The Text Mining Model for Lecturer Performance Evaluation: A Comparative Study

Anita Ratnasari[1]*, Vina Ayumi[2], Mariana Purba[3], Wachyu Hari Haji[4], Handrie Noprisson[5], Marissa Utami[6]
Faculty of Informatics and Engineering, Universitas Dian Nusantara, Indonesia[1, 2, 5]
Faculty of Computer Science, Universitas Sjakhyakirti, Indonesia[3]
Department of Management, Bina Nusantara University, Indonesia[4]
Faculty of Engineering, Universitas Muhammadiyah Bengkulu, Indonesia[6]

***Abstract*—To support the evaluation of the teaching and learning process in higher education institutions, it is necessary to develop a text mining (TM) model. The aim of this research is to compare the performance of Long Short-Term Memory using Word Embedding Text to Sequence (WETS-LSTM), WETS-BiLSTM, WETS-CNN1D, and WETS-RNN, using four dataset categories including pedagogic, professional, personality, and social competency. This research has five main steps, including literature study, dataset collection, TM model development, and evaluation. Dataset is collected from Universitas Sjakhyakirti, Institut Teknologi dan Bisnis Palcomtech, Universitas Muhammadiyah Palembang, Universitas Bina Darma, AMIK Bina Sriwijaya and Politeknik Darusalam. The questionnaire distribution process initially yielded 6,170 responses with 6,164 valid across four competency categories, with total of 24,656 text data for analysis. Model of WETS-LSTM obtained the best performance overall, achieved the train accuracy of 96.65% and the highest test accuracy of 82.92%. The CNN1D with Word Embedding Text to Sequence (WETS-CNN1D) demonstrated good train accuracy with 96.73% but obtained lower test performance with 80.67%. The WETS with Recurrent Neural Network (WETS-RNN) obtained the weakest results, with a train accuracy of 95.88% and a test accuracy of 77.99%.**

Keywords*—*Text mining; CNN; LSTM; RNN; text-to-sequence

I. INTRODUCTION

The results of the evaluation of lecturer teaching and university services show clear information about the state of a university and the services offered to students [1], [2]. Evaluation serves as a method to assess the quality of education and the efficiency of higher education services which is an aspect of international university ranking assessment [3]–[5]. Through the results of the evaluation, higher education institutions can find out the programs, procedures, and service structures that suit the needs of students [6]–[8].

Usually at the end of each semester, students are asked to fill out a questionnaire designed to gather students' perceptions and feedback about the university's learning experience and facilities in terms of classes, teaching quality, library, and other services [9]–[11]. However, these questionnaires consist of Likert scale or multiple-choice questions (yes or no) that are usually filled out quickly and do not present the actual state of affairs [12], [13].

For this reason, it is necessary to evaluate lecturer teaching and university services using a special system with a feature to store assessment feedback in the form of text [14]–[16]. However, the evaluation of qualitative text feedback from students is difficult to do if it is done by manual analysis, especially for large-scale feedback datasets [17]–[19]. In fact, the results of qualitative feedback data analysis can provide valuable insights into the evaluation of lecturer teaching and university services [20]–[22]. To solve this problem, it is necessary to develop an architecture and evaluation model for lecturer teaching and university services based on text mining [23]–[25]. Research on the evaluation model of lecturer teaching and university services based on text mining (TM) has been carried out with various methods in the last four years using the text mining (TM) approach [14], [26]–[30]

The previous works utilized a variety of machine learning models for text analysis, focusing on different linguistic datasets and methodologies. Research by Grönberg et al. (2021) applied Latent Dirichlet Allocation (LDA) for topic modeling in English, while Misuraca et al. (2021) leveraged Support Vector Machine (SVM) for text classification in English as well [26], [27]. Research by Grljević et al. (2022) utilized a combination of Term Frequency-Inverse Document Frequency (TF-IDF) and K-Nearest Neighbor (KNN) techniques on Serbian text data, whereas Ren et al. (2023) explored the application of Word Embedding and Long Short-Term Memory (LSTM) for text analysis in Chinese [28] [29]. Abdi et al. (2023) combined Word Embedding (WE) with Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) for English text classification [30].

