

Bidirectional long-short term memory and conditional random field for tourism named entity recognition

Annisa Zahra¹, Ahmad Fathan Hidayatullah², Septia Rani²

¹Undergraduate Program in Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta, Indonesia

²Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta, Indonesia

Article Info

Article history:

Received Nov 10, 2021

Revised Apr 20, 2022

Accepted May 19, 2022

Keywords:

Bidirectional long-short term memory

Conditional random field Feature

Named entity recognition

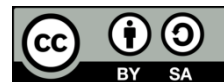
Tourism article

Word2vec

ABSTRACT

The common thing to do when planning a trip is to search for a tourist destination. This process is often done using search engines and reading articles on the internet. However, it takes much time to search for such information, as to obtain relevant information, we have to read some available articles. Named entity recognition (NER) can detect named entities in a text to help users find the desired information. This study aims to create a NER model that will help to detect tourist attractions in an article. The articles used for the dataset are English articles obtained from the internet. We built our NER model using bidirectional long-short term memory (BiLSTM) and conditional random fields (CRF), with Word2Vec as a feature. Our proposed model achieved the best with an average F1-Score of 75.25% compared to all scenarios tested.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ahmad Fathan Hidayatullah

Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia

Kaliurang St. km. 14.5, Sleman, Yogyakarta, Indonesia

Email: fathan@uii.ac.id

1. INTRODUCTION

Tourism growth happens almost every year. In 2019, worldwide international tourist arrivals reached 1.5 billion, increasing by 4% from the previous year [1]. With the advent of the web 2.0 age, internet users frequently share their travel experiences via websites [2]. A study conducted by Google Travel found that 74% of travelers plan their trips via the internet [3]. The search for tourist destinations is one of the steps that is generally carried out when planning a trip. Most people commonly search for information regarding tourism destinations through reviews, websites, and articles on the internet [4], [5]. However, searching, selecting, and reading details of each piece of information through travel guidebooks or portal sites is time-consuming [6], [7]. The time-consuming issue of getting travel information from texts can be solved by applying information extraction.

The named entity recognition (NER) task can be applied to extract information from texts in the natural language processing area. NER can be defined as a task of extracting entities from text documents, such as a person's name, location, or organization [8]–[10]. Various NER approaches are Rule-Based, Machine Learning which includes hidden markov model (HMM), maximum entropy, decision tree, support vector machines, conditional random fields (CRF), and Hybrid Approaches [11]. There are also other methods, such as recurrent neural network (RNN) and its variant, long short-term memory (LSTM) which has been successfully used in various prediction problems sequences, such as NER, language modeling, and speech recognition [12]. Many researchers have also studied NER in multiple fields, such as the geological domain [13]. They developed a NER system for geological text using the CRF method and IITKGP-GEOCORP dataset developed from article collections and scientific reports containing geology-

related information in India. In the biomedical domain, the BiLSTM-CRF (bidirectional long-short term memory-conditional random fields) model was used to recognize drug names in biomedical literature [14]. They used the concatenation of word embedding and character level embedding for the input vector. Their experiments obtained an F1-Score of 92.04% and achieved comparable performances with the state-of-the-art results on two datasets. Word embedding was used as a feature by [15] in named entity recognition research and produced a better result than the baseline without word embedding.

A study by [16] proposed BiLSTM-CRF for their NER system using three datasets, Penn Treebank POS Tagging, conference on computational natural language learning (CoNLL) 2000 chunking, and CoNLL 2003 named entity tagging. The research revealed that the BiLSTM-CRF model could efficiently use input features from the past and future because of the two-way LSTM component. Moreover, the CRF layer can provide the label information at the sentence level. From several scenarios used in that research, the BiLSTM-CRF model yielded the best results in almost all datasets. The merging of BiLSTM and CRF layers was also carried out in [17], and the results showed that such merging could solve the problem of the inability to handle strong dependencies of tags in a sequence. A Chinese dataset and BiLSTM-CRF model were used in [18]. It was found that using a dictionary produced better performance than using only the BiLSTM-CRF model. The work in [19], which used the BiLSTM-CRF model along with pre-trained word embeddings, character embeddings, and dictionary information, succeeded in improving the performance of the Disease-NER system. They used pre-trained word embeddings using skip-gram that combined domain-specific text (PMC and PubMed texts) with generic text (english wikipedia dump). The BiLSTM-CRF model can also be combined with bidirectional encoder representations from transformers (BERT) [20]. The combination of the three methods is called BBLC. They also compared the performance of the BBLC model with the BiLSTM-CRF model using the same dataset. As a result, the BBLC obtained a higher F1-Score than the BiLSTM-CRF on some entities, such as location, organization, and thing. On the other hand, the BiLSTM-CRF model achieved a higher F1-Score than BBLC on the time entity.

