Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 1 (2023) 1371 - 1377 ISSN Media Electronic: 2580-0760

Feature Extraction for Improvement Text Classification of Spam YouTube Video Comment using Deep Learning

Jasmir Jasmir[1*], Willy Riyadi[2], Pareza Alam Jusia[3]
[1,2]Computer Engineering, Universitas Dinamika Bangsa, Jambi, Indonesia
[3]Informatic Engineering Universitas Dinamika Bangsa, Jambi, Indonesia
[1]ijay_jasmir@yahoo.com, [2]wriyadi5@gmail.com, [3]parezaalam@gmail.com

Abstract

The proposed algorithms are Bidirectional Long Short Term Memory (BiLSTM) and Conditional Random Fields (CRF) with Data Augmentation Technique (DAT). DAT integrates spam YouTube video comments into the traditional TF-IDF algorithm and generates a weighted word vector. The weighted word vector is fed into BiLSTM CRF to capture context information effectively. The result of this study is a new classification model to spam YouTube comment videos and increase the computational value of its performance. This research conducted two experiments: the first using BiLSTM CRF without DAT and the second using BiLSTM CRF with DAT. The experimental results state that the evaluation score using BiLSTM CRF with DAT shows outstanding performance in text classification, especially in spam YouTube video comment texts, with accuracy = 83.3%, precision = 83.6%, recall = 83.3%, and F-measure = 83.3%. So the combination of the BiLSTM-CRF method and the Data Augmentation Technique is very precise, so it can be used to increase the accuracy of classification texts for spam YouTube video comments

Keywords: *Improvement; BLSTM; CRF; Data Augmentation; Feature Extraction*

1. Introduction

Currently, information and opinions flow freely since the emergence of social media. Through social media, everyone has the right to voice their opinion. This allows all of these rights to be expressed through social media platforms which are currently able to spread quickly and widely. One of the most popular social media platforms is YouTube, which has billions of active users who are also content creators[1], [2]. Apart from being a content creator, most social media platforms allow users to provide reactions such as commenting, also reacting by using the like, dislike, and share features[3].

One of the platforms discussed in this study is YouTube. Youtube is a video sharing platform for video owners and viewers to interact, such as liking or disliking videos and/or commenting on videos[4] . In the comments section, viewers can express opinions and emotions related to the video or even not related to the video. It is also a way for video viewers to interact with video creators[5] . The number of "likes" and the percentage of "likes" of a video, is important for content creators on the platform because videos with a high proportion of dislikes generally provide negative publicity[3] . With most user data available through social media, it is very possible to obtain information from users who express opinions through words or emotions that contain spam[6]. With so much YouTube content with various themes, this will produce a very large amount of text data that comes from user comments that interact with content creators.

With the increasing amount of data and data complexity as in the case above, deep learning offers an offer to solve this problem, with its high and almost perfect processing capabilities, deep learning[7] has achieved outstanding results in various fields, such as computer vision[8] , voice recognition [9] , and text classification [10] - [12].

Text classification is known as one of the most effective methods for an organization and set of information management about texts, and has been used in several fields such as the field of information sorting [13] sentiment analysis[14], [15], spam filtering[16], [17], etc. currently the most appropriate method for classifying is the deep learning method [18], [19]. The application of neural networks, including methods such as conditional random fields and recurrent neural

Accepted: 07-07-2023 | Received in revised: 21-12-2023 | Published: 26-12-2023

networks, has become a prevalent approach in text classification within the realm of deep learning[20].

RNN has attained remarkable achievement in handling serialization tasks[21]. Because the RNN has a repeating network structure that is able to store and maintain the information received, the RNN can integrate components into one information system more precisely and better. [22]. RNN has drawbacks such as frequent scattering and disappearance of gradients, so to avoid this problem, Long Short-Term Memory (LSTM)[23], [24] and other variants[21] has been created to perform memory enhancement and memory access. In this case, it is proven that LSTM has shown abilities that are in line with expectations, especially in natural language processing.

Conditional Random Fields (CRFs) is a deep learning technique that enhances the computational effectiveness of text classification[25]. CRF is a well-established model that predicts the most probable sequence of labels associated with a sequence of inputs. The objective of this research is to enhance the text classification of cancer clinical trials through the utilization of the Conditional Random Fields approach.

Conditional Random Fields (CRFs) are probabilistic models that are widely employed for labeling and segmenting sequential data[26], [27]. CRFs have diverse applications in various fields such as named entity recognition, part of speech tagging, gene prediction, text classification, noise reduction, and object detection, among others[28].