Based on research problem and previous works, several advancements were proposed in the development of a text mining-based evaluation system for lecturer teaching and university services. Researchers have focused on exploring various artificial intelligence models to analyze qualitative feedback, recognizing the limitations of traditional manual analysis, especially in large-scale datasets. The aim of this research was to develop and compare various text mining models for evaluating lecturer teaching and higher education services based on large-scale qualitative feedback. By utilizing different machine learning approaches, such as BiLSTM, LSTM, CNN-1D, and Recurrent Neural Network (RNN) with embedding text to sequence (WETS) techniques, the experiment was conducted to analyze a dataset of 6,164 student questionnaires across four categories, including pedagogic, professional, personality, and social competencies, into positive and negative interpretation result. The goal was to assess the effectiveness of these models for analyzing of lecturer performance and university services feedback by improving the accuracy of proposed model.

II. Related Work

Over the past four years, several studies investigated evaluation models of lecturer teaching and university services using the text mining (TM) approach. These studies employed various machine learning techniques on different linguistic corpora, including LDA for topic modeling and SVM for English text classification, TF-IDF with KNN classification on Serbian text, LSTM on Chinese data, and combinations of Word Embedding (WE) with CNN and BiLSTM for English text classification. The model has been implemented in the dataset of evaluation texts for lecturer teaching and university services can be seen in Table I.

TABLE I. State of the Art

Reference	Model	***Dataset***
[26]	Latent Dirichlet Analysis (LDA)	English
[27]	Support Vector Machine (SVM)	English
[28]	Term frequency–inverse document frequency (TF-IDF) – K-Nearest Neighbor (KNN)	Serbian
[29]	Word Embedding, Long Short-Term Memory (LSTM)	Chinese
[30]	Word Embedding, Convolutional Neural Networks (CNN)-BiLSTM	English
Proposed	Word Embedding Text to Sequence (WETS), LSTM	Indonesia

Based on previous, Grönberg et al. (2021) implemented Latent Dirichlet Allocation (LDA) for topic modeling to extract key themes from student feedback in English, while Misuraca et al. (2021) applied Support Vector Machine (SVM) for text classification in English, providing insights into the strengths and weaknesses of teaching and services. Similarly, Grljević et al. (2022) used a combination of Term Frequency-Inverse Document Frequency (TF-IDF) and K-Nearest Neighbor (KNN) techniques on Serbian text data, revealing valuable patterns and trends in the evaluation of academic services.

Moreover, other researchers, such as Ren et al. (2023), explored model of Word Embedding (WE) and Long Short-Term Memory (LSTM) for analyzing Chinese text data, further enhancing the ability to understand and interpret qualitative feedback and Abdi et al. (2023) advanced this approach by integrating Word Embedding with CNN-BiLSTM. This research proposed a text mining (TM) model that will be applied to the learning system. The proposed new text mining (TM) model comes from a combination of BiLSTM based on Word Embedding Text to Sequence (WETS).

III. Research Methodology

Data collection was based on the distribution of instruments at several universities in the city of Palembang, namely Universitas Sjakhyakirti, Institut Teknologi Bisnis Palcomtech, Universitas Muhammadiyah Palembang, Universitas Bina Darma, AMIK Bina Sriwijaya and Politeknik Darusalam. The questionnaire distribution process initially yielded 6,170 responses. After preprocessing, 6,164 valid questionnaires remained, each containing evaluations across four competency areas: pedagogic, professional, personality, and social. As a result, a total of 24,656 individual opinion data points were obtained for analysis.

The lecturer performance evaluation encompassed four key aspects: pedagogic, professional, personality, and social. The total evaluation included values for the overall performance, with 70% of the total representing the training data, 30% allocated for testing, and classifications for both positive and negative feedback. The composition of research dataset can be seen in Table II.

TABLE II. Research Dataset

Class	Train (70%)	Test (30%)	Total
Pedagogic	4,316	1,848	6,164
Professional	4,316	1,848	6,164
Personality	4,316	1,848	6,164
Social	4,316	1,848	6,164
Total	17,264	7,392	24,656

The initial stage in opinion mining is pre-processing with the aim of preparing text data which will later experience the next text data processing. This stage has five stages carried out, namely data cleansing, case folding, tokenizing, stop-word, tokenizing, and stemming. The summary of each of the five stages can be seen in Fig. 1.

