JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 8 Number 2 | December 2023: 242-251
DOI: 10.15575/join.v8i2.1205

Classification of Bulughul Maraam Categories: Prohibitions, Recommendations, and Information Using Extreme Learning Machine and Fasttext

Rissa Nurfitriana Handayani[1], Ina Najiyah[2*], Dirga Astya Wisnuwardana[3]
[1,2,3]Department of Information System, Universitas Adhirajasa Reswara Sanjaya, Bandung, Indonesia

Article Info

Article history:

Received October 18, 2023
Revised November 15, 2023
Accepted November 15, 2023
Published December 28, 2023

Keywords:

Bulughul Maraam Book
Classification of Shahih Hadith
Extreme Learning Machine for Classification
Text Mining

ABSTRACT

Hadith is the second source of Islamic law after the Quran. After the hadiths were compiled, Imam of Hadith created collections of hadiths, one of which is Imam Bukhari who compiled the book Bulughul Maraam, which is considered to have the highest level of authenticity. Digital collections of hadiths can now be found in the form of e-books and web pages, which help in the search for hadiths. The classification of hadiths is necessary to organize them by category, making it easier to search for hadiths based on their categories. Text mining is needed to classify hadiths because it can identify patterns in unstructured text. This research aims to improve the accuracy of classifying recommended, prohibited, and informational hadiths using a dataset of 7008 hadiths, which consists of primary data taken from the book Bulughul Maraam in the Indonesian language. Previously, similar research was conducted in 2017 that classified recommended, prohibited, and obligatory hadiths with an accuracy of 85%, but only for Sahih Bukhari hadiths. In this research, the same classification categories will be examined, proposing a different method, namely the Extreme Learning Machine method and Word2vec Fasttext for text representation with a larger dataset. The results of this research show a model accuracy of 86.31%, 86% precision, and 87% recall, indicating that the proposed model performs well in classifying hadiths.

Corresponding Author:

Ina Najiyah,
Information System Departement, Faculty of Technology and Information, Universitas Adhirajasa Reswara Sanjaya
Jl. Sekolah Internasional No 1-2 Antapani Bandung 40282, Indonesia
Email: inajiyah@ars.ac.id

1. INTRODUCTION

The book that serves as the primary guidance for the Islamic religion is the Quran, while hadiths are explanations of the Quranic verses used as companions or secondary guidance after the Quran[1]. The collection of hadiths narrated by renowned and prominent *Imam* has spread throughout the entire Muslim community for study, one of which is the collection of hadiths narrated by Imam Bukhari, which has the highest level of *Shahih* among other *Imam* [2] One of the books narrated by Imam Bukhari and several other transmitters, considered to have a high level of authenticity, is the book of Bukhari. Inside the Bukhari book, there are 16 chapters and 97 sub-chapters that encompass the foundations of Islamic law, including daily activities like the chapter on prayer, etiquette, and even specific activities like the chapter on marriage.

The classification of these chapters makes it easier for the Muslim community to find hadiths related to those categories [3]. After rapid technological advancements, especially in the era following COVID in the year 2020, society began to shift towards the digital world, both in terms of social media and the pursuit of knowledge [4], even digital collections of hadith books are very easy to find, especially in this technology-driven era, whether in the form of files, articles on a webpage, or even within mobile and desktop applications.

One of the online platforms that provides the Bulughul Maraam book is https://alquran-sunnah.com/kitab/bulughul-maram/index.html. On this website, users are facilitated in searching for hadiths based on their respective chapters.

Digital hadiths in the form of websites, as mentioned earlier, can facilitate the search for references or sources. However, a common issue encountered with the abundance of these hadith texts or documents is the difficulty in finding hadiths with different themes or different search objectives [5] For instance, if people want to search for hadiths concerning what prohibitions are to be avoided by Muslims or what Information are to be practiced by the Islamic community. The solution to this issue necessitates the classification of hadiths according to their themes, such as recommended hadiths, prohibitions, and informational ones, in order to facilitate Muslims in searching for hadiths based on their respective categories.

Text classification can be performed using text mining due to its ability to search for similarities or patterns [6] in the translated hadith text that distinguishes its categories, text classification of Bukhari Hadith Translation in Indonesian has been previously conducted using Recurrent Extreme Learning Machine (CRNN), with a performance result of 80.79% [7], A thesis-level research was also conducted to classify the categories of prohibitions, recommendations, and Information in hadith using the Naïve Bayes Classifier method, achieving an accuracy of 79.14% [8]. That research served as the primary reference for this study, and modifications were made to the model's working method to classify hadiths based on their categories by adding data preprocessing steps and experimenting with other algorithm models to improve accuracy. In this study, the Extreme Learning Machine model is proposed to assist in text classification. This model has been previously used in research for "Automated Rating Prediction Based on Restaurant Reviews," using textual data and achieving a fairly good accuracy of 80.01% [9].

