lucky

by dts digital

Submission date: 27-Aug-2024 08:00AM (UTC+0700)
Submission ID: 2438752842
File name: 8732-23650-1-SM.docx (302.4K)
Word count: 2205
Character count: 12576

JIKO (Jurnal Informatika dan Komputer)
Vol. x, No. x, April 2024, pp. x-x
DOI: 10.33387/jiko

Accredited KEMENDIKBUD RISTEK, No.105/E/KPT/2022
p-ISSN: 2614-8897
e-ISSN: 2656-1948

OPTIMIZING HADITH CLASSIFICATION WITH NEURAL NETWORKS: A STUDY ON BUKHARI AND MUSLIM TEXTS

Rasenda[1], Luky Fabrianto[2], Novianti Madhona Faizah[3]

Universitas Teknologi Muhammadiyah Jakarta[1]
Universitas Nusa Mandiri[2]
Universitas Tama Jagakarsa[3]

*Email: [1]rasenda@utmj.ac.id, [2]luky.lfb@nusamandiri.ac.id, [3]novianti@jagakarsa.ac.id

(Received: dd mmm yyyy, Revised: dd mmm yyyy, Accepted: dd mmm yyyy)

Abstract

In the book of Bukhari and Muslim hadith there are 7008 hadith sentences, of the 7008 sentences the Hadith is not yet known a hadith included in the category of prohibitions or orders. By doing the classification, it will be easier for readers to understand the hadith. The classification of hadiths is done in several stages, including: pre-processing text, the use of word vector features, and modeling of neural network architecture with multilayer perceptron. The use of layers in neural networks and feature extraction with word vectors has proven to provide good results for the classification of hadiths. The results showed a fairly high degree of accuracy that is equal to 97.72% by using two layers and 256 neurons, making this study can be used for very good classification of hadiths.

Keywords: Classification, Hadith, Multilayer Perceptron, Neural Network, Word Vector.

Corresponding Author: Author1

1. INTRODUCTION (UPPERCASE, 10pt, bold)

Text classification research on Hadith data is still relatively rare, even though Hadith data is now easily accessible through the internet or various applications. Processing Hadith data is crucial as it contains behavioral guidelines for Muslims in accordance with the Sunnah of Prophet Muhammad Shallallahu 'Alaihi Wasallam [1]. In the context of Hadith text processing, a command is referred to as "amar," and a prohibition is referred to as "nahi." According to the majority of usul al-fiqh scholars, "amar" is a command issued by a higher authority to a lower one, while "nahi" is a prohibition issued by a higher authority to a lower one [2].

Natural Language Processing (NLP) using WordVector has proven to be effective in encoding word relationships in vector space, which is beneficial for various NLP tasks [3]. One commonly used method in NLP is Neural Network, which mimics the workings of the human neural network for pattern recognition and classification [4]. In designing a Neural Network, it is crucial to determine the learning method, which involves updating synaptic weights based on input signals and the expected output. Neural Networks typically consist of several neurons as information processing units [5]. The classification model evaluation uses a confusion matrix that shows the actual number of cases from the observed class to be predicted [6].

Several studies related to text mining on Hadith show various methods and accuracy results. The study by [7], titled "Application of Particle Swarm Optimization on Feedforward Neural Network for Classification of Bukhari Hadith Text in Indonesian Translation," used BP-FNN and PSO-FNN methods on a Hadith classification dataset, achieving BP-FNN accuracy of 88.57% and PSO-FNN of 89.5%. [8] In their study "Classification of Suggestions, Prohibitions, and Information in Sahih Al-Bukhari Hadiths Based on Unigram Model Using Artificial Neural Network (ANN)" used n-gram and ANN models, yielding an f1-score of 85% for the categories of suggestions, prohibitions, and information. [9] In "Classification of Hadith Authenticity Based on Hadith Narrators Using Principal Component Analysis

(PCA) and Backpropagation Neural Network (BPNN)," PCA and BPNN were used on the categories sahih, hasan, and dha with an accuracy of 86.53%. The study [10] titled "Multi-Label Topic Classification of Hadith of Bukhari (Indonesian Language Translation) using Information Gain and Backpropagation Neural Network" used BPNN on the categories of suggestions, prohibitions, and commands with an accuracy of 88.42%. Meanwhile, [11] in "Classification of Hadith into Positive Suggestion, Negative Suggestion, and Information" used TF-IDF, Baseline, SVM, and ANN on the categories of commands, prohibitions, and information, achieving a Baseline f1-score of 69% and ANN 79%. These studies illustrate the use of various techniques in the classification of Hadith texts with varying results.