In his research, Zuheros[29] addresses the challenges associated with one of the deep learning methods, aiming to enhance system performance and computational efficiency. To tackle these issues, he proposes a neural network architecture that aims to improve both the overall system performance and computational efficiency. While Zhenjin Dai[30] once combined CRF with LSTM which produced good accuracy but still with low computational performance. The development of LSTM is BiLSTM. Zhiheng Huang[31] states that the combination of BiLSTM and CRF can produce sophisticated (or close) accuracy. Based on the problem above, the authors propose a model that is the BiLSTM-CRF model on the clinicaltrial.gov dataset to improve computational performance.

Apart from using the methods mentioned above, the classification value can also be increased by utilizing the Data Augmentation Technique. Several researchers, including Jason Wei, have explored and implemented this technique[32] In his research, Jason Wei used the Data Augmentation Technique as a feature to help add value to the performance evaluation of text classification up to 88.6%. Then Francisco J Moreno[33] also uses data augmentation techniques to increase the value of classification accuracy on small datasets. Based on the description above, we propose and build a new model, namely the BLSTM CRF model which is integrated with the Data Augmentation Technique as feature extraction to increase the evaluation value in the classification of spam youtube video comments.

Several studies related to the BiLSTM-CRF include Buzhou Tang[34], in this study, they propose a deep neural network (called LSTM-CRF) combining long short-term memory (LSTM) neural networks (a type of recurrent neural networks) and conditional random fields (CRFs) to recognize ADR mentions from social media inmedicine and investigate the effects of three factors on ADR mention recognition

Yao Chen[31], The objective of this research was to identify adverse drug reaction (ADR)-related entities from the unstructured text section of Chinese ADR reports. This extraction of entities aims to provide additional information to supplement the structured section of the reports, thereby enhancing the evaluation of ADRs.

Zhongbao Wan[35], this research paper presents a method that combines Bidirectional Long Short-Term Memory (BiLSTM) with an attention mechanism and incorporates word vector information. The proposed model operates in three main steps. Firstly, the model extracts character vector features from the text corpus, which are then fed into the BiLSTM model. Secondly, an attention mechanism is employed to calculate the relevance between the current input character and other input characters within the BiLSTM model. Lastly, the global feature is obtained based on the calculated relevance, and it is concatenated with the word vector feature. The Conditional Random Field (CRF) layer is then introduced to enforce mutual constraints between tags using this combined feature representation.

Zohaib Mushtaq[36] applied audio data augmentation (deformation) with CNN for environmental sound classification. Deformation involves stretching time (i.e., slowing down or speeding up the audio sample). Each deformation is applied directly to the audio signal before converting it into an input for training neural networks, then combined with data augmentation techniques, their proposed model produces cutting-edge results for environmental sound classification.

Data augmentation methods have gained significant popularity in various domains, including visual recognition. These techniques enable the generation of new data by applying straightforward image manipulation operations, thereby simplifying the process. Pei Wang et al.[37], while training MaxGrad, or Yanan Sun et al.[38], use the genetic algorithm to applied transformations to the input image to improve the performance of this model. Apart from the aforementioned researchers, Jason Wei[32] Georgios

DOI: https://doi.org/10.29207/resti.v7i6.5249

Rizos[39] and Anaıs Ollagnier[40] has also conducted research on enhancing the value of classification through the utilization of Data Augmentation Techniques.

2. Research Methods

In general, the sequence of steps in this research can be seen in the Figure 1. Starting with selecting a dataset, then proceeding with changing the categorical variables into numeric form or one hot vector. After that, the classification process was carried out with two experiments, namely using features and without using features and comparing the values.