Based on the research of [7] and [8], there are differences with the author's proposed research. In studies [7] and [8], the classified categories are the same, which are the classification of prohibition, recommendation, and information in hadiths. However, the data collected in those studies is limited to hadiths narrated by Bukhari, whereas in this research, the data is obtained from Bulughul Maraam, which includes hadiths narrated by various Imams, not just Imam Bukhari. On the other hand, in study [9], the difference with this research lies in its subject. In study [9], the subject is restaurant reviews, while in this research, the subject is hadiths, even though both studies use the same method, which is the Extreme Learning Machine. The extreme learning machine is used in this research because of its suitable working mechanism, which involves dividing the filtering process into several layers: the input layer, hidden layer, and output layer. This approach enhances the classification process, resulting in the highest level of accuracy.

Word2vec Fasttext, also known as FastText, is used to assist in word representation or converting word weights into one-hot vectors. The use of FastText to enhance accuracy in the classification of hadiths with extensive data is considered suitable because FastText can identify different words with the same meaning, as demonstrated in research showing that FastText outperforms the other two Word2vec FastText methods [10].

The aim of this research is to separate or classify hadiths based on the categories of prohibitions, recommendations, and Information. It is hoped that this research can make a significant contribution to the development of religious knowledge, particularly within Islam, and facilitate the search for hadiths, thereby reducing errors in the comprehension of their meanings. The writing structure or writing process involves identifying the raised issues and presenting them in the introduction. Subsequently, the methodology is designed, and all the stages are detailed in the discussion section. Once the discussion process is complete, conclusions can be drawn from the research. This research is conducted using technology that has proven its effectiveness, namely text mining, to maximize technology in the field of religion. The text classification process on hadith text data can assist the community in categorizing hadiths and contribute to the field of religion as a reference for further research.

2. METHOD

The process or stages involved in conducting this research are explained in this chapter, where the process starts from data collection and continues until the desired research output or objectives are achieved. The research stages are presented in Figure 1.

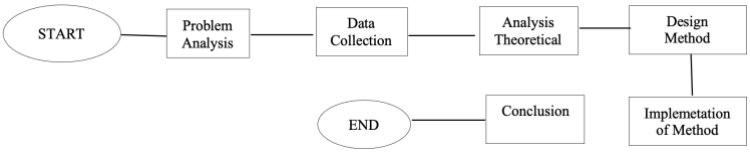

Figure 1. Research Stages

The stages explained in Figure 1 represent the step-by-step process of conducting this research. The explanation is as follows:

a. Problem analysis: In this stage, an analysis of the problem to be addressed is conducted, and solutions to resolve these issues are sought. This includes consulting with hadith experts, searching for supporting theoretical references in journals and books, and conducting questionnaires with a group of individuals to gain a more detailed understanding of the problem. The problem statements identified are as follows:
   1) Many Muslims who study hadith and wish to search for hadiths based on categories, such as what prohibitions are there in Islam.
   2) Previous research has categorized hadiths as prohibitions, recommendations, and information, but only from Sahih Bukhari, so the dataset needs to be expanded.
   3) Accuracy in previous hadith research can still be improved using alternative methods.
b. Analysis Therotical: In this stage, the author analyzes and gathers theories that support the research and align with the research objectives to create a sound research outcome. These theories are then applied in the research implementation process.
c. Data collection: After identifying the research problem, a dataset is collected, which includes hadiths found in the Bulughul Maraam book. Once the data has been successfully gathered, it is then labeled based on discussions with three hadith experts to create accurate data.
d. Design method: This process is carried out once all the data is collected and the proposed solutions are clear. In this stage, the Extreme Learning Machine method is designed, and text representation using FastText is designed, and they are represented using diagrams.
e. Implementation of Method: Once the method has been successfully designed and created, the proposed model is implemented. In this research, implementation is carried out using the Python programming language and adopting Python libraries that support the Extreme Learning Machine and FastText methods.
f. Conclusion: This stage is the final step in which it is concluded whether the proposed model achieves good performance and includes suggestions for improvements based on the research that has been proposed.

In addition to the previously described stages, the following is the theoretical framework in this research as shown in Figure 2.

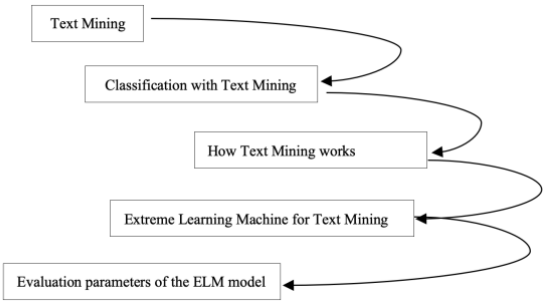

Figure 2. Theoretical Framework

1) *Text mining*

Text mining aims to uncover valuable information that is hidden, whether from structured or unstructured information sources. It is a process of knowledge discovery within text-based databases (also known as text data discovery or TDD). It can also be interpreted as the activity of unearthing or searching for text data that was previously unknown, so that the information can be understood, possess practical potential and patterns, and become new knowledge. This process focuses on large collections of text data or corpora (sets of text encompassing the use of language in complete and dense written form) that are inherently unstructured. [11]