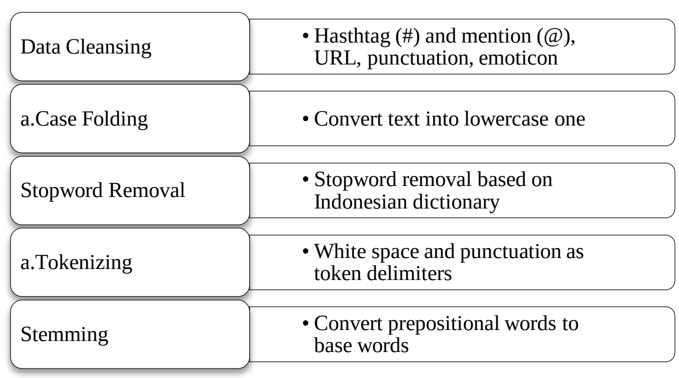

Fig. 1. Preprocessing phase.

At the data cleansing stage, the process of cleaning unnecessary words is carried out to reduce computational load, such as text containing html, links, and scripts. In addition, this stage also removes punctuation marks such as periods (.), commas (,) and also other punctuation marks. In this pre-processing process also applies the case folding method, which is the process of converting words into lower-case. The third stage is tokenization. This method is implemented to transform words contained in text into several sequences truncated by certain spaces or characters.

By applying a word that has a unique word from text data such as conjunctions, and possessive words. Less meaningful types of words will be removed, such as the words: I, and, or by using this method. The purpose of stop word removal is to reduce the burden on system performance, because the words to be taken are the words that are considered important. The last stage in the pre-process process is the stemming stage. The method at this stage is done by transforming the words in the text into basic words.

Next is to conduct experiments for model training and testing. The methods compared include BiLSTM, LSTM, CNN1D and RNN. Each model is carried out performance evaluation measure (PEM) or known as model performance evaluation measurement is an approach that aims to measure the performance or performance of the algorithm model. As for the experiment to be carried out as seen in Fig. 2.

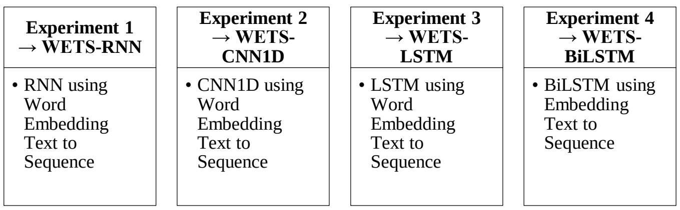

Fig. 2. Research experiment scenario.

The final stage is the evaluation stage that is carried out as a measurement of the performance value of the text mining ™ method that has been implemented. Performance evaluation measurement using confusion matrix, accuracy, precision, F1-score and recall to compare BiLSTM, LSTM, CNN-1D, and RNN with embedding text to sequence techniques performance.

IV. RESULT AND DISCUSSION

The study aimed to develop a text mining model for evaluating lecturer performance, with a specific focus on Indonesian text. Previous research primarily focused on datasets from other languages, such as English, Serbian, and Chinese, which evaluated different competencies. These studies incorporated personality, professional, pedagogical, and social competencies. This research differentiated itself by introducing a tailored evaluation model that includes all four competency categories and fill the gap in existing literature by providing an evaluation framework specifically designed for the Indonesian educational context as explained in Table III.

TABLE III. DATASET SPECIFICATION

Source & Dataset	Personality	Professional	Pedagogy	Social
English [26]		•	•	
English [27]		•	•	
Serbian [28]		•	•	•
Chinese [29]		•	•	
English [30]		•	•	•
Indonesia [Proposed]	•	•	•	•

The first experiment used RNN using Embedding Text to Sequence (WETS-RNN). The RNN model itself is in the form of a neural network designed to process the text sequence on the learning evaluation opinion. Using WETS is a dataset of learning opinion texts converted into numerical representations so that can be processed by the RNN model. The WETS-RNN model obtained a training accuracy of 0.9588 of the correct predictions which showed that the model was very good at recognizing patterns from the training data from the learning evaluation dataset and the test accuracy of 0.7799 which was lower than the training which showed that the WETS-RNN model was not fully generalized well to the test data.