Furthermore, NER can extract meaningful information from tourism websites by identifying the named entities. In the tourism domain, identified entities can be the names of tourist attractions, places of lodging, facilities, and locations. Identifying related entities is expected to make it easier for potential tourists to find tourist destinations via the internet. However, many NER studies in the tourism domain did not focus on categorizing the characteristics of tourist attractions. We argue that classifying the characteristics of the tourist attraction, such as natural, heritage, or purposefully built, is essential to help users make decisions for future utilization of our NER system. Thus, in this study, we present a NER system to aid tourists in finding tourist destinations from articles by extracting tourist attractions into four categories such as “natural”, “heritage”, “purposefully built” (artificial), and “outside”. This study proposes a combination of Word2vec and BiLSTM-CRF approaches to building the NER system for tourism. The implementation of Word2vec in our research is inspired by [21]. Additionally, we explore the performance of Word2vec using two different Word2vec algorithms called skip-gram and continuous bag-of-words (CBOW).

There is a previous study with different tags/labels for tourism NER. Saputro *et al.* [22] used five labels, namely “nature”, “place”, “city”, “region”, and “negative” as the named entity. The proposed system has scored 70.43% of accuracy, with an F-Score of 69%. However, some of the labels used in this study are not specific enough to identify the characteristics of the tourist attractions. For example, when our goal is to extract the name of a tourist attraction and its characteristics (whether they are natural, heritage, or purposefully built), the labels “city” and “region” are too broad. Another study proposed a corpus for tourism NER in the Mongolian language [23]. Thus, although the studies are similar, direct comparison is impossible because the labels and the language are different.

The rest of this paper is structured as follows. Section 2 describes the methodology of our study. In section 3, the result and discussion of this study are presented. Finally, section 4 provides the conclusion and future work.

2. METHOD

This section provides the methodology conducted in our study. Our work consists of six steps such as data retrieval, pre-processing, data labeling, feature extraction, classification, and evaluation. A more detailed explanation of each step is described later.

2.1. Data retrieval

The data for this study was gathered via web scraping techniques from English tourism articles. We searched the articles using two distinct methods: keyword searches and scraping articles directly from predetermined websites. The articles were indexed using the following keywords: *top tourist cities*, *best places to visit*, *top world heritage sites*, *world heritage list*, *best destinations for nature lovers*, and *best natural tourist attractions*. Scraping was accomplished using Python's Newspaper module, which is capable

of extracting and parsing text from website articles. Each article was processed independently and then saved in a file with the *.txt extension. We gathered our dataset from 24 websites and obtained 92 articles containing 8,500 sentences, 17,137 unique words, and 183,507 tokens.

2.2. Pre-processing

This pre-processing task aims to eliminate all meaningless characters and preserve the remaining valuable words [24]. The pre-processing techniques carried out in this research were removing URLs, emoticons, and tokenization. URLs and emoticons were removed because they do not significantly affect recognizing a named entity. Moreover, tokenization was done by dividing the sentences into smaller units called tokens.

2.3. Data labelling

We manually labeled the tourist attractions into four categories: natural, heritage, purposefully built (artificial), and outside. We adopted the BIO tagging format in the labeling process, where a token is tagged as B-label if it is the beginning of a named entity, I-label if it is within a named entity but not the first, and O-label represents otherwise [25]. In this study, the B-prefix represents the first word of a tourist attraction's name, while the I-prefix represents the second through the last word of a tourist attraction's name. We considered natural products to be labeled as a natural category in this study, which is open to the public and provides natural views such as waterfalls, mountains, caves, rivers, and glaciers. Tourist attractions that have been around for a long time are ancient, historic, and often cultural, or tend to represent culture and heritage, and places of worship fall into the heritage category. Ruins, monuments, temples, forts, castles, mosques, and cathedrals are categorized as heritage structures. Tourist attractions that are purposefully built to attract visitors, such as museums, markets, and amusement parks, will be classified as purposefully built. The final category, outside, is for a word that is not considered a tourist attraction.