2) Classification with Text mining

Classification is a method in data mining used to group data based on how closely it is associated with example data [12]

3) The Operation of Text Mining

In the explanation of the Concept of Text Processing Representation, it is found that the overall structure of text processing involves two main components: text cleaning steps that transform unstructured text into a pre-defined format, and an information filtering process (knowledge reduction) that represents patterns or information from the existing model [13]. The illustration in Figure 3 demonstrates a Text Processing Sequence [14]

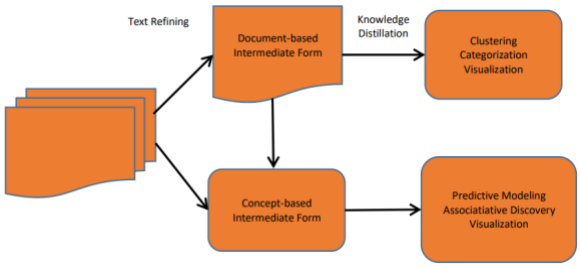

Figure 3. Framework of Text Mining

4) Extreme Learning Machine Method

ELM is a feedforward artificial neural network that implements the idea of the Single Hidden Layer Feedforward Neural Networks (SLFNs). This is because ELM uses only one hidden layer in its structure. The purpose of creating ELM is to address the weaknesses of previous artificial neural networks in terms of learning speed [15]. By determining parameters such as input, weights, and hidden biases randomly, ELM can enhance learning speed, achieve good generalization performance, and avoid potential issues like overtraining. The application of ELM aims to accommodate the growth of data every month and the addition of necessary external factors to achieve accurate forecasting results. The processes within the ELM method can be described as follows: [16]

a) Data Normalization

Data normalization is a transformation step that scales data into a more detailed value range, typically from 0 to 1, with the goal of aligning input data with output data. The normalization technique applied involves dividing each parameter value in the original data by its maximum value [17]. The steps for data normalization are explained by Equation 1:

$$v^1 = \frac{v}{max} \quad (1)$$

Explanation :

v^1 = new value

v = original data

max = maximum data for each parameter

b) Training process

The step aimed at training using the training data. Through this training, weight values that yield optimal results will be obtained. The training process consists of the following steps:

- Random values are generated within the range [0,1] for the Wmn matrix, which will be used as input weights. This matrix is organized as an array with a size of m (number of hidden neurons) times n (number of input neurons). Furthermore, random values are created for the bias matrix b with a size of 1 times m (number of hidden neurons), also within the range [0,1] [18].
- The output matrix from the hidden layer is calculated using equation 2. To calculate b (ones(itrain,1)), this matrix is expanded to match the number of available training data [18].

$$H = \frac{1}{1+exp\left(-\left(x_{train}w^T + b_{ones}\left(i_{train,1}\right),:\right)\right)} \quad (2)$$

Explanation:

H = output matrix of the hidden layer

X_{train} = input matrix of the normalized training data

W^T = transpose matrix of the weights

i_{train} = number of training data

b = bias matrix

In the training process of this research, the steps are executed following a consistent sequence. This is done because the primary focus of this research is to test the performance of the applied method and to identify the impact of adding additional variables (external factors) on the ability to produce accurate predictions.

c) Testing Phase After completing the training process, the next step involves the testing phase using test data. The purpose of this testing phase is to assess the results of the training that has been conducted, with the intention of evaluating how accurate the program is.

d) Denormalization of Data The function of this process is to return the normalized values to their original form [19]. Model Evaluation Parameters for ELM There are three metrics used to measure the performance of the classification system that has been constructed: precision, recall, and accuracy. Precision value is

an indicator of the system's sensitivity or accuracy in providing correct information related to the negative or positive class of data. On the other hand, recall measures how accurately the system can correctly identify negative class data or text with positive content. The calculation of precision and recall values can be determined using the formulas in equations 3 and 4, as follows [20]

$$precision = \frac{1}{2}\ x\ \left(\frac{negative\ -\ F.N}{negative\ -\ F.N\ +\ F.P} + \frac{positive\ -\ F.P}{positive\ -\ F.P\ +\ F.N}\right)\ x\ 100\% \quad (3)$$

$$recall = \frac{1}{2}\ x\ \left(\frac{negative\ -\ F.N}{negative} + \frac{positive\ -\ F.P}{positive}\right)\ x\ 100\% \quad (4)$$

The accuracy value represents the comparison between the number of tweets correctly detected in the test data. Furthermore, accuracy indicates how closely the system's predictions align with predictions provided by humans. The accuracy value can be calculated using the following Equation 5[21]:

$$accuration = \frac{negative\ +\ positive\ -\ F.P\ -\ F.N}{negative\ +\ positive}\ x\ 100\% \quad (5)$$

3. RESULT AND DISCUSSION

The research results produce an output that can be analyzed to obtain meaningful information. Here is a description of the outcomes of the conducted researc.