The second experiment used CNN1D using Word Embedding Text to Sequence (WETS-CNN1D). The CNN1D model is based on a neural network used to process one-dimensional sequence data of student comment texts on learning. By using WETS, the students' comments were converted into vector representations so that the CNN1D model could capture the semantic relationships between words in the dataset text. The WETS-CNN1D model obtained a training accuracy of 0.9673 of the correct predictions indicating that the WETS-CNN1D model was very good at recognizing patterns from the training dataset while the test of 0.8067 showed good performance on the submission dataset being judged correct.

The third experiment used LSTM using Word Embedding Text to Sequence (WETS-LSTM). Model of LSTM model leverages neural networks designed to address sequence problems and long-term dependencies in processing the sequence of opinion and comment texts by capturing temporal dependencies. Then, the WETS is used to convert the order of the text numbers into a lower-dimensional vector that can be processed by the LSTM model. The training accuracy performance of the WETS-LSTM model is very good, which is 0.9665 which means that the model shows excellent performance on the training data, with 96.65% of the predictions correct while the test performance is 0.8292 or 82.92% which shows that the model has good generalization ability to comment text data that is different from the training data.

The fourth experiment used WETS-BiLSTM. The BiLSTM model is a variation of the LSTM that has the ability to process learning comment text data in two directions, namely from start to end and from end to end. The training accuracy performance of the WETS-BiLSTM model obtained a training accuracy of 0.9688 which shows good performance in the training data with 96.88% of the training data prediction results being declared correct. Then, the test accuracy obtained a value of 0.8234 or 82.34%, which is lower than the training accuracy but this value is the best value of the other models.

Based on experimental results, WETS-LSTM has good accuracy among other methods. The WETS -LSTM model is a model based on long short-term memory which can be used in WETS which allows the model to understand the meaning of a sentence based on word order. The results of the comparison of model can be seen in Table IV.

TABLE IV. THE PERFORMANCE COMPARISON OF MODEL

Method	Abbreviation	Train	Test
BiLSTM using Word Embedding Text to Sequence	WETS-BiLSTM	96.88%	82.34%
LSTM using Word Embedding Text to Sequence	WETS-LSTM	**96.65%**	**82.92%**
CNN1D using Word Embedding Text to Sequence	WETS-CNN1D	96.73%	80.67%
RNN using Word Embedding Text to Sequence	WETS-RNN	95.88%	77.99%

The initial experiment commenced with tokenization, where sentences were segmented into individual words. Each word was then assigned an index from a predefined vocabulary. For example, the sentence "Indonesian language" was divided into "Indonesian" and "language," which were mapped to specific indices within the vocabulary. This mapping transformed the text data into numerical form as part of the WETS phase. The WETS phase improved the performance of data processing and classification by utilizing models such as BiLSTM, LSTM, CNN1D, and RNN, with performance results, particularly accuracy, depicted in Fig. 3.

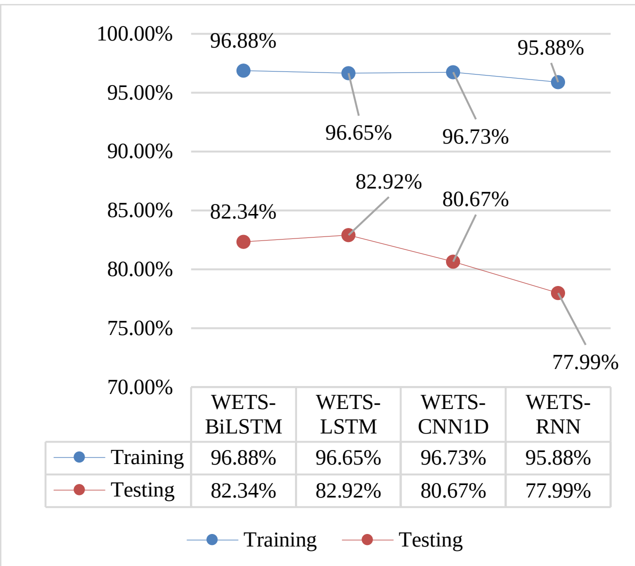

Fig. 3. Model performance results comparison.