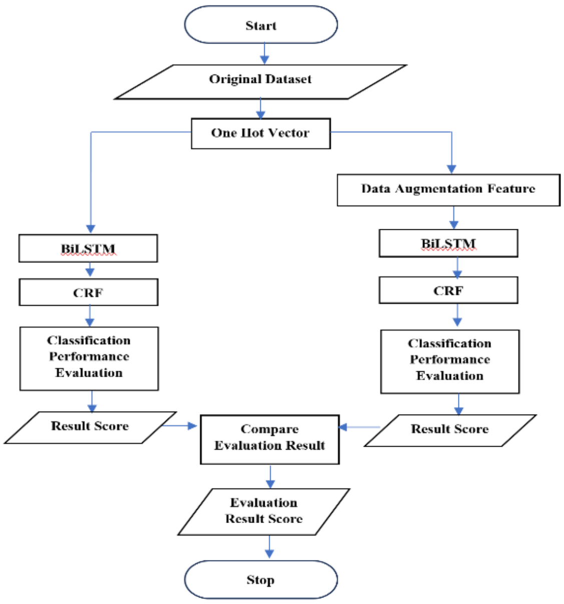

Figure 1. Research Framework

The dataset used was taken from the UCI Machine Learning website with the link https://archive.ics.uci.edu/ml/datasets/Youtube+Spam+Collection, that were trained using the python jupyter notebook software. This dataset is a collection of public comments collected for spam research. This dataset has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed videos in the collection period. The dataset in question is as shown in the Table 1.

Table 1. Table of dataset spam youtube video comments

Dataset	Youtube ID	Spam	Ham	Total
Psy	9bZkp7q19f0	175	175	350
KatyPerry	CevxZvSJLk8	175	175	350
LMFAO	KQ6zr6kCPj8	236	202	438
Eminem	uelHwf8o7_U	245	203	448
Shakira	pRpeEdMmmQ0	174	196	370
Total		1005	951	1956

2.1. Construction of Weighted Word Vector

In natural language processing, words or sentences must be converted into a vector form known as word to vectors or commonly abbreviated as word2vec. The work process from the word or language used is then mapped into the vector space. In a vector space each word is represented by a real number. The CBOW model predicts target words based on a context distribution. For the word wk, the context is stated as Formula 1.

$$Context\ (W_k) = \{w_{k-t}, w_{k-(t-1)}, .., w_{k+(t-1)}, w_{k+t}\} \quad (1)$$

TF IDF is defined as a calculation of how related a word in the corpus is to a sentence that accompanies it. The relational value will increase rationally with the number of occurrences of the word in the sentence. The TF-IDF formula is as shwn in Formula 2 and 3.

$$w(ti, d) = \frac{tf\ (ti,d) x\ idf\ (ti)}{\sqrt{\sum_{ti \epsilon d}[tf\ (ti,d) x\ idf\ (ti)]^2}} \quad (2)$$

$$idf(ti) = \log\left(\frac{N}{Nti}\right) + 1 \quad (3)$$

The weight calculation method for word vectors is as shown in Formula 4 and 5.

$$wi = tf - idfi\ .e \quad (4)$$

$$e = \begin{cases} \alpha, & ti\ is\ not\ eligibility\ class \\ 1, & ti\ is\ eligibility\ class \end{cases} \quad (5)$$

2.2. Data Augmentation

Data Augmentation Technique (DAT) is the practice of applying transformations to existing training data sets to develop them synthetically [41]. In this work, data augmentation was the addition of several other types of spam comment and insertion them into a spam statement corpus for each criterion [42]. In this case the data augmentation were taken from some of the youtube video comments from the link https://statso.io/youtube-comments-data/. The data used are 44 comments, so the total YouTube video comment data are 2000 comments

Here, we present the full DAT details. For certain sentences in the training set, we randomly select and perform one of the following operations: 1. Synonym Replacement (SR): Randomly select n words from non-stop sentences. Replace each of these words with one of its synonyms is chosen at random; 2.Random Insertion (RI): Find random synonyms of a random word in a sentence that is not a stop word. Insert these synonyms at random positions in the sentence. Do this n times; 3. Random Swap (RS): Select two random words in a sentence and swap their positions. Do this n times; 4. Random Deletion (RD): Randomly deletes every word in a sentence with probability p.

Because long sentences have more words than short sentences, they can absorb more noise while retaining their original class label. To compensate, we vary the number of words changed, n, for SR, RI, and RS based on sentence length l with the formula n = α l, where α is the parameter that shows the percentage of words in a sentence is changing (we use p = α for RD). Next, for

DOI: https://doi.org/10.29207/resti.v7i6.5249

each original sentence, we generate an additional sentence.

2.3. BLSTM Layer

Long Short-term Memory (LSTM)[10] is an enhanced version of the recurrent neural network (RNN) that addresses its limitations. Among various modifications of RNN, LSTM has gained significant popularity. The primary purpose of LSTM is to overcome the inability of RNN to effectively predict words based on long-term stored information. LSTM is designed to remember and manage a comprehensive set of information stored over an extended period. It also has the capability to discard irrelevant information. This makes LSTM more efficient in processing, predicting, and classifying data with temporal dependencies.