3.1. Problem Analysis

Bulughul Maraam is a book that was compiled by various Imams as a guide for seeking shahih hadiths. In the book of Bulughul Maraam, there are a total of 7008 hadiths organized into 16 chapters and 97 sub-chapters, categorized as follows:

a. The discussion on purification
b. The discussion on prayer
c. The discussion on funeral rites
d. The discussion on almsgiving (zakat)
e. The discussion on fasting
f. The discussion on the pilgrimage (hajj)
g. The discussion on trade and transactions
h. Regarding marriage
i. Regarding criminal offenses
j. Regarding punishments
k. The discussion on food
l. The discussion on oaths and vows
m. The book of settling disputes (al-Qadha')
n. The discussion on freeing slaves
o. The discussion on a collection of various chapters
p. The discussion on comprehensiveness

From all those chapters, the mixing of which ones carry prohibitions and which do not remains. This causes difficulties for hadith scholars in searching for specific categories. Therefore, there is a need for new categories that align with the meaning of the hadith, whether it contains prohibitions, commands, or information.

3.2. Document Collection

The data used in this research is primary data. Primary data is data collected and obtained directly by the researcher. The primary data is obtained from the authentic book of Bukhari, consisting of 7008 hadiths and written in Bahasa Indonesia. The following is the hadith data from the book of Bulughul Maraam as the initial data in Figure 4.

No	Arab	Terjemah
1	حَدَّثَنَا الْحُمَيْدِيُّ عَبْدُ اللَّهِ بْنُ الزُّ...	Telah menceritakan kepada kami Al Humaidi Abdullah...
2	حَدَّثَنَا عَبْدُ اللَّهِ بْنُ يُوسُفَ قَالَ أَخْبَ...	Telah menceritakan kepada kami Abdullah bin Yusuf ...
3	حَدَّثَنَا يَحْيَى بْنُ بُكَيْرٍ قَالَ حَدَّثَنَا ...	Telah menceritakan kepada kami Yahya bin Bukair be...
4	حَدَّثَنَا مُوسَى بْنُ إِسْمَاعِيلَ قَالَ حَدَّثَنَ...	Telah menceritakan kepada kami Musa bin Isma'il di...
5	حَدَّثَنَا عَبْدَانُ قَالَ أَخْبَرَنَا عَبْدُ اللَّ...	Telah menceritakan kepada kami Abdan dia berkata, ...
6	حَدَّثَنَا أَبُو الْيَمَانِ الْحَكَمُ بْنُ نَافِعٍ...	Telah menceritakan kepada kami Abu Al Yaman Al Hak...
7	حَدَّثَنَا عُبَيْدُ اللَّهِ بْنُ مُوسَى قَالَ أَخْ...	Telah menceritakan kepada kami Abdullah bin Musa d...
8	حَدَّثَنَا عَبْدُ اللَّهِ بْنُ مُحَمَّدٍ الْجُعْفِ...	Telah menceritakan kepada kami Abdullah bin Muhamm...
9	حَدَّثَنَا آدَمُ بْنُ أَبِي إِيَاسٍ قَالَ حَدَّثَنَ...	Telah menceritakan kepada kami Adam bin Abu Iyas b...
10	حَدَّثَنَا سَعِيدُ بْنُ يَحْيَى بْنِ سَعِيدٍ الْقُ...	Telah menceritakan kepada kami Sa'id bin Yahya bin...
11	حَدَّثَنَا عَمْرُو بْنُ خَالِدٍ قَالَ حَدَّثَنَا ا...	Telah menceritakan kepada kami 'Amru bin Khalid be...
12	حَدَّثَنَا مُسَدَّدٌ قَالَ حَدَّثَنَا يَحْيَى عَنْ...	Telah menceritakan kepada kami Musaddad berkata, t...
13	حَدَّثَنَا أَبُو الْيَمَانِ قَالَ أَخْبَرَنَا شُعَ...	Telah menceritakan kepada kami Abu Al Yaman berkat...

Figure 4. Original Dataset in Bahasa Indonesia

The initial data consists of IDs and the content of hadiths. The content of the hadiths includes both the Arabic text and the Indonesian translation. Essentially, the content of the hadiths is divided into two parts, namely the "sanad" and "matan." This research will only use the Indonesian text as its feature. The following is the hadith data after removing the Arabic text in Table 1.

Table 1. Translation of Hadith Data in Indonesian

Id	Translation
1	Al Humaidi Abdullah bin Az Zubair told us, he said, Sufyan told us that Yahya bin Sa'id Al Anshari told us that Muhammad bin Ibrahim At Taimi informed us that he heard Alqamah bin Waqash Al Laitsi say: I heard Umar bin Al Khaththab say on the pulpit, 'I heard the Messenger of Allah, may peace and blessings be upon him, say, 'Actions are but by intention, and every man shall have only that which he intended. Thus, he whose migration was for Allah and His messenger, his migration was for Allah and His messenger; and he whose migration was to achieve some worldly benefit or to take some woman in marriage, his migration was for that for which he migrated
2	Abdullah bin Yusuf told us, he said, Malik informed us from Hisyam bin 'Urwah, from his father, from Aisyah, the Mother of the Believers, that Al Harits bin Hisyam asked the Messenger of Allah, may peace and blessings be upon him, 'O Messenger of Allah, how does revelation come to you?' The Messenger of Allah, may peace and blessings be upon him, replied, 'Sometimes it comes to me like the ringing of a bell, and this is the most difficult for me, but then it stops, and I understand what is conveyed. And sometimes an Angel comes in the form of a man and speaks to me, and I follow what he says.' Aisyah said, 'I have indeed seen the revelation coming to him, the Messenger of Allah, may peace and blessings be upon him, on a very cold day, and then it stopped, and I saw sweat on his forehead