Based on the experiment, WETS-LSTM was selected as the preferred text mining (TM) model for lecturer performance evaluation. The phase WETS involved padding to ensure uniformity in the input sequence length, as the model required inputs of a fixed length. In the given example, a sentence of length 2 was padded with zeros to extend the sequence to a length of 4. After padding, the sentence was represented as [1, 2, 0, 0]. The padded sequence was then processed through an embedding layer, where each token (word index) was mapped to a dense vector of fixed dimensions. This transformation converted the tokenized sequence into numerical representations suitable for the WETS-LSTM. The complete phase and result of data processing using WETS can be seen in Fig. 4.

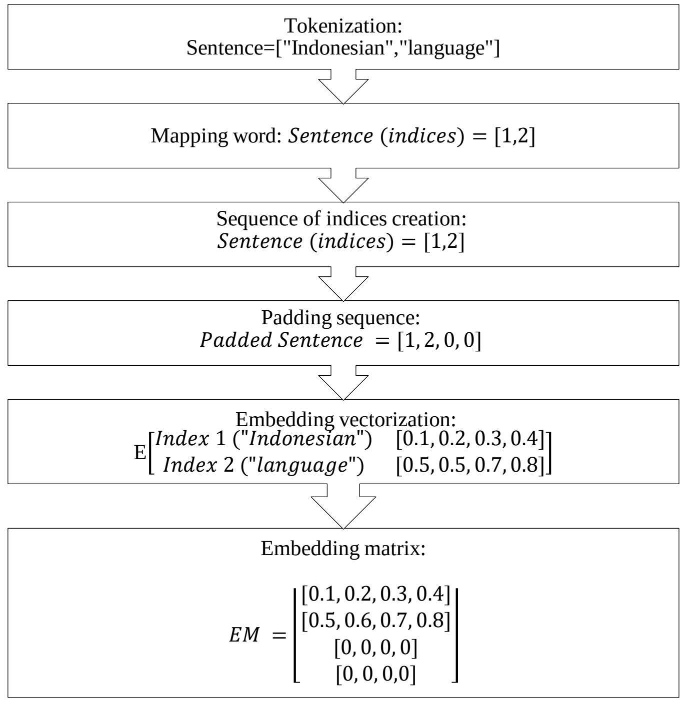

Fig. 4. Word Embedding Text to Sequence (WETS).

The input was processed through the LSTM, which operated on sequences and is capable of retaining long-term dependencies due to its forget f_i, input i_i, cell state c_i, and output gates o_l. At each time step *t*, the LSTM processed the current word embedding x_i, the previous h_{i-1} hidden state and the previou c_{i-1} cell state. The forget gate determined which information from the previous cell state should be weighted W_f, allowing the model to decide what aspects of the past memory to retain or forget for the current time step by using sigmoid activation function σ and bias term b_f. The explanation of LSTM can be expressed as in Eq. (1).

$$f_i = \sigma(W_f.[h_{i-1}, x_i] + b_f \quad (1)$$

$$h_0 = [0, 0, 0, 0],\ c_0 = [0, 0, 0, 0]$$

$$f_i = [0.6, 0.7, 0.5, 0.4]$$

Assuming that after calculating the forget gate at *t*=1 for the first word "Indonesian," the next step involved the input gate and the $\widetilde{C}_l$ cell candidate. The i_i input gate determined the amount of new information to be incorporated into the cell state, while the cell candidate generated potential new values for updating the cell state. The calculation of i_i and $\widetilde{C}_l$ for *t=1* can be expressed as in Eq. (2) and Eq. (3).

$$i_i = \sigma(W_i.[h_{i-1}, x_i] + b_i = [0.4, 0.3, 0.2, 0.1], \quad (2)$$

$$\widetilde{C}_l = tanh(W_c.[h_{i-1}, x_i] + b_c = [0.5, 0.4, 0.6, 0.7] \quad (3)$$

At *t=1*, for the first word "Indonesian," the cell state c_i was updated by combining the results of the forget gate f_i and the input gate i_i. This update process allowed the model to decide which parts of the previous cell state should be retained and which new information should be incorporated. The o_l output gate then determined which part of the updated cell state would be output as the hidden state. The h_l hidden state captured the relevant information from the current word and the context provided by the preceding words. The calculation of o_l and $\widetilde{h_l}$ for *t=1* can be expressed as in Eq. (4) and Eq. (5).