Systems utilizing LSTM can handle time series data, enabling processing, prediction, and classification based on the temporal sequence. By retaining and discarding data selectively, LSTM ensures more complete and up-to-date information management. The key distinction between LSTM and RNN lies in the structure of the network. While RNN allows each neuron to process one input and produce one output, LSTM operates differently. It incorporates multiple gateways to augment and integrate information.

LSTM consists of four essential gates: forget gate, input gate, input modulation gate, and output gate. These gates serve distinct functions and responsibilities in gathering, categorizing, and processing data. In addition to the gates, LSTM also includes an internal cell state that stores relevant information from previous iterations.

2.4. CRF Layer

The inclusion of Conditional Random Fields (CRF) within the model serves the purpose of incorporating constraints on the label relationships. This ensures the validity of the predicted labels and aids in identifying the most optimal sequence of labels[35], [26].

2.5. BiLSTM-CRF.

For this research, we employ LSTM-CRF as the modeling approach for this problem. The architecture of LSTM-CRF is depicted in Figure 3, comprising three layers: (1) input layer, (2) LSTM layer, and (3) CRF layer. The input layer is responsible for receiving various types of token embeddings. In this study, we utilize word embeddings, character-level embeddings, subject-related embeddings, and knowledge-based embeddings as the input embeddings. The LSTM layer utilizes bidirectional LSTM neural networks to produce hidden context representations for each position..

The complete work structure of the classification process from this study can be seen in Figure 2.

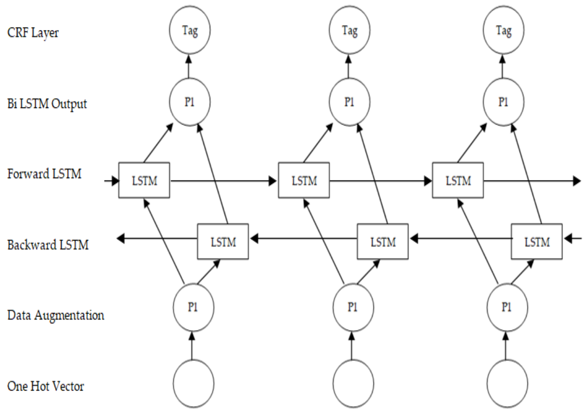

Figure 2. BiLSTM CRF Structure

Figure 2 shows the structure of the combined BiLSTM CRF process. Initially, one hot vector is used as input and then data augmentation is carried out using certain techniques. After that, the data is given to the BiLSTM model or two-way LSTM which consists of two parts, namely backward and forward. This model performs a detrain process (or commonly called backward training) on the data. The results of the BiLSTM process become the output of the model. The final step is the tagging process using CRF to produce the final output.

3. Results and Discussions

We have carried out two experiments, namely an experiment using the BLSTM CRF model without the Data Augmentation Technique, and the second experiment, namely an experiment using the BLSTM CRF with the Data Augmentation Technique. The results of the process from these two experiments have shown results that are in accordance with the aim of this research, namely increasing the evaluation value of classification performance, especially in evaluating accuracy, precision, recall and f1-score. We divide the training value data and test value data by 80 : 20 and use 10-fold cross validation. The differences in the results of the experimental process can be seen in the following information.

Table 1. Accuracy BLSTM-CRF without DAT

Summary	Value	Percent
Correctly Classified Instances	1554	77,74%
Incorrectly Classified Instances	445	22,26%
Total Number of Instances	1999	
Ignore Class Unknown Instances	1	

Table 1 explains that the results of the evaluation using the BLSTM-CRF without DAT yielded correctly classified Instances values of 1554 data which were considered correct with an accuracy value of 77.74%, out of 2000 data there was 1 data that could not be processed.

Table 2 describes the results of precision, recall and F-Measure, with an average value of precision = 77.8%; recall = 77.7%; and F-Measure = 77.7%

DOI: https://doi.org/10.29207/resti.v7i6.5249

Table 2. Value of Evaluation BLSTM-CRF without DAT

Detailed Accuracy by class			
Class	Precision	Recall	F-Measure
1	0,768	0,808	0,788
0	0,788	0,745	0,766
Weighted Avg	0,778	0,777	0,777

Table 3. Confusion Matrix BLSTM-CRF without DAT

Confusion Matrix		
a	b	<-- Classified as
825	196	a=1
249	729	b=0

Table 3 explains the confusion matrix, as shown above the results of the confusion matrix are true positive values = 825; False Negatives = 196; False Positive = 249 and True Negative = 729.