The number of tokens for each label is shown in Table 1. Our dataset was imbalanced since the label O dominated our dataset with 171,728 tokens. The I-PURPOSE and I-HERITAGE labels consist of 2,874 and 2,051 tokens, respectively. The number of B-NATURAL, I-NATURAL, and B-PURPOSE were almost the same. The number B-HERITAGE label was the lowest, with 1,401 tokens.

Table 1. Number of tokens for each label

Label	Number of Tokens
O	171,728
B-NATURAL	1,789
I-NATURAL	1,853
B-HERITAGE	1,401
I-HERITAGE	2,051
B-PURPOSE	1,811
I-PURPOSE	2,874

2.4. Feature extraction using word embeddings

This study applied word embeddings called Word2vec as the feature extraction method. Word2vec has two different approaches, continuous bag-of-words (CBOW) and skip-gram. In our experiments, we compared the application of CBOW and skip-gram to obtain the best result. The CBOW algorithm predicts a target word based on its context, whereas the Skip-gram algorithm predicts the target context based on a word. The CBOW model aims to predict the middle word by combining the representation of the surrounding words. The Skip-gram model generates a word vector representation capable of predicting the context of the word.

Additionally, both models require little training time and can be applied to a large corpus [26]. The textual input is converted to vectors using Word2Vec and then trained to generate a dictionary. The dictionary contains the same number of words as the dataset's unique words, and each word has its vector. In the following step, the dictionary containing the pre-trained vectors will be used as weights for the embedding layer.

2.5. Classification

We used StratifiedKfold from Scikit-Learn to divide our data into ten subsets. One out of ten subsets were used as the test data, and the other subsets acted as the train data. The total sample in this study was 8,500, and the total sample is the total sentences in the dataset. After splitting the data, we have 7,650 train data and 850 test data. During training, we used 20% of the train data to be used as validation data, so

that the train data ended up being 6,120 samples and the validation data was 1,530 samples. The average number of tokens used for each label in each fold for training data is shown in Table 2, and for test data is presented in Table 3. A large number of O labels result from the post padding sequences performed on each sample with O as the padding value.

Table 2. The average number of tokens for each label in each fold (training data)

Label	Average Number of Tokens
O	678,186
B-NATURAL	1,602
I-NATURAL	1,659
B-HERITAGE	1,233
I-HERITAGE	1,803
B-PURPOSE	1,554
I-PURPOSE	2,459

Table 3. The average number of tokens for each label in each fold (test data)

Label	Average Number of Tokens
O	75,354
B-NATURAL	178
I-NATURAL	184
B-HERITAGE	137
I-HERITAGE	200
B-PURPOSE	172
I-PURPOSE	273

The input data were classified into predefined categories by combining two methods, BiLSTM and CRF. BiLSTM is made up of two LSTMs, forward LSTM and backward LSTM. As a modification of RNN, LSTM has a memory cell that can store information for a long period. When dealing with long sequential data, the vanishing gradient problem encountered in RNN can be addressed with LSTM by utilizing gates that control the information entering the memory [27].

LSTM is useful to be applied in sequential labeling cases since its capability to gain the information from both front and back sides of the texts. The hidden state in the LSTM, on the other hand, only retrieves information from the previous part, leaving the next part unknown. To overcome these issues, BiLSTM can be applied [28]. In BiLSTM, the combination of forward LSTM and backward LSTM will capture information from both directions. The output of forward and backward LSTM then be combined using the sigmoid function (σ) as shown in (1). It can be a concatenation function, an addition function, an average function, or a multiplication function. The y_t represents the output at time t , while \vec{h} represents the hidden state from the forward layer and \overleftarrow{h} represents the hidden state from the backward layer.

$$y_t = \sigma(\vec{h}, \overleftarrow{h}) \quad (1)$$

The proposed BiLSTM-CRF architecture in this research is depicted in Figure 1. The input layer receives the words. These words are represented by a vector of integer values. Each word's value was generated using the pre-trained word2vec. Additionally, the embedding layer's output serves as the input for the following BiLSTM layer. Moreover, a decision function based on the CRF layer was used to generate the label sequence. CRF is a method to obtain global optimum predictions using a conditional probability distribution model [29]. The CRF layer labels the sequence using the surrounding labels. The labels preceding and following the current word can aid in predicting labels for the current word. There are two types of scoring calculations in the CRF method: emission and transition scores. The emission scores in this model are derived from the output score matrix for the preceding layer. While transition scores are initially assigned randomly, they will be updated throughout the training process. The two scores will be used to predict the final output sequence of labels.