$$o_l = \sigma(W_o.[h_{i-1}, x_i] + b_o = [0.5, 0.6, 0.7, 0.8] \quad (4)$$

$$h_l = o_l.tanh\,(c_l) = [0.0985, 0.0714, 0.0833, 0.056 \quad (5)$$

This process was repeated for the second word "language" at *t=2* based value of embedding matrix. The previous hidden state $h_{l-1}$and cell state c_{l-1} were passed forward, and the new word embedding x_l for the current word was processed. Model LSTM generated a prediction based binary classification, which distinguished between positive and negative lecturer evaluations from the LSTM will be passed through a sigmoid activation function. This function will output a probability between 0 and 1. The sigmoid function is defined as Eq. (6).

$$\sigma(y) = \frac{1}{1+e^{-y}} \quad (6)$$

The output will be a value σ(y) between 0 and 1. If *σ(y)* is greater than or equal to 0.5, the feedback is classified as positive. Otherwise, the data is classified as negative, distinguishing between positive and negative lecturer evaluations. Based on this classification, the accuracy results were calculated for each category. The accuracy per epoch for the social feedback data was presented in Fig. 5.

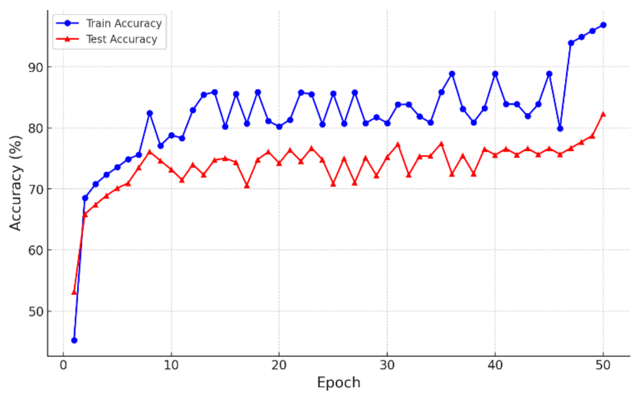

Fig. 5. Accuracy WETS-LSTM for dataset of social feedback.

The data used for the WETS-LSTM classification consisted of four distinct categories, i.e. pedagogic, professional, personality, and social, with each category having positive and negative samples. For the pedagogic category, the training set contained 2,389 positive and 1,927 negative samples, while the testing set had 1,024 positive and 824 negative samples. In the Professional category, the training data included 2,500 positive and 1,816 negative samples, with 946 positive and 902 negative samples in the testing data. The dataset of social category had 2,307 positive and 2,009 negative samples for training. Accuracy results were computed for each category with the detail of dataset as detailed in Table V.

TABLE V. THE DATASET FOR WETS-LSTM EVALUATION

Category	**Train**		**Test**	
	Positive	**Negative**	**Positive**	**Negative**
Pedagogic	2,389	1,927	1,024	824
Professional	2,500	1,816	946	902
Personality	2,290	2,026	975	873
Social	2,307	2,009	967	881
Total	7,486	7,778	3,912	3,480

The text mining model for evaluating lecturers calculated the total accuracy as the average of accuracy of each dataset category. The pedagogic accuracy obtained 85.00%, the professional accuracy obtained 80.00%, the personality accuracy obtained 81.00%, and the social accuracy obtained 83.36%. The total average accuracy was determined by calculating the mean of all dataset category accuracy, as in Eq. (7).

$$acc_{avg} = \frac{acc_{pedagogic} + acc_{professional} + acc_{personality} + acc_{social}}{4} \quad (7)$$

$$acc_{avg} = \frac{85.00\% + 80.00\% + 81.00\% + 83.36\%}{4} = 82.34\%$$

The calculation was applied to other models to assess the testing accuracy. The accuracy results for each model were then obtained. The BiLSTM using Word Embedding Text to Sequence (WETS-BiLSTM) model achieved an accuracy of 82.34%. The LSTM using Word Embedding Text to Sequence (WETS-LSTM) model performed slightly better, with an accuracy of 82.92%. Additionally, the CNN1D using Word Embedding Text to Sequence (WETS-CNN1D) and RNN using Word Embedding Text to Sequence (WETS-RNN) models achieved accuracy rates of 80.67% and 77.99%, respectively.