Table 4. Accuracy BLSTM-CRF with DAT

Summary	Value	Percent
Correctly Classified Instances	1666	83,3417
Incorrectly Classified Instances	333	16,6583
Total Number of Instances	1999	
Ignore Class Unknown Instances	1	

Table 4 explains that the results of the evaluation using the BLSTM-CRF with the DAT feature yielded correctly classified Instances values of 1666 data which were considered correct with an accuracy value of 83.34%, out of 2000 data there was 1 data that could not be processed

Table 5. Value of Evaluation BLSTM-CRF with DAT

Detailed accuracy by class			
Class	Precision	Recall	F-Measure
1	0,81	0,881	0,844
0	0,863	0,784	0,822
Weighted Avg	0,836	0,833	0,833

Table 5 describes the results of precision, recall and F-Measure, with an average value of precision = 83,6%; recall = 83,3%; and F-Measure = 83,3%.

Table 6. Confusion Matrix BLSTM-CRF with DAT

Confusion matrix		
a	b	<-- Classified as
899	122	a = 1
211	767	b = 0

Table 6 explains the confusion matrix, as shown above the results of the confusion matrix are true positive values = 899; False Negatives = 122; False Positive = 211 and True Negative = 767

Table 7. Table of BILSTM CRF without DAT

Evaluation	BiLSTM CRF (%)
Accuracy	77,7
Precision	77,8
Recall	77,7
F-Measure	77.7

Table 7 and Figure 3 are the results of a collection of BiLSTM CRF evaluation values without DAT. The evaluation results from the BiLSTM CRF without DAT stated Accuracy = 77.7; Precision = 77.8; Remember = 77.7; and F-1 score = 77.7.

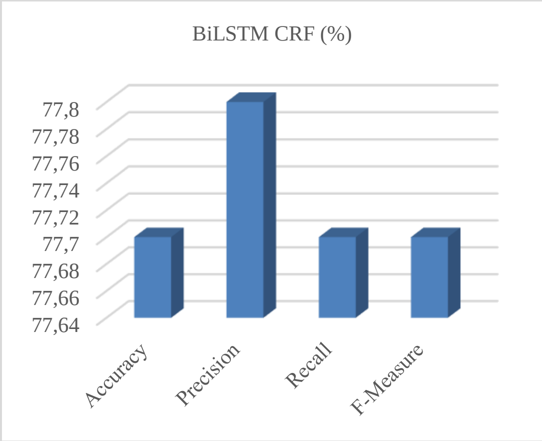

Figure 3 Graph of BiLSTM CRF without DAT

Table 8 and Figure 4 are the results of a collection of BiLSTM CRF evaluation values with DAT. The evaluation results from the BiLSTM CRF with DAT stated Accuracy = 83,3; Precision = 83,6; Recall = 83,3; and F-1 Score = 83,3

Table 8. Table of BILSTM with DAT

Evaluation	BiLSTM CRF +DAT (%)
Accuracy	83,3
Precision	83,6
Recall	83,3
F-1 Score	83,3

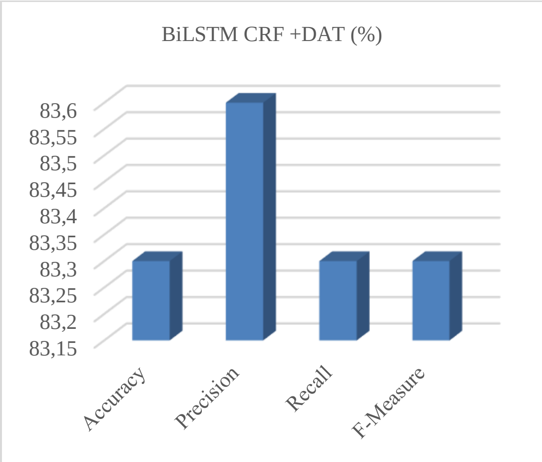

Figure 4. Graph of BiLSTM CRF with DAT

The classification evaluation process has been completed, the effect of feature extraction to increase the evaluation value has been seen and the results have been obtained. As seen in the table and graph above. The combination of BiLSTM CRF with DAT has proven that the proposed method is able to increase the evaluation value. The evaluation value using BiLSTM CRF with DAT outperformed the evaluation value for BiLSTM CRF without DAT. To clarify the differences in the evaluation results between the use of BiLSTM

DOI: https://doi.org/10.29207/resti.v7i6.5249

without DAT and the use of BiLSTM with DAT in the spam youtube video comments dataset, we present them in Table 9 and Figure 5.