3.3. Design Method

3.3.1. Feature Selection

Matan is the content of the hadith, which represents the words of the hadith, while the sanad is the chain of narrators [22]. The author only uses the matan in classifying hadiths into recommendations, prohibitions, and information. The following is the hadith data with the sanad removed in Table 2.

Table 2. Data of Hadith Matan

Id	*Matan*
1	All deeds depend on their intentions, and the recompense for each person is based on what they intended. Whoever migrates with the intention of worldly gain or to marry a woman, their migration is for what they intended
2	O Messenger of Allah, how does revelation come to you?' The Messenger of Allah, may peace and blessings be upon him, replied, 'Sometimes it comes to me like the sound of ringing bells, and this is the most difficult for me. It stops, and then I understand what is conveyed. And sometimes an Angel in the form of a man comes and speaks to me, and I follow what he says.' Aisyah said, 'I have indeed seen the revelation coming to him, the Messenger of Allah, may peace and blessings be upon him, on a very cold day. It stopped, and I saw sweat on his forehead

3.3.2. Categorization of Hadith

The data from Sahih Bukhari hadiths is categorized by the author into three categories: hadiths containing recommendations, prohibitions, and information. This classification is not arbitrary. It is conducted under the guidance and direction of three hadith experts. The hadith data includes the matan, which is the translation of the hadith in Indonesian, along with the labels that have been created. The creation of this dataset takes into account the words and meanings contained in each matan comprehensively, allowing for accurate categorization. The number of datasets and their categories selected by the author are explained in Table 4.

Table 4. Total of Dataset

No	Class/Categories	Total
1	Recommendations	659
2	Informations	5830
3	Prohibitions	519
	Total	7008

3.4. Implementation of the Proposed Method

3.4.1. Processing of Initial Data

The collected data is not processed immediately but will be separated again in the range of proportions from 10% to 90%. This grouping is done for the purpose of training data testing. This way, optimal results from variations in the proportion of document divisions can be obtained.

The stage that follows data collection is preprocessing. It includes the following steps.

a. Tokenizer

A tokenizer is a process that breaks sentences into words. An example of the tokenizer's output in this research.

Sentence:

The Prophet, may peace and blessings be upon him, prohibited prayer after Fajr until sunrise and after Asr until sunset"

After the tokenizer process:

"The" "Prophet," "may" "peace" "and" "blessings" "be" "upon" "him," "prohibited" "prayer" "after" "Fajr" "until" "sunrise" "and" "after" "Asr" "until" "sunset"

b. Feature Representation Using Word2Vec FastText

The translation of Sahih Bukhari hadith text data in this research includes a large amount of data, totaling 7008 data with 201,222 texts. To facilitate the classification process and achieve maximum accuracy, feature representation is used, specifically Word2Vec FastText. Word2Vec FastText is suitable for this research because it contains many different words that actually have the same meaning. The feature representation using Word2Vec FastText or Word2Vec is as follows:

1) *One-hot* Matrix

The initial step in the process is to determine the weight or encoding of each element in a document. Here's an example of the one-hot matrix step:

Table 5. One-Hot Matrix

Prohibitions	Information	Recommendations
The Prophet, may peace and 0 0 0 0 0	The Prophet, may peace and 0 0 0 0 0	Worship Allah SWT perfectly 1 0 0 0
blessings be upon him, prohibited 0 0 0 0 1	blessings be upon him, once 0 0 0 0 1	without associating partners 0 0 0
prayer after Fajr until sunrise and 0 0 0 0 0 0	recited the Quran while his head 1 0 0 0 0 0	with Him, and establish the 0 0 0 0 0
after Asr until sunset 0 0 0 0	was resting on my lap, and I was 0 0 0 0 0 0 1 0	prayer 0
	in a state of menstruation 0 0 0 0 0	
Don't intentionally perform the 1 0 0 0		Dispose of the mouse carcass 1 0 0 0 0
prayer when the sun is rising or 0 0 0 0 0 0 0		and the fat around it 0 0 0 0 0
when it is setting. 0 0 0 0		

2) *Continuous Bag-of-Words(CBOW)*

Model Continuous Bag-of-Words (CBOW) often used in deep learning. The model attempts to predict words given the context of several words before and after the target word. CBOW is not sequential and does not have to be probabilistic.