2.6. Evaluation

To measure the performance of NER, [15] argues that the measurement using the F1-Score is more suitable than the accuracy. This is because most of the NER data labels are labeled as O, which refers to tokens that are not an entity named (named entity), and thus high accuracy can be obtained. Therefore, this study will use the F1-Score as a parameter for measuring model performance. F1-Score is the harmonic mean of precision

and recall as shown in (2). The best score that the F1-Score can achieve is 1, while the worst is 0. This value can also be represented in a percentage ranging from 0-100%, which will also be used in this study.

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (2)$$

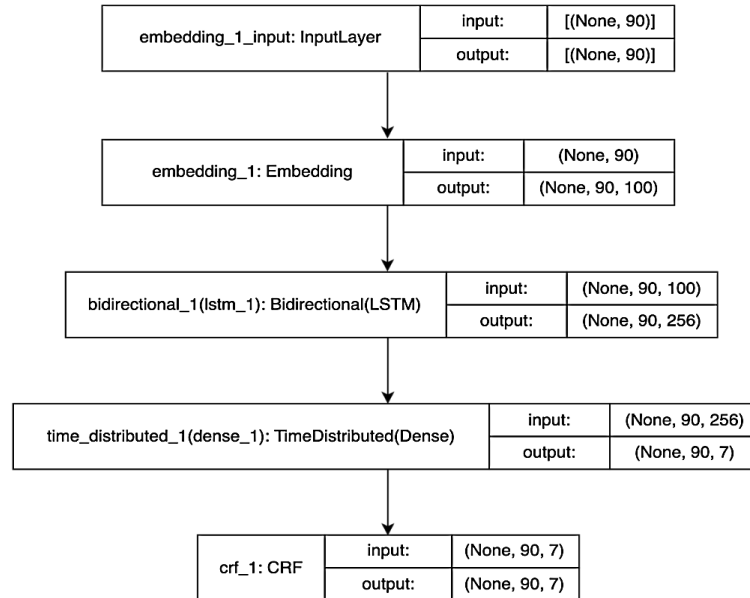


Figure 1. BiLSTM-CRF architecture

3. RESULTS AND DISCUSSION

In this study, we set the hyperparameter with various values to obtain the best scenario. The initial scenario in this study used the Skip-gram algorithm for Word2Vec, 128 LSTM units, the dropout in the LSTM layer is 0.5, TanH as the activation function in the dense layer, Adam optimization function, 32 batch size, and 30 epochs. To get the best model performance for NER in the tourism domain, we made seven different scenarios, as shown in Table 4. We conducted scenarios 1 to 7 sequentially. The configuration that produces the best performance in each scenario will be used in the subsequent scenarios.

Table 4. Hyperparameter setting

Scenario	Hyperparameter	Configuration
1	Learning rate	0.01
		0.001
		0.0001
2	Word2Vec algorithm	Skip-gram
		CBOw
3	Dense layer's activation function	TanH
		ReLU
		Linear
4	Batch size	32
		64
5	LSTM unit	100
		128
6	Epoch	30
		50
7	Optimization function	Adam
		Nadam

Based on all scenarios that have been done, the model with the best scenario obtained an average F1-Score of 75.25% and used configurations shown in Table 5. In addition, the accuracy and average F1-Score generated for each scenario are shown in Table 6. Table 7 presents the best F1-Score for each type of

tourist attraction. The best model is then used to detect named entities in new data. The detected entities are tourist attractions classified into three categories: Heritage Attraction, Purposeful Built (Man-Made) Attraction, and Natural Attraction. Meanwhile, words that are not tourist attractions will be included in the outside category.

