convolutional Neural Network

Convolutional Neural Network (CNN) is used in this study because of its ability to recognize local patterns in sequential data such as hadith texts. CNN utilizes convolution operations to capture important features from word sequences and has proven effective in text classification with high accuracy [15]. CNN is capable of processing one-dimensional data such as text by relying on convolutional layers, pooling layers, and dense layers that work together to extract and classify features [16].

In this study, two different Convolutional Neural Network (CNN) architectures were used to classify hadith topics. The first model is a baseline model used as a basic design. This model uses a single convolutional path consisting of embedding, pooling, dropout, and two dense layers. This configuration reflects the general form of a one-dimensional CNN for text classification.

Unlike the first model, the second model uses a more diverse architecture. The second model is designed to capture word patterns with a wider range of n-gram variations. This model uses hyperparameter tuning techniques and several convolutional filters running simultaneously to enrich the CNN structure. Embedding layers are connected to three Conv1D layers with varying kernel sizes, and the results are combined before proceeding to the pooling stage and dense layer.

Both architectures were evaluated on the task of classifying hadith topics. The classification output was then combined with TF-IDF and SBERT in the hadith search stage. The results were then compared to see which approach was the most effective for semantic representation and search in the context of searching for English-language hadiths.

### 2.6 Evaluate Model

The model evaluation aims to assess the CNN's ability to classify hadith topics. The metrics used to evaluate the model are accuracy, which shows how many predictions are correct compared to incorrect ones, then there are precision and

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 3, December 2025 Page: 2017−2024
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8861

recall, which indicate how accurate and complete the classification is for each topic, while F1-Score is a combined metric that balances precision and recall.

These evaluation results were generated using Keras Tokenizer for accuracy measurement and SCIKIT-learn to obtain a detailed class matrix. This evaluation approach is widely used in classification research because it shows model performance not only overall but also for each class, especially when the data distribution is uneven [17].

### 2.7 TF-IDF

One of the text representation methods used in this study is Term Frequency–Inverse Document Frequency (TF-IDF). TF-IDF is a statistical approach that assigns weights to words in a document based on their frequency in that document and how rarely they appear in the entire corpus. The more frequently a word appears in a document and the rarer its occurrence in other documents, the higher its weight. Thus, TF-IDF helps highlight words that are considered important and relevant for distinguishing between documents [18].

In its application, this study uses the TfidfVectorizer library in scikit-learn with the parameter max_features = 5000 to limit the number of unique words, and stop_words = 'english' to remove insignificant everyday words. The representation of this transformation is applied at the semantic search stage, which is the search for hadiths based on the similarity of meaning between the user's query and the hadith collection. In this context, TF-IDF representation is used as a substitute for keyword-based retrieval before being contrasted with meaning-based retrieval using Sentence-BERT.

### 2.8 Sentence-BERT

Sentence-BERT (SBERT) is an extension of the standard BERT architecture designed specifically to generate semantic and meaningful sentence representations (sentence embeddings). Unlike conventional BERT approaches that use the [CLS] token for classification, SBERT leverages the Siamese Network structure and applies pooling strategies (such as mean pooling) to contextually merge word representations into efficient sentence representations. This approach makes SBERT highly effective for tasks requiring measurement of semantic similarity between sentences, such as semantic search and text matching [19].

In this study, SBERT is used as a meaning-based text representation method to support the semantic search process for hadiths. The all-MiniLM-L6-v2 model created by Reimers & Gurevych was selected due to its ability to make quick inference while also being very low in computational cost, while facing minimal trade-off in embedding quality. Using SBERT allowed for cleanedText column extracted hadiths to be converted into fixed-dimensional vector forms. This is made it possible to calculate consine similarity for input from user with the aim of finding the semantically most relevant hadiths

### 2.9 Semantic Search

Semantic search is an information retrieval technique that aims to go beyond simple keyword matching by focusing on the integration of term semantics along with their relationships. This technique is based on semantic representations for textual data with an exploration of relationships between concepts with a focus on improving the relevance of retrieved data. Keyword-based contextual search has limitations in the field of information management and digital libraries. Semantic search is seen as the solution because it is able to recognize the semantic relationships between words that appear different on the surface [20] [21].