3) Determining vector dimensions

The previous stage resulted in words with a value of 1, which are words like "melarang," "janganlah," and "*mengharamkan*" for the classification of prohibitions. Meanwhile, in the Classification of Matan Hadith Recommendations, words with a value of 1 include "buanglah" and "hendaklah." Therefore, if these words are searched for their vector values, they will yield similar numbers. In this stage, the number of vector dimensions will be determined. For example, the vector dimensions of the words "don't" and " prohibited" total 50 dimensions. The 50-dimensional vectors are as follows:

don't – 0,90962 -0,16793 -0,28244 1,2006 0,334477 -0,054996 -1,2784, 0,056494 -0,7309 -0,65440 0,084254 -0,31206 -0,0033612 0,0821280 0,32966 -1,01 0,11907 -0,087034 0,99843 -1,5931 3.4081 1.0488 0,20391 -0,43841 0,1119 0,87173 0,62297 0,62392 0,2883 -0,97984 -0,016959 -0, 39827 -0,97197 0,59103 -0,57006 0,96217 0,76927 0,40248 -0,013602 0,10745

prohibitted – 0,66488 -0,11391 0,67944 0,17951 0,6828 -0,47797 -0,30761 0,17489 -0,70512 -0,55607 -0,68297 0,037364 0,70266 1,9093 -0,61483 -0,83329 -0,3023 -1,1110 -1,55 0,2604 0,22957 -1,0375 3,5091 -0,25071 1,0151 0,65927 -0,18231 -0,75859 -0,30927 -0,91678 1.0633 -0,66761 -0,37464 -0,29143 0,65606 -0,44642 -0,075495 -1,0552 -0,60501 0,73582 1,0139 -0,27749

c. Implementation of Word2Vec FastText with Python

1) Determining Input and Output Layers (Word2vec Fasttext Process)
The process of determining the input and output layers in IPython using the Keras library. Keras.Sequential is a simple type of neural network consisting of a stack of layers executed sequentially. You can create a stack with only two layers (input and output) to complete the neural network; without hidden layers, it will not be considered a deep neural network. The input and output layers are the most crucial because they define the overall shape of the neural network. Most of the network consists of Dense Layers, which are linear neural net layers comprising inputs, weights, and outputs. In this research case, sentences to be converted into a one-hot matrix have a maximum length of 3000 words. The output from these layers is 512, and the type of maximization or "activation" to be used needs to be determined. Activation functions are used when training the network; they inform the network how to evaluate when the weights for a particular node have created a good fit. In the first layer, ReLU is used (also for fun). Between the input and output layers, there is another Dense Layer and two Dropout layers. Dropout is used to randomly remove nodes, which can help prevent overfitting. Overfitting can occur when the training data is the same or too similar. In the training data, accuracy may remain stable or even decrease instead of increasing.

2) Determining the Accuracy Matrix for the Training Model Stage
The final step before training is to configure the network model for accuracy.

3.4.2. Experimentation and Method Testing

a. Classification Algorithm (Extreme Learning Machine)

The main focus of this study revolves around the classification process performed using the Extreme Learning Machine algorithm. The pre-selected features will be utilized as inputs by the Extreme Learning Machine in its computations, aiming to classify documents. In this step, the training documents are used as inputs for this process. The sequence of algorithms applied to the Extreme Learning Machine is as follows:

1) Training the Created ELM Model
This stage is the core phase that requires a considerable amount of time for training using the training data. After training with the data, the next step involves grouping by batch size, then checking the network's accuracy, adjusting node weights, and running through another batch. Small batches allow for faster network training. Epochs determine how many times this batch-by-batch separation is performed. In this research, the author tried 5 epochs and 10 epochs, with 10 epochs showing better accuracy and no overfitting compared to 5 epochs. Validation_split refers to how much input is desired for testing data. It's crucial to assess the network's accuracy at that point. The recommended ratio for training documents is 80:20 or 90:10.

2) Saving the Model
After training on the data and obtaining accuracy, the next step is to save the model, so that you don't need to start from scratch when you want to train it again.

3) Apply Model
This step is used to execute the previously assembled model. Using the training documents as test data, a re-evaluation will be conducted to assess the success of the classification during the training phase. Validation for the training and testing phases will be done through 10 iterations of epochs.

b. *Performance Measure*

After the model has been implemented, the next step is to measure the performance of the executed model and carry out validation. The output of this step includes precision, recall, and, of course, accuracy values. These values will be compared to determine the model with the best performance. All results from the validation process will yield a model and performance evaluation outcomes. Subsequently, these results will be presented in the form of a confusion matrix, and at the same time, the trained model will be saved. The saved model will be used to test different document samples.

c. Measurement Value

It is the final result in the form of a table (confusion matrix) or a graph that displays the values of precision, recall, and accuracy.