Figure 2 illustrates the examples of the named entity detection results from our dataset. The result shows that our best model was able to detect some named entities correctly. However, our NER model still makes some mistakes while predicting the label. For example, several types of tourist attractions were still wrongly detected as tourist attractions. This mistake may happen due to the lack of tourist attractions in our dataset. Therefore, our model fails in predicting the entity.

Table 5. Hyperparameter setting for the best model

Hyperparameter	Configuration
Learning rate	0.001
Word2Vec algorithm	Skip-gram
Dense layer's activation function	Linear
Batch size	32
LSTM unit	128
Epoch	50
Optimization function	Adam

Table 6. Accuracy and average F1-Score for each scenario

Scenario	Hyperparameter	Configuration	Training Accuracy (%)	Validation Accuracy (%)	Testing Accuracy (%)	Average F1-Score (%)
1	Learning rate	0.01	98.51	98.52	98.55	14.27
		0.001	99.67	98.80	99.43	68.68
		0.0001	98.97	98.80	98.94	38.83
2	Word2Vec algorithm	Skip-gram	99.67	98.80	99.43	68.68
		CBOW	99.50	98.48	99.14	53.01
3	Dense layer's activation function	TanH	99.67	98.80	99.43	68.68
		Linear	99.71	98.79	99.46	71.66
		ReLU	99.37	98.7	99.19	54.45
4	Batch size	32	99.71	98.79	99.46	71.66
		64	99.61	98.79	99.37	67
5	LSTM unit	100	99.56	98.77	99.34	66
		128	99.71	98.79	99.46	71.66
6	Epoch	30	99.71	98.79	99.46	71.66
		50	99.82	98.73	99.52	75.25
7	Optimization function	Adam	99.82	98.73	99.52	75.25
		Nadam	99.82	98.79	99.51	74.61

Table 7. The best F1-Score for each type of tourist attraction

Label	F1-Score										Average
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	
HERITAGE	0.45	0.40	0.60	0.76	0.82	0.85	0.86	0.79	0.83	0.76	0.712
NATURAL	0.47	0.58	0.68	0.79	0.86	0.87	0.85	0.86	0.87	0.89	0.772
PURPOSE	0.48	0.54	0.66	0.77	0.85	0.88	0.85	0.84	0.90	0.89	0.766

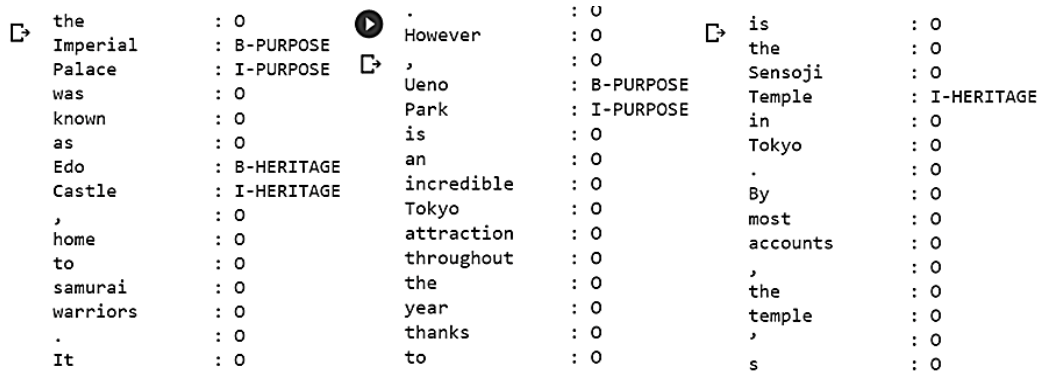


Figure 2. The example of entity detection results

4. CONCLUSION

In this study, a named entity recognition system for tourism has been presented. We focused on extracting tourism entities based on their categories: natural attraction, heritage attraction, and purposefully built (artificial) attraction. The experiments have shown that the proposed BiLSTM-CRF algorithm has demonstrated promising results in identifying named entities from the tourism dataset. We also found that the application of word2vec with skip-gram can improve the performance of the named entity system. This research has produced a model that could predict new data quite well, but there were still some mistakes in the detection. We experimented with various scenarios, and the best model produced an average F1-Score of 75.25%. For future work, applying any other word representation models can be considered to improve the performance of named entity detection. In addition, we suggest adding more entity labels, such as country, city location, and tourism name.