Based on previous research, neural network algorithms have proven effective in classifying Hadith texts. Neural networks are considered superior due to their non-linear nature, making them suitable for high-complexity problems, and their adaptability, allowing them to effectively map inputs to outputs. Additionally, neural networks are fault-tolerant, continuing to function even with some degree of error, and capable of generalization, processing new data based on learned experience [4][12]. Therefore, this study will use neural networks to classify Hadiths narrated by Bukhari and Muslim into categories of prohibitions or commands.

2. RESEARCH METHOD

Here are the research stages, as shown in Figure 1 below.

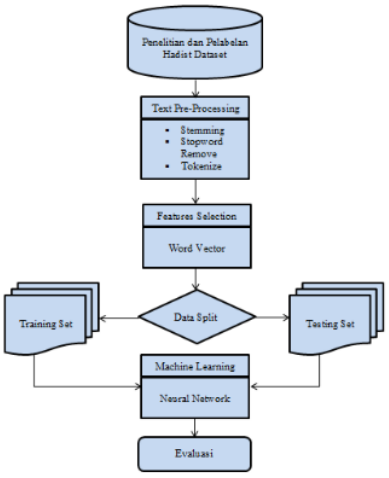

Figure 1. Research stages

This research utilizes a dataset of Hadith from Shahih Bukhari and Muslim, consisting of 7008 matan, labeled according to the classification of either prohibition or command. During the pre-processing stage, the StemmerFactory library from the Sastrawi module in Python 3.7 was used. The pre-processing process involves three stages: stemming, stopword removal, and tokenization. Stemming was performed using the StemmerFactory method from the Sastrawi library, executed on sentences with the create stemmer function. Stopwords were removed using StopWordRemoverFactory from the Sastrawi library, executed with create stop word remover. Tokenization was done using word_tokenize from the NLTK library.

The feature used is the word vector from the Gensim library in Python 3.7, which was employed to calculate vocabulary and similarity values from the Hadith dataset. Word vectors were then used for one-hot encoding to convert text into binary values to be input into the neural network. Data was split into training and testing sets using the scikit-learn library. The neural network model was built using the Keras library in Python 3.7 by adding layers with the dense function. Input was determined based on the similarity size in the word vector, with the relu activation function used for the input and hidden layers, and the Sigmoid activation function for the output layer. The Adam optimizer was employed with binary cross entropy as the loss function, and the epoch and batch size values were set accordingly. Figure 2 illustrates the architecture for building the neural network.

Input layer (Dense)	→	*Aktivasi Relu*
Size Input	→	*Size Similiarity in Word Vector*
Hidden layer (Dense)	→	*Aktivasi Relu*
Output layer (Dense)	→	*Aktivasi Sigmoid*
Optimizer	→	*Adam*
Loss	→	*Binary Crossentropy*
Epoch	→	*Value Epoch*
Batch Size	→	*Value Batch Size*

Figure 2. Developing Neural Network

The final stage of this research involves evaluating the experiment using Confusion Matrix, as shown in Table 1. The accuracy can then be calculated using the formula provided below [13]:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \; x\; 100\ \% \qquad (1)$$

Tabel 1. *Confusion Matrix*

	Actual Positive Class	Actual Negative Class
Predictive Positive Class	True Positive (TP)	False Negative (FN)
Predictive Negative Class	False Positive (FP)	True Negative (TN)

Keterangan : TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

3. RESULT AND DISCUSSION

An example of the pre-processing process involves a Hadith that reads, "All actions are judged by intentions, and each person will be rewarded

according to what they intended. Whoever migrates for worldly gain or to marry a woman, his migration is for that which he intended." After undergoing stemming using the Sastrawi library, stopword removal, and tokenization with the NLTK library, the Hadith is transformed as shown in Figure 3 below.