In this study, this approach was used to improve the English-language hadith search system so that the results would be more semantically relevant. Two text representation methods were applied to form semantic vectors, namely Term Frequency-Inverse Document Frequency (TF-IDF), which focuses on word frequency, and Sentence-BERT (SBERT), which focuses on capturing sentence meaning. The effectiveness of these two techniques was then tested in the process of using cosine similarity to measure the accuracy between user queries and hadith texts. This approach allows for retrieving hadiths semantically related to a given question despite differences in terminology.

## 3. RESULT AND DISCUSSION

### 3.1 Dataset Description

This study uses an English Hadith dataset obtained from sunnah.com. Although the general characteristics of the dataset have been discussed in Chapter 2, this section focuses on the dataset as applied in the experimental workflow. From the original 42,694 hadiths, only sahih hadiths narrated by Bukhari and Muslim were selected to ensure data reliability. After filtering, a total of 12,266 hadiths were used in this study. These data were then subjected to topic labeling, text preprocessing, model training, and semantic search experiments as illustrated in Figure 1.

### 3.2 Results of Hadith Topic Classification Using CNN

This study employs a Convolutional Neural Network (CNN) architecture to classify a topic obtained from a set of hadiths into seven unique categories. The architecture that is relatively straightforward and has been embraced for thisstudy consistsof an embedding layer, followed by a Conv1D layer, then a GlobalMaxPooling1D layer, in addition

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 3, December 2025 Page: 2017−2024
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8861

to two dense layers linked by a dropout layer. Training of the network lasts for a duration of six epochs, with an early stopping strategy being employed to eliminate any possible risk of overfitting.

The training results showed a remarkable training accuracy improvement from 39.6% to 97.8%, while the best validation score was achieved for validation accuracy of 92.18%. Furthermore, the loss value for the validation data notably fell from 0.67 to 0.38. Evaluation of the test data showed an accuracy of 91.44%. Based on the classification report, the baseline model recorded the highest precision and recall values in class 0 with an F1-score of 0.95, while the lowest score was in class 6 with an F1-score of 0.79. The average macro F1-score achieved was 0.89, and the weighted F1-score was 0.92. Figure 2 shows the training and validation accuracy and loss of the baseline CNN model. The results indicate that the model performs classification reasonably well, despite the class imbalance in the minority class.

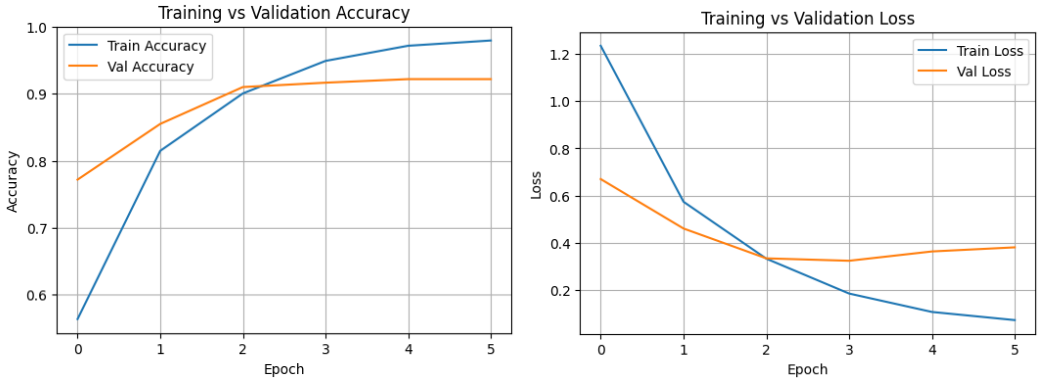


**Figure 2.** CNN Baseline Result

Figure 2 shows the accuracy and loss graphs of the baseline model training process. The left graph shows a significant increase in training and validation accuracy until it stabilizes above 90%. The right graph shows a decrease in loss on training and validation data. Although there was a slight increase in validation loss after the third epoch, the overall training trend shows that the model is able to learn well and only experiences slight overfitting.