3.4.3. Performance Report

The model proposed in this study is implemented by adopting the Extreme Learning Machine algorithm. The Word2Vec Fasttext approach is applied as an optimization step in the feature representation phase. The results of this representation are then passed on to the classification process by the Extreme Learning Machine. Evaluation is conducted to calculate accuracy, precision, and recall. After successfully identifying the highest accuracy, the next step involves classification testing. This aims to formulate the most ideal algorithmic structure in solving the encountered issues. Once the preprocessing and feature selection steps are

completed, the next phase involves the classification process through validation. The evaluation and validation processes use the proposed model, and to optimize compatibility with the machine, Keras and TensorFlow are also employed. The results can be summarized as follows:

a. Results of Testing with Word2vec Fasttext and ELM

To determine whether the proposed model by the author achieves the best accuracy, the author conducted experiments with several other methods, namely Random Forest, Linear SVC, and Naïve Bayes Classifier using TF-IDF as the data processing technique. The results obtained from the classification process using various proportions of training documents, ranging from 10% to 90%, are documented in Table 6.

Table 6. Accuracy Results and Comparison with Other Methods

No	Algorithms	Documents Proportion	Total of Documents	Result		
				Precision	***Recall***	***Accuracy***
1	RCNN	90-10%	7008	76,32%	79,12%	84,2%
2	NAÏVE BAYES	90-10%	7008	77,59%	81,07%	82,5%
3	*WORD2VEC - ELM*	90-10%	7008	**86%**	**87%**	**86,31%**

From Table 6, it can be seen that the proportion of training documents that yielded the highest accuracy is 80%. With an accuracy result of 86.31%. It can be observed that Word2vec, Fasttext, and ELM produce the highest accuracy.

b. Confusion matrix.

The confusion matrix is a useful tool to assess how well a classifier identifies tuples from different classes. The results of the confusion matrix in this study are illustrated in Figure 5.

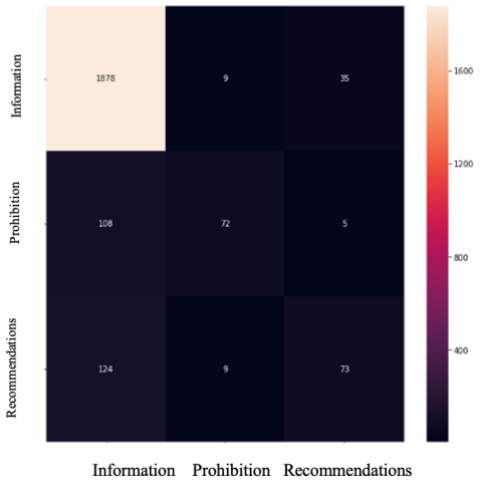

Figure 5. *Confussion Matrix*

Figure 5 explains the results of the confusion matrix from the proposed model using Python tools. The confusion matrix illustrates the identification errors made by the designed machine. The x-axis represents the labels of the actual classes, and the y-axis represents the predicted classes. Based on the figure, there are 124 hadith data that should be identified as the information class, but the designed machine incorrectly identifies them as recommendations, 108 as prohibitions, and correctly predicts 1878 hadith as the information class. Furthermore, for the actual prohibition class data, there are 9 instances wrongly predicted as recommendation hadith and 9 hadith wrongly predicted as information, while 72 hadith are correctly detected as prohibitions. Lastly, for the recommendation class hadith, there are 5 hadith wrongly identified as prohibitions and 35 hadith wrongly identified as information, whereas 35 hadith are correctly identified according to their actual class, which is the recommendation class. Based on the figure, it can be concluded that the proposed machine does not have significant errors in the identification process and has been able to maximize identification and classification.

4. CONCLUSION

The author's discussion on the Classification of Bulughul Maram into Prohibition, Recommendation, and Information Categories using Extreme Learning Machine and Fasttext has been completed. The conclusions of this research that this research successfully assisted in classifying hadiths based on categories

of prohibitions, recommendations, and commands using Python technology in the field of text mining and the classification using Extreme Learning Machine and Word2Vec resulted in higher accuracy than previous research using RCNN and Naive Bayes. This research achieved an accuracy of 86.31%.

ACKNOWLEDGEMENTS

This research undoubtedly has many limitations and would not have progressed smoothly without the support of all parties involved. The authors express their gratitude to the hadith experts who have contributed to the successful execution of this study, in line with the initial goal, "Jazakallah." Additionally, we would like to extend our thanks to the Indonesian government for providing Operational Assistance to State Universities (BOPTN) to the author's team.

REFERENCES

[1] H. Anam, M. A. Yusuf, and S. Saada, "Kedudukan Al-Quran Dan Hadis Sebagai Dasar Pendidikan Islam," *Al-Tarbawi Al-Haditsah jurnal pendidikan agama islam* , vol. 7, no. 22, pp. 1–17, Dec. 2022, doi: 10.32939/ishlah.v1i2.46.

[2] D. Darussamin, *KULIAH ILMU HADIS*, KALMEDIA. RIAU: KALMIEDIA, 2020.