This increase in accuracy occurs when combined with the data augmentation feature due to the selection of several appropriate parameters, namely the parameter of data variation by adding data. By having more sample data, the model will have more opportunities to understand underlying patterns and capture important features of the data.

Table 9. Table of Compare BILSTM CRF Without DAT And BILSTM CRF with DAT

Evaluation	BiLSTM CRF (%)	BiLSTM CRF + DAT (%)
Accuracy	77,7	83,3
Precision	77,8	83,6
Recall	77,7	83,3
F-1 Score	77.7	83,3

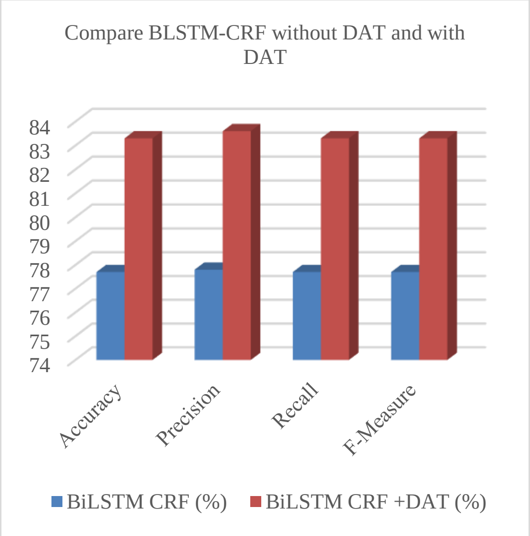

Figure 5. Graph of Compare BiLSTM without DAT and BiLSTM with DAT

Data variation can also help reduce overfitting. Models that are trained on disparate data sets will be more commonly used and better at recognizing overall patterns than models that over-adjust to small details in the training dataset

The Context BiLSTM CRF with DAT achieves superior performance because it leverages contextual information acquired specifically from spam youtube comment videos. This allows the classifier to encounter fewer uncertainties when confronted with other types of contextual knowledge, such as general text. Consequently, this model demonstrates a higher accuracy in correctly identifying disease concepts among the predicted disease name mentions.

Our paper seeks to tackle the absence of a standardized data augmentation approach in the field of Natural Language Processing (NLP), in comparison to the field of computer vision. To address this gap, we propose a set of straightforward operations that can serve as a foundation for future explorations. Considering the rapid advancements in NLP research in recent years, we anticipate that researchers will soon uncover a more effective and user-friendly data augmentation technique that yields even better performance.

4. Conclusion

This paper introduces a novel model called BiLSTM CRF with DAT for analyzing spam comments on YouTube videos. The experiment conducted in this study compares the performance of BiLSTM CRF without DAT to that of BiLSTM CRF with DAT. The results demonstrate that the utilization of BiLSTM CRF with DAT leads to improved evaluation outcomes. Furthermore, the model's performance is evaluated by utilizing it to classify comments on YouTube videos as either spam or non-spam. A comparison between different models reveals that our proposed model is better suited for detecting spam in YouTube video comments, namely the BLSMT model with DAT. Moreover, the evaluation demonstrates that our BiLSTM CRF with the DAT model achieves impressive results and surpasses several basic models in terms of performance.

References

[1] R. Lozano-Blasco, M. Mira-Aladrén., and M. Gil-Lamata, “Social media influence on young people and children: Analysis on Instagram, Twitter and YouTube,” *Comunicar*, vol. 30, no. 74, pp. 117–128, 2023, doi: 10.3916/C74-2023-10.

[2] I. D. and I. Tabak, “An empirical analysis of knowledge co-construction in YouTube comments Ilana,” 2020.

[3] W. W. Sitompul, “Penelitian Tentang Youtube,” *... J. Perpust. dan Inf.*, vol. 2275, 2022.

[4] H. A. R. Harpizon, R. Kurniawan, Iwan Iskandar, R. Salambue, E. Budianita, and F. Syafria, “Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes,” *... Di YouTube ...*, vol. 5, no. 1, pp. 131–140, 2022.