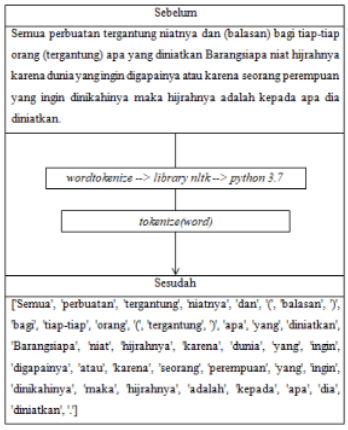

Figure 3. Tokenize result

The Wordvector used in this research was configured with a min_count of 5, resulting in a vocabulary of 3,393 words. Figure 4 shows a visualization of the vocabulary after setting min_count to 1000, which reduces the vocabulary to 65 words.

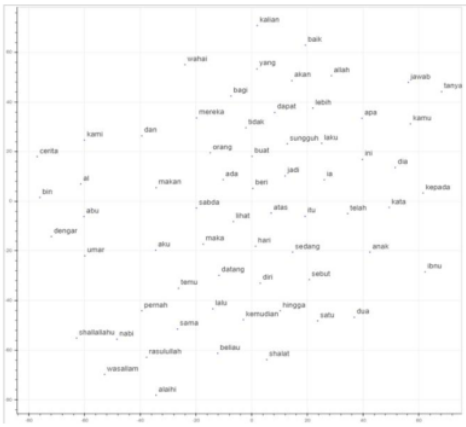

Figure 4. Visualization of Vocabulary Results

The Wordvector used to find similarity in this research was configured with windows = 5 and size = 200. Table 2 provides an example of the similarity results for the word "hari" (day). Figure 5 illustrates the visualization of the similarity for the word "hari."

Tabel 2. Similiarity word 'Hari' result

-0.998	-0.15458	-1.60573	-1.92564	0.171547
0.2764	1.009715	-0.74964	-0.66444	1.068439
0.4179	-0.04054	-0.70896	0.424856	1.530188
.....	.....	.....	.....	.....
0.8793	-0.12627	-0.93852	-0.77086	-1.67769
2.7103	0.79881	1.67535	0.608546	0.101442
-0.178	-1.23129	0.161919	1.066376	-0.04765

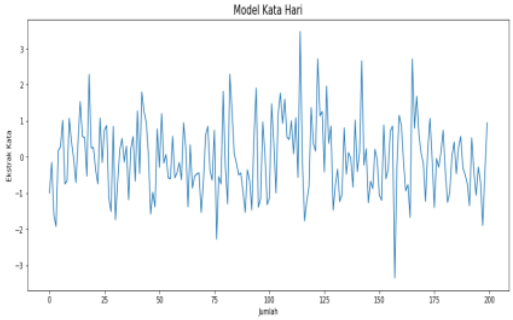

Figure 5. Visualisasi Similiarity 'Hari' result

The top 5 most similar words to "hari" are shown in Table 3, and their visualization is presented in Figure 6.

Tabel 3. *Most_similiarity* Top 5 *word 'hari' result*

Word	Most_Similiarity
pagi	0.453531414
bulan	0.453418076
siang	0.450212121
raya	0.431948364
malam	0.430161893

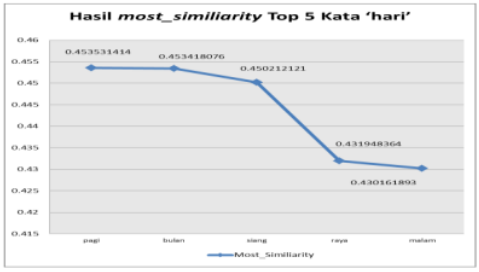

Figure 6. Visualization of similiarity 'hari' result

Next is an example of the top 5 most similar words to "hari" + "islam" - "kiamat," as shown in Table 4. The visualization of these top 5 most similar words is presented in Figure 7.