REFERENCES

- [1] UNWTO, "UNWTO World Tourism Barometer and Statistical Annex, January 2020," *UNWTO World Tourism Barometer*, vol. 18, no. 1, pp. 1–48, Jan. 2020, doi: 10.18111/wtobarometereng.2020.18.1.1.
- [2] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews," *Journal of Computational Science*, vol. 27, pp. 386–393, Jul. 2018, doi: 10.1016/j.jocs.2017.11.006.
- [3] "The 2014 Traveler's Road to Decision," 2014. Accessed: June 26, 2020. [Online]. Available: <https://www.thinkwithgoogle.com/consumer-insights/2014-travelers-road-to-decision/>.
- [4] R. A. N. Nayoan, A. F. Hidayatullah, and D. H. Fudholi, "Convolutional Neural Networks for Indonesian Aspect-Based Sentiment Analysis Tourism Review," Aug. 2021, doi: 10.1109/icoict52021.2021.9527518.
- [5] K. H. Yoo and U. Gretzel, "What Motivates Consumers to Write Online Travel Reviews?," *Information Technology & Tourism*, vol. 10, no. 4, pp. 283–295, Dec. 2008, doi: 10.3727/109830508788403114.
- [6] C. Chantrapornchai and A. Tunsakul, "Information Extraction based on Named Entity for Tourism Corpus," Jul. 2019, doi: 10.1109/jcsse.2019.8864166.
- [7] A. Ishino, H. Nanba, and T. Takezawa, "Automatic Compilation of an Online Travel Portal from Automatically Extracted Travel Blog Entries," in *Information and Communication Technologies in Tourism 2011*, Springer Vienna, 2011, pp. 113–124.
- [8] R. Alfred, L. C. Leong, C. K. On, and P. Anthony, "Malay Named Entity Recognition Based on Rule-Based Approach," *International Journal of Machine Learning and Computing*, vol. 4, no. 3, pp. 300–306, Jun. 2014, doi: 10.7763/ijmlc.2014.v4.428.
- [9] A. Das, D. Ganguly, and U. Garain, "Named Entity Recognition with Word Embeddings and Wikipedia Categories for a Low-Resource Language," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 16, no. 3, pp. 1–19, Apr. 2017, doi: 10.1145/3015467.
- [10] G. Aguilar, S. Maharjan, A. P. L. Monroy, and T. Solorio, "A Multi-task Approach for Named Entity Recognition in Social Media Data," 2017, doi: 10.18653/v1/w17-4419.
- [11] D. Kaur and V. Gupta, "A survey of named entity recognition in English and other Indian languages," *International Journal of Computer Science Issues (IJCSI)*, vol. 7, no. 6, pp. 239–245, 2010.
- [12] C. Lyu, B. Chen, Y. Ren, and D. Ji, "Long short-term memory RNN for biomedical named entity recognition," *BMC Bioinformatics*, vol. 18, no. 1, Oct. 2017, doi: 10.1186/s12859-017-1868-5.
- [13] N. V. Sobhana, P. Mitra, and S. K. Ghosh, "Conditional Random Field Based Named Entity Recognition in Geological text," *International Journal of Computer Applications*, vol. 1, no. 3, pp. 143–147, Feb. 2010, doi: 10.5120/72-166.
- [14] D. Zeng, C. Sun, L. Lin, and B. Liu, "LSTM-CRF for Drug-Named Entity Recognition," *Entropy*, vol. 19, no. 6, p. 283, Jun. 2017, doi: 10.3390/e19060283.
- [15] M. Seok, H.-J. Song, C.-Y. Park, J.-D. Kim, and Y. Kim, "Named Entity Recognition using Word Embedding as a Feature," *International Journal of Software Engineering and Its Applications*, vol. 10, no. 2, pp. 93–104, Feb. 2016, doi: 10.14257/ijseia.2016.10.2.08.
- [16] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," *arXiv preprint*, no. 1508.01991, 2015.
- [17] H. Wei *et al.*, "Named Entity Recognition From Biomedical Texts Using a Fusion Attention-Based BiLSTM-CRF," *IEEE Access*, vol. 7, pp. 73627–73636, 2019, doi: 10.1109/access.2019.2920734.
- [18] Q. Wang, Y. Zhou, T. Ruan, D. Gao, Y. Xia, and P. He, "Incorporating dictionaries into deep neural networks for the Chinese clinical named entity recognition," *Journal of Biomedical Informatics*, vol. 92, p. 103133, Apr. 2019, doi: 10.1016/j.jbi.2019.103133.
- [19] H. A. Nayel and H. L. Shashirekha, "Integrating dictionary feature into a deep learning model for disease named entity recognition," *arXiv preprint*, no. 1911.01600, 2019.
- [20] L. Xue, H. Cao, F. Ye, and Y. Qin, "A Method of Chinese Tourism Named Entity Recognition Based on BBLC Model," Aug. 2019, doi: 10.1109/smartworld-uic-atc-scalcom-iop-sci.2019.00307.
- [21] S. K. Siencnik, "Adapting word2vec to named entity recognition," in *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*, 2015, pp. 239–243.
- [22] K. E. Saputro, S. S. Kusumawardani, and S. Fauziati, "Development of semi-supervised named entity recognition to discover new tourism places," Oct. 2016, doi: 10.1109/icstc.2016.7877360.
- [23] X. Cheng, W. Wang, F. Bao, and G. Gao, "MTNER: A Corpus for Mongolian Tourism Named Entity Recognition," in *Communications in Computer and Information Science*, Springer Singapore, 2020, pp. 11–23.
- [24] A. F. Hidayatullah and M. R. Ma'arif, "Pre-processing Tasks in Indonesian Twitter Messages," *Journal of Physics: Conference Series*, vol. 801, p. 12072, Jan. 2017, doi: 10.1088/1742-6596/801/1/012072.
- [25] Z. Dai, X. Wang, P. Ni, Y. Li, G. Li, and X. Bai, "Named Entity Recognition Using BERT BiLSTM CRF for Chinese Electronic Health Records," Oct. 2019, doi: 10.1109/cisp-bmei48845.2019.8965823.
- [26] T. Mikolov, Q. v Le, and I. Sutskever, "Exploiting similarities among languages for machine translation," *arXiv preprint*, no. 1309.4168, 2013.