[5] R. Handayani, “Youtube Sebagai Media Komunikasi Dalam Berdakwah Di Tengah Pandemi,” *Hikmah*, vol. 15, no. 1, pp. 123–138, 2021, doi: 10.24952/hik.v15i1.3569.

[6] Rusydi Umar, Sunardi, and M. N. A. Nuriyah, “Comparing the Performance of Data Mining Algorithms in Predicting Sentiments on Twitter,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 4, pp. 817–823, 2023, doi: 10.29207/resti.v7i4.4931.

[7] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, “Deep learning in spiking neural networks,” *Neural Networks*, vol. 111, pp. 47–63, 2019, doi: 10.1016/j.neunet.2018.12.002.

[8] K. G. Nalbant and Ş. Uyanik, “Computer Vision in the Metaverse,” *J. Metaverse*, vol. 1, no. 1, pp. 9–12, 2021.

[9] Ł. Brocki and K. Marasek, “Deep Belief Neural Networks and Bidirectional Long-Short Term Memory Hybrid for Speech Recognition,” vol. 40, no. 2, pp. 191–195, 2015, doi: 10.1515/aoa-2015-0021.

[10] K. S. Tai, R. Socher, and C. D. Manning, “Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks”.

[11] A. Elnagar, R. Al-debsi, and O. Einea, “Arabic text classification using deep learning models,” vol. 57, no. April 2019, 2020.

DOI: https://doi.org/10.29207/resti.v7i6.5249

[12] J. Jasmir, S. Nurmaini, and B. Tutuko, "Fine-grained algorithm for improving knn computational performance on clinical trials text classification," *Big Data Cogn. Comput.*, vol. 5, no. 4, 2021, doi: 10.3390/bdcc5040060.

[13] J. Zhang, F. Liu, W. Xu, and H. Yu, "Feature fusion text classification model combining CNN and BiGRU with multi-attention mechanism," *Futur. Internet*, vol. 11, no. 11, 2019, doi: 10.3390/fi11110237.

[14] A. Alsaeedi and M. Z. Khan, "A study on sentiment analysis techniques of Twitter data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 2, pp. 361–374, 2019, doi: 10.14569/ijacsa.2019.0100248.

[15] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," *IEEE Access*, vol. 7, pp. 51522–51532, 2019, doi: 10.1109/ACCESS.2019.2909919.

[16] H. Baaqeel and R. Zagrouba, "Hybrid SMS spam filtering system using machine learning techniques," *Proc. - 2020 21st Int. Arab Conf. Inf. Technol. ACIT 2020*, 2020, doi: 10.1109/ACIT50332.2020.9300071.

[17] H. Yang, Q. Liu, S. Zhou, and Y. Luo, "A spam filtering method based on multi-modal fusion," *Appl. Sci.*, vol. 9, no. 6, 2019, doi: 10.3390/app9061152.

[18] J. Jasmir, W. Riyadi, S. R. Agustini, Y. Arvita, D. Meisak, and L. Aryani, "Bidirectional Long Short-Term Memory and Word Embedding Feature for," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 6, no. 4, pp. 505–510, 2022, [Online]. Available: https://jurnal.iaii.or.id/index.php/RESTI/article/view/4005/606

[19] J. Jasmir, S. Nurmaini, R. F. Malik, and D. Zaenal, "Text Classification of Cancer Clinical Trials Documents Using Deep Neural Network and Fine Grained Document Clustering," vol. 172, no. Siconian 2019, 2020.

[20] M. Zulqarnain, R. Ghazali, Y. M. M. Hassim, and M. Rehan, "A comparative review on deep learning models for text classification," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 1, pp. 325–335, 2020, doi: 10.11591/ijeecs.v19.i1.pp325-335.

[21] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, no. March, pp. 1–43, 2020, doi: 10.1016/j.physd.2019.132306.

[22] W. K. SARI, D. P. RINI, R. F. MALIK, and I. S. B. AZHAR, "Sequential Models for Text Classification Using Recurrent Neural Network," vol. 172, no. Siconian 2019, pp. 333–340, 2020, doi: 10.2991/aisr.k.200424.050.

[23] A. Gulli and S. Pal, "Long Short Term Memory - LSTM," *Deep Learn. with Keras*, pp. 187–195, 2017, doi: 10.1144/GSL.MEM.1999.018.01.02.

[24] A. Darmawahyuni, S. Nurmaini, and Sukemi, "Deep Learning with Long Short-Term Memory for Enhancement Myocardial Infarction Classification," *Proc. 2019 6th Int. Conf. Instrumentation, Control. Autom. ICA 2019*, no. August 2019, pp. 19–23, 2019, doi 10.1109/ICA.2019.8916683.