In addition, the CNN architecture was improved by introducing a multi-kernel layout into the convolutional layer to capture broader features. The model uses Early Stopping as well as the Model Checkpointing to prevent overfitting, and to keep the best model in place through training. Therefore, the improved model does veify a better performance on the validation set than the base model, which is 95%. While the weighted F1-score equalled 0.94, showing that our model was consistent in performance in all classes. The improved model achieved a slightly higher validation accuracy than the base model, being in 95%. The weighted averaged F1-score was similar with a value of 0.94, concluding that the performance was quite reliable for each class.

Modifications of CNN architecture usually results in higher testing accuracy, the good model performance and slightly reduction of the overfitting. This is illustrated by a more linear training graph and lower validation loss compared to the standard model.

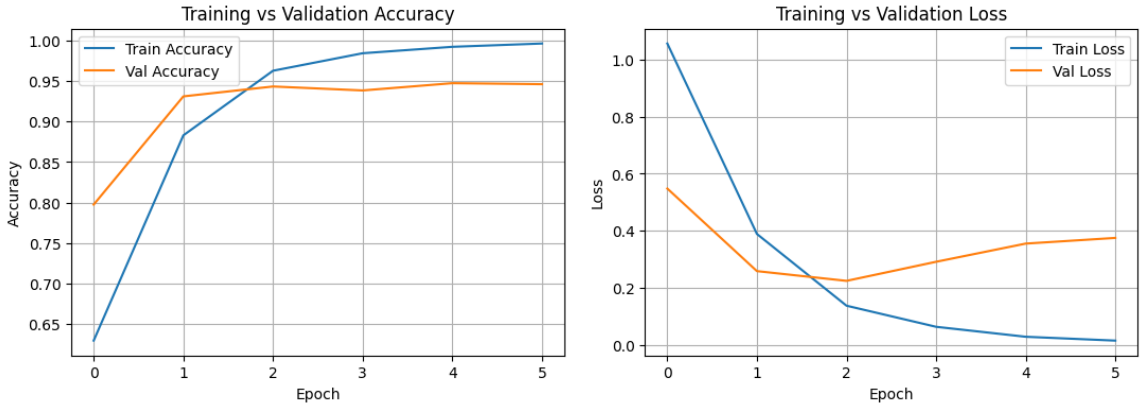


**Figure 3.** CNN Tuned Result

The training graph in Figure 3 indicates a more consistent trend in accuracy and loss when compared to the earlier model. The left graph shows that training accuracy increases rapidly and approaches 99% at the end of training, while validation accuracy remains high and stable after the second epoch. In the right graph, it can be seen that training loss decreases dramatically, while validation loss is also low and stable despite slight fluctuations. The gap between training and validation loss is smaller than the baseline model, indicating better model generalization and lower risk of overfitting. A complete comparison between the two models is shown in Table 2 below:

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 3, December 2025 Page: 2017−2024
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8861

**Table 2.** Result of CNN baseline and CNN tuned

| Model | Accuracy | Precision | Recall | Macro F1-Score | Weighted F1-Score |
|---|---|---|---|---|---|
| CNN (Baseline) | 0.92 | 0.88 | 0.90 | 0.89 | 0.92 |
| CNN (Tuned) | 0.94 | 0.92 | 0.94 | 0.94 | 0.92 |

Table 2 shows a comparison of the performance of the two models. The table shows that the second model performs slightly better than the first model. Although the difference is not significant, the model that has undergone hyperparameter tuning still shows better results. This occurs because the model receives structural adjustments, such as through the use of multi-kernel convolutions. These architectural adjustments help model capture more complex patterns or features. This result is evident from the model's improved performance when tested on new data.

### 3.2.1 CNN Performance Analysis

The tuned CNN model shows more optimal results than the baseline model. The accuracy of the tuned model reached 94%, while the baseline achieved 92%. In terms of recall and precision, the tuned model achieved 92% and 94% respectively, while the untuned model only reached 88% and 90%. This proves that the tuning process has better pattern recognition capabilities.

The tuned CNN model has a major advantage due to its ability to ensure training stability. This is shown in Figure 3, where the training loss and validation loss are close or do not fluctuate. This means that the model prevents overfitting and is able to learn efficiently due to the application of additional techniques such as multi-kernel convolutional layers, early stopping, and checkpoints. However, the baseline model shows an imbalance between training loss and validation loss, indicating a tendency for overfitting even though the model's accuracy is high.