- [27] N. K. Manaswi, *Deep learning with applications using python: chatbots and face, object, and speech recognition with tensorflow and keras*. Apress, 2018.
- [28] X. Ma and E. Hovy, "End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF," 2016, doi: 10.18653/v1/p16-1101.
- [29] Z. Zhao *et al.*, "Chinese named entity recognition in power domain based on Bi-LSTM-CRF," 2019, doi: 10.1145/3357254.3357283.

BIOGRAPHIES OF AUTHORS






Annisa Zahra    is a machine learning engineer at Lenna.ai. She obtained her Bachelor's degree in Informatics in 2021 from Universitas Islam Indonesia, Indonesia. She has done several machine learning projects in the domain of natural language processing. She can be contacted at email: az.annisazahra@gmail.com



Ahmad Fathan Hidayatullah    is an assistant professor at the Department of Informatics, Faculty of Industrial Engineering, Universitas Islam Indonesia. He is currently pursuing his PhD in Computer Science Programme, Faculty of Science, Universiti Brunei Darussalam with The UBD Graduate Scholarship (UGS). He earned his Master's degree in computer science (M.Cs.) from Universitas Gadjah Mada, Indonesia, in 2014. Previously, he completed his Bachelor's degree (B. Eng) in Informatics in 2010 from Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia. His research interests are mainly related to text mining, natural language processing, and data science. He is also actively involved in several research in Center of Data Science, Universitas Islam Indonesia. He can be contacted at email: fathan@uii.ac.id



Septia Rani    is a lecturer at the Department of Informatics Universitas Islam Indonesia. She obtained her Master's degree in Computer Science in 2014 from Universitas Gadjah Mada, Indonesia. Her research interests are mainly related to artificial intelligence and data science. Currently, she is working on a research project related to the development of a smart application for travel itinerary planning using artificial intelligence techniques. The methods include optimization and the use of recommendation systems. She is also working on several data science projects in the domains of computer vision and social science. She can be contacted at email: septia.rani@uii.ac.id