[25] C. Jiang, M. Maddela, W. Lan, Y. Zhong, and W. Xu, "Neural CRF model for sentence alignment in text simplification," *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pp. 7943–7960, 2020, doi: 10.18653/v1/2020.acl-main.709.

[26] J. Jasmir, S. Nurmaini, R. F. Malik, and B. Tutuko, "Bigram feature extraction and conditional random fields model to improve text classification clinical trial document," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 19, no. 3, pp. 886–892, 2021, doi: 10.12928/TELKOMNIKA.v19i3.18357.

[27] L. Yang, Y. Li, J. Wang, and Z. Tang, "Post text processing of chinese speech recognition based on bidirectional LSTM networks and CRF," *Electron.*, vol. 8, no. 11, p. 1249, 2019, doi: 10.3390/electronics8111248.

[28] Q. H. Pham, T. Nguyen, B. S. Hua, G. Roig, and S. K. Yeung, "JSIS3D: Joint semantic-instance segmentation of 3D point clouds with multi-task pointwise networks and multi-value conditional random fields," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 8819–8828, 2019, doi 10.1109/CVPR.2019.00903.

[29] C. Zuheros, S. Tabik, A. Valdivia, E. Martínez-cámara, and F. Herrera, "Deep recurrent neural network for geographical entities disambiguation on social media data," *Knowledge-Based Syst.*, 2019, doi: 10.1016/j.knosys.2019.02.030.

[30] 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," *Proc. - 2019 12th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2019*, pp. 0–4, 2019, doi: 10.1109/CISP-BMEI48845.2019.8965823.

[31] K. Y. Huang, Zhiheng ; Wei Xu, "Bidirectional LSTM-CRF Models for Sequence Tagging," 2019.

[32] J. Wei and K. Zou, "EDA: Easy data augmentation techniques for boosting performance on text classification tasks," *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 6382–6388, 2020, doi: 10.18653/v1/d19-1670.

[33] F. J. Moreno-Barea, J. M. Jerez, and L. Franco, "Improving classification accuracy using data augmentation on small data sets," *Expert Syst. Appl.*, vol. 161, 2020, doi: 10.1016/j.eswa.2020.113696.

[34] B. Tang, J. Hu, X. Wang, and Q. Chen, "Recognizing Continuous and Discontinuous Adverse Drug Reaction Mentions from Social Media Using LSTM-CRF," *Wirel. Commun. Mob. Comput.*, vol. 2018, 2018, doi: 10.1155/2018/2379208.

[35] Z. Wan, J. Xie, W. Zhang, and Z. Huang, "BiLSTM-CRF Chinese Named Entity Recognition Model with Attention Mechanism," *J. Phys. Conf. Ser.*, vol. 1302, no. 3, 2019, doi: 10.1088/1742-6596/1302/3/032056.

[36] Z. Mushtaq and S. F. Su, "Environmental sound classification using a regularized deep convolutional neural network with data augmentation," *Appl. Acoust.*, vol. 167, 2020, doi: 10.1016/j.apacoust.2020.107389.

[37] P. Wang, K. Nagrecha, and N. Vasconcelos, "Gradient-based Algorithms for Machine Teaching," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1387–1396, 2021, doi: 10.1109/CVPR46437.2021.00144.

[38] Y. Sun, B. Xue, M. Zhang, and G. G. Yen, "Evolving Deep Convolutional Neural Networks for Image Classification," *IEEE Trans. Evol. Comput.*, vol. 24, no. 2, pp. 394–407, 2020, doi: 10.1109/TEVC.2019.2916183.

[39] G. Rizos and K. Hemker, "Augment to Prevent : Short-Text Data Augmentation in Deep Learning for Hate-Speech Classification," pp. 991–1000, 2019.

[40] H. W. Anaıs Ollagnier, "Text Augmentation Techniques for Clinical Case Classification." 2020.

[41] M. Papadaki, "Data Augmentation Techniques for Legal Text Analytics by," no. October, pp. 1–33, 2017.

[42] S. Aiyar and N. P. Shetty, "N-Gram Assisted Youtube Spam Comment Detection," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 174–182, 2018, doi 10.1016/j.procs.2018.05.181.

DOI: https://doi.org/10.29207/resti.v7i6.5249