V. CONCLUSION

The aim of this research was to compare the performance of WETS-BiLSTM, WETS-LSTM, WETS-CNN1D, and WETS-RNN and involved four dataset categories, i.e. pedagogic, professional, personality, and social. The WETS-LSTM model exhibited the best performance overall, with a slightly lower training accuracy of 96.65%, but it achieved the highest test accuracy of 82.92%. The WETS-CNN1D model demonstrated good training accuracy with 96.73% but exhibited lower test performance, achieving 80.67%. The WETS-RNN produced the weakest results, with a training accuracy of 95.88% and a testing accuracy of 77.99%. While WETS-LSTM methods proved to be the most effective for both training and testing, WETS-RNN showed less satisfactory performance. Further analysis is required to improve the robustness of the proposed WETS-LSTM on larger datasets, and while the current study relied on limited data from a few private universities, efforts were undertaken to collect data from additional universities for more comprehensive analyses.

ACKNOWLEDGMENT

Author thanks to Universitas Dian Nusantara and Direktorat Riset, Teknologi, dan Pengabdian Kepada Masyarakat, Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia through research funding with 105/E5/PG.02.00.PL/2024;792/LL3/AL.04/2024,11/133/HSPK/VI/2024.

REFERENCES

[1] M. Purba et al., "Effect of Random Splitting and Cross Validation for Indonesian Opinion Mining using Machine Learning Approach," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 9, 2022.

[2] J. R. Hanaysha, F. B. Shriedeh, and M. In'airat, "Impact of classroom environment, teacher competency, information and communication technology resources, and university facilities on student engagement and academic performance," Int. J. Inf. Manag. Data Insights, vol. 3, no. 2, p. 100188, 2023.

[3] M. A. Fauzi, C. N.-L. Tan, M. Daud, and M. M. N. Awalludin, "University rankings: A review of methodological flaws," Issues Educ. Res., vol. 30, no. 1, pp. 79–96, 2020.

[4] L. A. Rozak et al., "Empirical evaluation of educational service quality in the current higher education system," Emerg. Sci. J., vol. 6, no. Special Issue, pp. 55–77, 2022.

[5] H. Noprisson, "A Survey of the Online Learning Implementation During COVID-19 Outbreak," Int. J. Recent Contrib. from Eng. Sci. IT, vol. 8, no. 4, p. 18, Dec. 2020.

[6] M. Purba, E. Ermatita, A. Abdiansah, V. Ayumi, H. Noprisson, and A. Ratnasari, "A Systematic Literature Review of Knowledge Sharing Practices in Academic Institutions," in 2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS, 2021, pp. 337–342.

[7] V. Kaushal, D. Jaiswal, R. Kant, and N. Ali, "Determinants of university reputation: conceptual model and empirical investigation in an emerging higher education market," Int. J. Emerg. Mark., vol. 18, no. 8, pp. 1846–1867, 2023.

[8] B. Thornhill-Miller et al., "Creativity, critical thinking, communication, and collaboration: assessment, certification, and promotion of 21st century skills for the future of work and education," J. Intell., vol. 11, no. 3, p. 54, 2023.

[9] P. Warfvinge, J. Löfgreen, K. Andersson, T. Roxå, and C. Åkerman, "The rapid transition from campus to online teaching–how are students' perception of learning experiences affected?," Eur. J. Eng. Educ., vol. 47, no. 2, pp. 211–229, 2022.

[10] F. M. Van der Kleij and A. A. Lipnevich, "Student perceptions of assessment feedback: A critical scoping review and call for research," Educ. assessment, Eval. Account., vol. 33, pp. 345–373, 2021.

[11] H. Singh, M.-H. Tayarani-Najaran, and M. Yaqoob, "Exploring computer science students' perception of ChatGPT in higher education: A descriptive and correlation study," Educ. Sci., vol. 13, no. 9, p. 924, 2023.

[12] Y.-J. Seo and K.-H. Um, "The role of service quality in fostering different types of perceived value for student blended learning satisfaction," J. Comput. High. Educ., vol. 35, no. 3, pp. 521–549, 2023.