However, the baseline model is lighter and faster to train due to its less complex architecture. However, the baseline model is less capable of capturing complex patterns due to its simple architecture. This is demonstrated by a lower macro F1-score of 89% compared to the tuned model, which achieved 94%. This means that the tuned model is suitable for this dataset.

In general, the tuned model provides better results. This is demonstrated by better accuracy, balanced classification between classes, and stable training performance. These findings show that the choice of CNN architecture and the tuning process have a significant impact on the model's ability to classify hadith topics. Despite the imbalance class distribution, it is important to ensure that the grouping of hadith topics is not only accurate for the majority categories but also effective for rare classes.

### 3.3 Application of Semantic Search

This section explains the process of searching for hadiths using a semantic search approach. The TF-IDF and SBERT approaches are used as text representations. TF-IDF analyzes the occurrence of words in the hadith text, while SBERT uses deep learning to understand concepts and meanings.

After both approaches were implemented, the results of the hadith search were evaluated using five user queries that are frequently used in Islamic studies. The queries included “ablution water,” “fasting in Ramadan,” “inheritance rules,” “forgiveness of sins,” and “helping the poor.” The results of both approaches were evaluated manually to determine their semantic relevance.

Based on the results of manual evaluation, it can be seen that the cosine similarity value in the TF-IDF method is generally higher than that in the SBERT method. However, a high cosine similarity value does not always reflect the semantic relevance of the hadith to the user's query. This can be seen from the manual relevance scores (0: irrelevant, 1: relevant), which show that the TF-IDF method obtained a relevance score of 3 out of a total of 5 hadiths (60%), while the SBERT method obtained a relevance score of 4 out of a total of 5 hadiths (80%). For example, for the queries “rules of inheritance” and “helping the poor,” the TF-IDF method produced hadiths with fairly high cosine similarity values, but after in-depth analysis, these hadiths were not semantically relevant to the context of the user's query. On the other hand, the SBERT method is capable of producing semantically relevant hadiths for all queries tested, although the similarity values tend to be lower than those obtained using the TF-IDF method. To clarify the difference in similarity scores visually, Figure 3 illustrates the cosine similarity obtained from both methods.

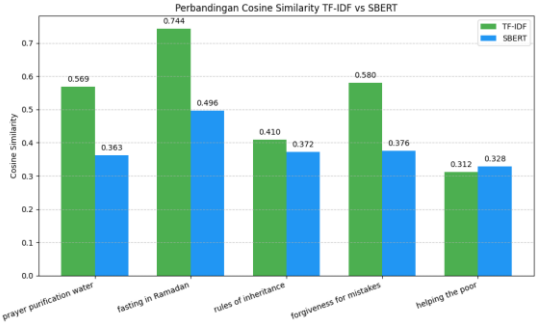


**Figure 4.** Visualization of Cosine Similarity TF-IDF and SBERT

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 3, December 2025 Page: 2017−2024
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8861

Figure 4 above shows a comparison of cosine similarity values between the TF-IDF and SBERT methods for five user queries. Overall, it can be seen that the TF-IDF method consistently produces higher similarity scores than the SBERT method. However, as explained earlier, high cosine similarity values in the TF-IDF method do not always reflect high semantic relevance. On the other hand, although the similarity values obtained by SBERT tend to be lower, this method is able to provide more semantically and contextually relevant hadith search results. This is because the SBERT embedding representation is able to capture the meaning of sentences comprehensively, not limited to literal word matching alone.

### 3.3.1 Analysis of the Relevance of Hadith Search Result

From a manual evaluation of the hadiths obtained, several advantages and disadvantages of each search method were found. The TF-IDF method has advantages in terms of computational speed and simplicity of data representation. This method works very well when searching for hadiths using explicit and literal keywords. The primary drawback of TF-IDF is its inability to grasp semantic connections and the broader context of a sentence. As a result, some selected hadiths show high numerical similarity scores but may not always have semantic meaning.