Tabel 4. Most_similiarity
Top 5 'hari'+ 'islam' – 'kiamat' result

Word	Most_Similiarity
masa	0.379982561

talak	0.372045457
baiat	0.36673528
unjung	0.362578362
pulang	0.34503299

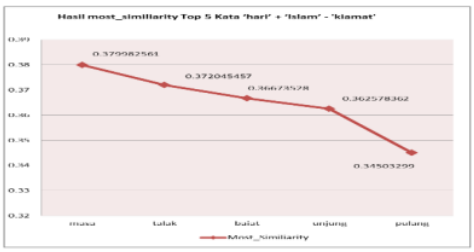

Figure 7. Visualization 'hari' + 'islam'– 'kiamat' result

Data was divided using cross-validation, consisting of 90% training data (6307 samples) and 10% testing data (701 samples) from a total of 7008 samples. The neural network type used is a multilayer perceptron, which is effective for classification by constructing several simple yet effective layers. The model employed is a sequential model with dense layers for adding layers. The input size is 200 derived from feature selection, with ReLU as the activation function for the first and intermediate layers, and Sigmoid for the output layer. The optimizer used is Adam, with binary cross entropy as the loss function, 25 epochs, and a batch size of 1024. The parameters adjusted include the number of layers and neurons in the hidden layers to achieve the best accuracy. Table 5 summarizes the accuracy results from three different neural network experiments.

Tabel 5. Experiment Result Recapitulation

Experiment	*Layer*	*Neurons*	Accuracy
A	2	128	0.9757
	5	128	0.9714
	10	128	0.9643
B	**2**	**256**	**0.9772**
	5	256	0.9686
	10	256	0.9728
C	2	512	0.9757
	5	512	0.9757
	10	512	0.9715

The confusion matrix results from Experiment B are shown in Table 6 below.

Tabel 6. Confusion Matrix result

	Actual 0	Actual 1
Pred. 0	22 (TN)	13 (FP)
Pred. 1	3 (FN)	663 (TP)

From the confusion matrix results, the accuracy calculated is 97.72%. Figure 8 illustrates the visualization of the accuracy history from this research.

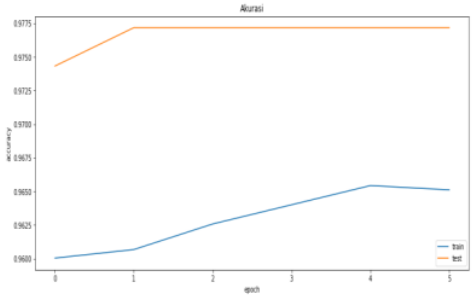

Figure 8. Accuracy History result

The accuracy results show that the training accuracy is lower than the testing accuracy, indicating that the model is not overfitting and performs well on the test data.

4. CONCLUSION

The conclusions of this research indicate that effective text pre-processing yields optimal results in word analysis within sentences. The use of word vectors as feature selection aids the neural network in extracting the meaning of sentences. Neural networks are effective for classifying Hadith into categories of prohibition or command. Experiment B architecture, which employs two layers and 256 neurons in the hidden layer, achieved the highest accuracy of 97.72%.

5. REFERENCE

[1] D. Sholehat and D. A. Alfiani, "Pengaruh Pola Asuh Permisif dari Orang Tua Terhadap Prestasi Belajar Siswa Pada Mata Pelajaran Al-Qur'an Hadits Kelas IV di Madrasah Ibtidaiyah," *Indones. J. Elem. Educ.*, vol. 1, no. 1, pp. 1–12, 2019.

[2] H. M. Abdelaal, B. R. Elemary, and H. A. Youness, "Classification of Hadith According to Its Content Based on Supervised Learning Algorithms," *IEEE Access*, vol. 7, pp. 152379–152387, 2019, doi: 10.1109/ACCESS.2019.2948159.

[3] S. Park, J. Byun, S. Baek, Y. Cho, and A. Oh, "Subword-level word vector representations for Korean," *ACL 2018 - 56th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long*

Pap., vol. 1, pp. 2429–2438, 2018, doi: 10.18653/v1/p18-1226.

[4] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[5] M. W. Habibi, "Permainan edukatif puzzle surat Al-Qur'an (Q-Puzzle) menggunakan backpropagation sebagai penentu level permainan," Jun. 2018.

[6] I. Dwita, S. Tarigan, R. Habibi, R. Nuraini, and S. Fatonah, "Evaluasi Algoritma Klasifikasi Machine Learning Kategori Nilai Akhir Tunjangan Kinerja Pegawai," *J. Sist. Cerdas*, vol. 6, no. 3, pp. 251–261, Dec. 2023, doi: 10.37396/JSC.V6I3.246.