[13] D. S. Asudani, N. K. Nagwani, and P. Singh, "Impact of word embedding models on text analytics in deep learning environment: a review," Artif. Intell. Rev., vol. 56, no. 9, pp. 10345–10425, 2023.

[14] A. H. Wani and R. Hashmy, "A supervised multinomial classification framework for emotion recognition in textual social data," Int. J. Adv. Intell. Paradig., vol. 24, no. 1–2, pp. 173–189, 2023.

[15] T. Shaik et al., "A review of the trends and challenges in adopting natural language processing methods for education feedback analysis," Ieee Access, vol. 10, pp. 56720–56739, 2022.

[16] N. A. Dahri, W. M. Al-Rahmi, A. S. Almogren, N. Yahaya, M. S. Vighio, and Q. Al-Maatuok, "Mobile-based training and certification framework for teachers' professional development," Sustainability, vol. 15, no. 7, p. 5839, 2023.

[17] A. V. Y. Lee, "Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation," Stud. Educ. Eval., vol. 77, p. 101250, 2023.

[18] J. P. Bernius, S. Krusche, and B. Bruegge, "Machine learning based feedback on textual student answers in large courses," Comput. Educ. Artif. Intell., vol. 3, p. 100081, 2022.

[19] R. Nawaz et al., "Leveraging AI and machine learning for national student survey: actionable insights from textual feedback to enhance quality of teaching and learning in UK's higher education," Appl. Sci., vol. 12, no. 1, p. 514, 2022.

[20] F. Gurcan and N. E. Cagiltay, "Research trends on distance learning: a text mining-based literature review from 2008 to 2018," Interact. Learn. Environ., vol. 31, no. 2, pp. 1007–1028, 2023.

[21] M. Á. Escotet, "The optimistic future of Artificial Intelligence in higher education," Prospects, vol. 54, no. 3, pp. 531–540, 2024.

[22] I. Nurhaida, I. Dhamanti, V. Ayumi, F. Yakub, and B. Tjahjono, "Hospital quality classification based on quality indicator data during the COVID-19 pandemic," Int. J. Electr. Comput. Eng., vol. 14, no. 4, pp. 4365–4375, 2024.

[23] K. Okoye, A. Arrona-Palacios, C. Camacho-Zuñiga, J. A. G. Achem, J. Escamilla, and S. Hosseini, "Towards teaching analytics: a contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification," Educ. Inf. Technol., pp. 1–43, 2022.

[24] S. Z. Salas-Pilco, K. Xiao, and X. Hu, "Artificial intelligence and learning analytics in teacher education: A systematic review," Educ. Sci., vol. 12, no. 8, p. 569, 2022.

[25] M. E. Dogan, T. Goru Dogan, and A. Bozkurt, "The use of artificial intelligence (AI) in online learning and distance education processes: A systematic review of empirical studies," Appl. Sci., vol. 13, no. 5, p. 3056, 2023.

[26] N. Grönberg, A. Knutas, T. Hynninen, and M. Hujala, "Palaute: An online text mining tool for analyzing written student course feedback," IEEE Access, vol. 9, pp. 134518–134529, 2021.

[27] M. Misuraca, G. Scepi, and M. Spano, "Using Opinion Mining as an educational analytic: An integrated strategy for the analysis of students' feedback," Stud. Educ. Eval., vol. 68, p. 100979, 2021.

[28] O. Grljević, Z. Bošnjak, and A. Kovačević, "Opinion mining in higher education: a corpus-based approach," Enterp. Inf. Syst., vol. 16, no. 5, p. 1773542, 2022.

[29] P. Ren, L. Yang, and F. Luo, "Automatic scoring of student feedback for teaching evaluation based on aspect-level sentiment analysis," Educ. Inf. Technol., vol. 28, no. 1, pp. 797–814, 2023.

[30] A. Abdi, G. Sedrakyan, B. Veldkamp, J. van Hillegersberg, and S. M. van den Berg, "Students feedback analysis model using deep learning-based method and linguistic knowledge for intelligent educational systems," Soft Comput., vol. 27, no. 19, pp. 14073–14094, 2023.