The SBERT method offers the primary advantage of deeply understanding semantic meaning and can recognize relationships among different semantically related ideas. This allows SBERT to be highly effective in generating contextually relevant hadiths, as seen in questions such as “inheritance regulations” and “assisting the needy.”. However, SBERT has the drawback of higher computational complexity and produces relatively lower cosine similarity scores compared to TF-IDF. This happens because of the compact embedding representation in SBERT, resulting in increased spread in distances between vectors.

Based on the overall review, we could infer from the overall assessment that the SBERT approach performs reasonably well in generating semantically relevant hadiths. Moreover, the time required to build the indexing structure is also faster than the model presented in this paper, making it potentially useful as an interactive baseline. However, it may not fully support deep semantic exploration. Therefore, the SBERT approach is highly recommended for developing a hadith search system based on semantic search to achieve results that better match users’ needs.

## 4. CONCLUSION

In summary, this research developed an English language hadith search system by implementing automated topic classification using a CNN model and two text representation methods, i.e., TF-IDF (Term Frequency-Inverse Document Frequency) and Sentence-BERT (SBERT), to enhance semantic search. The CNN models categorized hadith into seven thematic categories through an optimized CNN framework, resulting in a simplified data structure, and the classification accuracy is higher. Once classified, hadith searches are conducted by the system based on users' queries by using two text representation techniques: TFIDF (statistical method driven by frequency) and SBERT (deep learning technique to capture the meaning and the context of a sentence). The results of the experiment indicate that while TF-IDF yields higher numerical cosine similarity scores, SBERT excels at retrieving hadiths that are semantically contextual to the users’ queries. Compared to TF-IDF, SBERT is able to provide search results that are more relevant to the context of the user's query and is a better choice because SBERT is able to understand the meaning of sentences better than TF-IDF. However, the study still has several points that can be further developed, such as the small number of test questions, the relevance assessment is still manual/subjective, and the dataset only covers English translations of hadiths. Future developments could include cross-language search, including Arabic, testing queries with more diverse user scenarios, and further integration between topic classification and semantic search for more adaptive and accurate results.

## REFERENCES

[1] M. Amin, “EASY STEPS TO SEARCH AND USE HADITH FOR THE GENERAL PUBLIC,” *AQWAL Journal of Qur’an and Hadis Studies*, vol. 3, no. 2, pp. 140–162, Jan. 2023, doi: 10.28918/aqwal.v3i2.3149.

[2] K. I. Siregar, “Pelatihan Memahami Hadits ‘Hidupkanlah Saya Miskin’ dengan Metode Semantik di Pesantren Modern Nurul Huda-Cikarageman-Setu-Bekasi,” *Satwika : Jurnal Pengabdian kepada Masyarakat*, vol. 1, no. 1, pp. 1–5, Dec. 2021, doi: 10.21009/satwika.010101.

[3] D. C. Hidayati, S. Al Faraby, and A. Adiwijaya, “Klasifikasi Topik Multi Label pada Hadis Shahih Bukhari Menggunakan K-Nearest Neighbor dan Latent Semantic Analysis,” *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, p. 140, Feb. 2020, doi: 10.30865/jurikom.v7i1.2013.

[4] Yudi Widhiyasana, Transmissia Semiawan, Ilham Gibran Achmad Mudzakir, and Muhammad Randi Noor, “Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 4, pp. 354–361, Nov. 2021, doi: 10.22146/jnteti.v10i4.2438.

[5] B. Juarto and A. Suganda Girsang, “Neural Collaborative with Sentence BERT for News Recommender System,” *JOIV : International Journal on Informatics Visualization*, vol. 5, no. 4, p. 448, Dec. 2021, doi: 10.30630/joiv.5.4.678.

[6] A. Ramadhani, N. Safaat, S. Agustian, I. Iskandar, and S. Sanjaya, “Perbandingan Performa Metode Klasifikasi Teks Multilabel Hadis Terjemahan Bukhari Menggunakan Support Vector Machine dan Long Short Term Memory,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, pp. 896–907, Jul. 2025, doi: 10.57152/malcom.v5i3.2051.

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 3, December 2025 Page: 2017−2024
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8861

[7] W. Darmalaksana, C. Slamet, W. B. Zulfikar, I. F. Fadillah, D. S. adillah Maylawati, and H. Ali, "Latent semantic analysis and cosine similarity for hadith search engine," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 1, pp. 217–227, Feb. 2020, doi: 10.12928/TELKOMNIKA.V18I1.14874.