[7] M. Ghufran, A. Adiwijaya, and S. Al-Faraby, "Penerapan Particle Swarm Optimization Pada Feedforward Neural Network Untuk Klasifikasi Teks Hadis Bukhari Terjemahan Bahasa Indonesia," *J. Media Inform. Budidarma*, vol. 2, no. 4, p. 165, 2018, doi: 10.30865/mib.v2i4.951.

[8] A. Y. Prathama *et al.*, "Klasifikasi Anjuran , Larangan dan Informasi pada Hadis Sahih Al-Bukhari berdasarkan Model Unigram menggunakan Artificial Neural Network (ANN)," *J. Teknosains*, vol. 2, no. 3, pp. 1130–1139, 2018, [Online]. Available: http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2573%0Ahttp://eprints.undip.ac.id/35760/.

[9] U. Nuha and N. Rochmawati, "Klasifikasi Kesahihan Hadits Berdasarkan Perawi Hadits Menggunakan Principal Component Analysis (PCA) dan Backpropagation Neural Network (BPNN)," *J. Informatics Comput. Sci.*, vol. 1, no. 03, pp. 138–143, Jan. 2020, doi: 10.26740/JINACS.V1N03.P138-143.

[10] M. Y. Abu Bakar, Adiwijaya, and S. Al Faraby, "Multi-Label Topic Classification of Hadith of Bukhari (Indonesian Language Translation)Using Information Gain and Backpropagation Neural Network," *Proc. 2018 Int. Conf. Asian Lang. Process. IALP 2018*, pp. 344–350, Jul. 2018, doi: 10.1109/IALP.2018.8629263.

[11] S. Al Faraby, E. R. R. Jasin, A. Kusumaningrum, and Adiwijaya, "Classification of hadith into positive suggestion, negative suggestion, and information," *J. Phys. Conf. Ser.*, vol. 971, no. 1, p. 012046, Mar. 2018, doi: 10.1088/1742-6596/971/1/012046.

[12] T. S. Gunawan *et al.*, "Development of video-based emotion recognition using deep learning with Google Colab," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 18, no. 5, pp. 2463–2471, Oct. 2020, doi: 10.12928/TELKOMNIKA.V18I5.16717.

[13] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf. Sci. (Ny).*, vol. 507, pp. 772–794, Jan. 2020, doi: 10.1016/J.INS.2019.06.064.

lucky

ORIGINALITY REPORT

15%	13%	10%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ejournal.unkhair.ac.id Internet Source	3%
2	www.nature.com Internet Source	2%
3	123dok.com Internet Source	1%
4	Submitted to Universitas Khairun Student Paper	1%
5	jtiik.ub.ac.id Internet Source	1%
6	Submitted to Universitas Amikom Student Paper	1%
7	core.ac.uk Internet Source	1%
8	Jingjing Zhao, Xiaoming Wang. "Visual Pattern Recognition: Identity Authentication Technology Based on Eye Movement Patterns", Journal of Physics: Conference Series, 2022 Publication	1%

9	link.springer.com Internet Source	1 %
10	Rahmadi Yotenka, Sekti Kartika Dini, Achmad Fauzan, Atina Ahdika. "Exploring the relationship between hadith narrators in Book of Bukhari through SPADE algorithm", MethodsX, 2022 Publication	1 %
11	doi.nrct.go.th Internet Source	1 %
12	elibrary.nusamandiri.ac.id Internet Source	1 %
13	insightsociety.org Internet Source	1 %
14	jurnal.iaii.or.id Internet Source	1 %
15	acikerisim.cumhuriyet.edu.tr Internet Source	<1 %
16	medinform.jmir.org Internet Source	<1 %
17	repository.universitasbumigora.ac.id Internet Source	<1 %
18	"Advances in Robotics, Automation and Data Analytics", Springer Science and Business Media LLC, 2021	<1 %

Publication

19	Ni Wayan Sumartini Saraswati, I Putu Krisna Suarendra Putra, I Dewa Made Krishna Muku, Gede Dana Pramitha. "Support Vector Machine For Hoax Detection", SINTECH (Science and Information Technology) Journal, 2023 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		