[3] H. Noor, "KITAB HADIS DI PESANTREN: BIOGRAFI KITAB AL-BULUGH AL-MARAM," *Jurnal Pendidik Islam Bahsun Ilmy*, vol. 01, pp. 1–10, 2020, [Online]. Available: https://irtaqi.net/2016/09/01/introduksi-kitab-bulughul-maram-ii/

[4] S. M. Pd, E. Rahwanto, S. Kom, and R. Komala, "E-COMMERCE DORONG PEREKONOMIAN INDONESIA, SELAMA PANDEMI COVID 19 SEBAGAI ENTREPRENEUR MODERN DAN PENGARUHNYA TERHADAP BISNIS OFFLINE," *JURNAL MANAJEMEN DANBISNIS (JUMANIS) PRODI KEWIRAUSAHAAN*, vol. 02, no. 2686–5939, pp. 111–124, 2020, doi: 10.47080.

[5] S. S. Ummah, "DIGITALISASI HADIS (Studi Hadis di Era Digital)," *Diroyah: Jurnal Ilmu Hadis 4*, vol. 1, pp. 1–10, 2019.

[6] D. Marini, U. Atmaja, and R. Mandala, "Analisa Judul Skripsi untuk Menentukan Peminatan Mahasiswa Menggunakan Vector Space Model dan Metode K-Nearest Neighbor," *IT FOR SOCIETY*, vol. 04, no. 2503–2224, pp. 1–10, 2019.

[7] 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, 2021, doi: 10.25126/jtiik.202183750.

[8] I. F. Rozi, M. H. Ratsanjani, and R. A. Afandi, "KLASIFIKASI ANJURAN, LARANGAN, DAN INFORMASI PADA HADITS MENGGUNAKAN NAIVE BAYES CLASSIFIER," in *SEMINAR INFORMATIKA APLIKATIF POLINEMA (SIAP)*, 2020, pp. 303–310.

[9] D. Tania Ananda Paramitha, I. Cholissodin, and C. Dewi, "Prediksi Rating Otomatis Berdasarkan Review Restoran pada Aplikasi Zomato dengan menggunakan Extreme Learning Machine (ELM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 4687–4693, pp. 4687–4693, 2019, [Online]. Available: http://j-ptiik.ub.ac.id

[10] D. A. Sani and M. Z. Sarwani, "Koreksi Jawaban Esai Berdasarkan Persamaan Makna Menggunakan Fasttext dan Algoritma Backpropagation," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 2, pp. 92–111, Aug. 2022, doi: 10.23887/janapati.v11i2.49192.

[11] A. Firdaus and W. I. Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan)," 2021.

[12] A. Y. Muniar, Pasnur, and K. R. Lestari, "PENERAPAN ALGORITMA K-NEAREST NEIGHBOR PADA PENGKLASIFIKASIAN DOKUMEN BERITA ONLINE," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 10, no. 2, pp. 137–144, 2020.

[13] S. Ulya, A. Ridwan, W. Cholid Wahyudin, F. Maisa, and H. D. Ab, "TEXT MINING SENTIMEN ANALISIS PENGGUNA APLIKASI MARKETPLACE TOKOPEDIA BERDASAR RATING DAN KOMENTAR PADA GOOGLE PLAY STORE," 2022.

[14] M. Afdal and L. R. Elita, "PENERAPAN TEXT MINING PADA APLIKASI TOKOPEDIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 8, no. 1, pp. 78–87, 2022.

[15] N. A. Izati, B. Warsito, and T. Widiharih, "PREDIKSI HARGA EMAS MENGGUNAKAN FEED FORWARD NEURAL NETWORK DENGAN METODE EXTREME LEARNING MACHINE," *JURNAL GAUSSIAN*, vol. 8, no. 2339–2541, pp. 171–183, 2019, [Online]. Available: http://ejournal3.undip.ac.id/index.php/gaussian

[16] A. N. Alfiyatin, W. F. Mahmudy, C. F. Ananda, and Y. P. Anggodo, "Penerapan Extreme Learning Machine (ELM) untuk Peramalan Laju Inflasi di Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 2, p. 179, 2019, doi: 10.25126/jtiik.201962900.

[17] D. Azzahra Nasution, H. H. Khotimah, and N. Chamidah, "PERBANDINGAN NORMALISASI DATA UNTUK KLASIFIKASI WINE MENGGUNAKAN ALGORITMA K-NN," *CESS (Journal of Computer Engineering System and Science)*, vol. 4, no. 2502–7131, pp. 2502–7131, 2019.

[18] S. C. Nayak and B. B. Misra, "Extreme learning with chemical reaction optimization for stock volatility prediction," *Financial Innovation*, vol. 6, no. 1, Dec. 2020, doi: 10.1186/s40854-020-00177-2.

[19] H. Mukhtar, R. Gunawan, A. Hariyanto, Syahril, and Wide Mulyana, "Peramalan Kedatangan Wisatawan ke Suatu Negara Menggunakan Metode Support Vector Machine (SVM)," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 274–282, Dec. 2022, doi: 10.37859/coscitech.v3i3.4211.

[20] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.

[21] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 24–33, 2021, doi: 10.57152/malcom.v1i1.20.

[22] S. Kritik Matan Hadits Al and A. Diana Devi UIN Sunan Kalijaga Yogyakarta, "Studi Kritik Matan Hadits," *Desember Tahun*, vol. 14, no. 2, pp. 293–312, 2020, doi: 10.24042/digunakan.