[8] I. Akbar, M. Faisal, and T. Chamidy, "Penerapan Long Short-Term Memory untuk Klasifikasi Multi-Label Terjemahan Al-Qur'an dalam Bahasa Indonesia," *Journal of Information Technology and Computer Science)*, vol. 9, no. 1, pp. 41–54, Jun. 2024, doi: 10.31328/jointecs.v8i1.5291.

[9] E. T. Luthfi, Z. I. M. Yusoh, and B. M. Aboobaider, "Enhancing the Takhrij Al-Hadith based on Contextual Similarity using BERT Embeddings," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, p. 2021, 2021, doi: 10.14569/IJACSA.2021.0121133.

[10] M. Yuslan Abu Bakar and Adiwijaya, "KLASIFIKASI TEKS HADIS BUKHARI TERJEMAHAN INDONESIA MENGGUNAKAN RECURRENT CONVOLUTIONAL NEURAL NETWORK (CRNN)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 8, no. 5, pp. 907–918, Oct. 2021, doi: 10.25126/jtiik.202183750.

[11] D. Saputra, F. Firman, A. Rafika, and S. A. Yasti, "Historical Bacground Dan Manhaj Kepenulisan Kitab Al Jami' Al Bukhari," *Ushuly: Jurnal Ilmu Ushuluddin*, vol. 3, no. 1, pp. 37–54, Jan. 2024, doi: 10.52431/ushuly.v3i1.2321.

[12] F. P. Utami, "ANALISIS KUALITAS HADIS RIWAYAT SHAHIH MUSLIM NO. 2700: KAJIAN SANAD DAN MATAN," *At-Thullab : Jurnal Mahasiswa Studi Islam*, vol. 7, no. 1, Jan. 2025, doi: 10.20885/tullab.vol7.iss1.art9.

[13] M. Lamba and M. Madhusudhan, "Text Pre-Processing," in *Text Mining for Information Professionals*, Cham: Springer International Publishing, 2022, pp. 79–103. doi: 10.1007/978-3-030-85085-2_3.

[14] M. U. Albab, Y. K. P., and M. N. Fawaiq, "Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic," *Jurnal Transformatika*, vol. 20, no. 2, pp. 1–12, Jan. 2023, doi: 10.26623/transformatika.v20i2.5374.

[15] M. R. Athallah and K. M. Lhaksmana, "Hadith Text Classification Based on Topic Using Convolutional Neural Network (CNN) and TF-IDF," *Journal of Renewable Energy, Electrical, and Computer Engineering*, vol. 5, no. 1, pp. 30–36, Mar. 2025, doi: 10.29103/jreece.v5i1.20354.

[16] C. C. Aggarwal, *Neural Networks and Deep Learning : a Textbook*. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-94463-0.

[17] Ž. Vujović, "Classification Model Evaluation Metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.

[18] A. H. Dani and E. Y. Puspaningrum, "Studi Performa TF-IDF dan Word2Vec Pada Analisis Sentimen Cyberbullying," *Router : Jurnal Teknik Informatika dan Terapan*, vol. 2, no. 2, pp. 94–106, 2024, doi: 10.62951/router.v2i2.76.

[19] O. Galal, A. H. Abdel-Gawad, and M. Farouk, "Rethinking of BERT sentence embedding for text classification," *Neural Comput Appl*, vol. 36, no. 32, pp. 20245–20258, Nov. 2024, doi: 10.1007/s00521-024-10212-3.

[20] M. A. Rifai and I. G. Anugrah, "Semantic Search for Scientific Articles by Language Using Cosine Similarity Algorithm and Weighted Tree Similarity," *Journal of Development Research*, vol. 5, no. 2, pp. 106–114, Nov. 2021, doi: 10.28926/jdr.v5i2.150.

[21] D. W. Eller, "Transparency and the future of semantic searching in academic libraries," *Inf Serv Use*, vol. 42, no. 3–4, pp. 453–461, Aug. 2022, doi: 10.3233/ISU-220175.