ITM Web of Conferences

All issues   Series   Forthcoming   About

Search   Menu

Table of Contents

Free Access to the whole issue

ITM Web of Conferences
Volume 71 (2025)

International Conference on Mathematics, its Applications and Mathematics Education (ICMAME 2024)

Export the citation of the selected articles Export
Select all

Preface of the Conference Proceeding of the 2nd International Conference on Mathematics, its Applications, and Mathematics Education (ICMAME) 2024

We are pleased to present the proceedings of the 2nd International Conference on Mathematics, its Applications, and Mathematics Education (ICMAME) 2024. This esteemed event was a collaborative effort organized by the Department of Mathematics, Faculty of Science and Technology, and the Department of Mathematics Education, Faculty of Teacher Training and Education, Universitas Sanata Dharma, Yogyakarta, Indonesia.

The conference was held on September 21, 2024, at Universitas Sanata Dharma, Indonesia, under the theme *"21st Century Mathematics and Mathematics Education."* This theme reflects the urgent need for innovative approaches to mathematics and mathematics education, aligned with the demands and opportunities of the 21st century. ICMAME 2024 brought together mathematicians, educators, and researchers to discuss emerging trends and share insights in these critical areas.

We were privileged to have distinguished keynote and invited speakers who enriched the conference with their expertise and perspectives. Our heartfelt thanks go to:

- **Assoc. Prof. Dr. Wanty Widjaya** (Deakin University, Australia)
- **Dr. rer. nat. Wolfgang Bock** (Linnaeus University, Sweden)
- **Dr. Martianus Frederic Ezerman** (Nanyang Technological University, Singapore)
- **Asst. Prof. Dr. Parkpoom Phetpradap** (Chiang Mai University, Thailand)
- **Veronica Fitri Rianasari, M.Sc., Ph.D.** (Universitas Sanata Dharma, Indonesia)

We extend our deepest gratitude to the organizing committee for their dedication and hard work, to the authors for their invaluable contributions, and to the reviewers for their diligent evaluations and constructive feedback. Without their combined efforts, this conference would not have been possible.

We hope this collection of proceedings, featuring contributions from authors across various countries (Indonesia, Philippines, Thailand, Japan, etc.), serves as a significant resource for researchers, educators, and practitioners. May it inspire future innovations, foster collaboration, and contribute to the advancement of mathematics and mathematics education for the betterment of our global community.

Thank you for being a part of the 2nd ICMAME 2024. We look forward to continuing this journey together and to exploring new horizons in the fascinating realms of mathematics and mathematics education.

Yogyakarta, 15 January 2025
Eko Budi Santoso
Chair, The 2nd ICMAME 2024

Statement of Peer review

In submitting conference proceedings to ITM Web of Conferences, I certify to the Publisher that I adhere to the **Policy on Publishing Integrity** of the journal in order to safeguard good scientific practice in publishing.

1. All articles have been subjected to peer review administered by the proceedings editors.

2. Reviews have been conducted by expert referees, who have been requested to provide unbiased and constructive comments aimed, whenever possible, at improving the work.

3. Proceedings editors have taken all reasonable steps to ensure the quality of the materials they publish and their decision to accept or reject a paper for publication has been based only on the merits of the work and the relevance to the journal.

International Conference on Mathematics, its Applications and Mathematics Education (ICMAME), 21 st September 2024, Yogyakarta, Indonesia

Proceedings editor(s): Hartono, Ph.D., Eko Budi Santoso, Ph.D., Prof.Dr.rer.nat Herry Pribawanto

January 29, 2025

Long Short-Term Memory and Bidirectional Long Short-Term Memory Algorithms for Sentiment Analysis of Skintific Product Reviews

Laurensia Simanihuruk[1]*,* and *Hari* Suparwito[1]*

[1]Department of Informatics, Sanata Dharma University, Indonesia

Abstract. In the era of ever-evolving digital technology, conducting customer sentiment analysis through product reviews has become crucial for businesses to improve their offerings and increase customer satisfaction. This research aims to analyze the sentiment of SKINTIFIC skincare products on the Shopee online store platform using advanced deep learning models: Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (Bi-LSTM). These models were evaluated using learning rate, number of units, and dropout rate. The dataset consists of 9,184 product reviews extracted through the Shopee API. The reviews were pre-processed using stemming, normalization, and stopword removal techniques. The Bi-LSTM model showed superior performance, achieving an average accuracy of 95.91% and an average F1 score of 95.82%, compared to the standard LSTM model. The optimal configuration for Bi-LSTM included a learning rate 0.01, 64 units, and a dropout rate 0.2. These findings underscore the effectiveness of Bi-LSTM in understanding and classifying consumer sentiment toward specific products.

1 Introduction

In the digital age, online shopping through e-commerce has become one of the main shopping trends. Modern society encourages online shopping for several reasons. One reason people choose to shop online is the ability to see customer reviews posted on online stores, which can help them consider buying products based on customer reviews [1].

The Shopee online store platform is one of the e-commerce platforms that offer a wide variety of products and services, and it has become the leading destination for consumers looking for various kinds of things they need [2]. One of the trending products in e-commerce is skincare products. This research will analyze user sentiment towards SKINTIFIC products on Shopee by looking at these online shopping trends. SKINTIFIC is one of the popular brands in Indonesia, and its flagship product is ceramide moisturizer. This brand has become popular among teenagers and adults. SKINTIFIC originated in Canada and was founded in 1957 by Kristen Tveit and Ann Kristin Stokke. Its products use skincare industry technology that creates innovative products with pure active ingredients, brilliant formulations, and Trilogy Triangle Effect (TTE) Technology [3].

---

* Corresponding author: shirsj@jesuits.net

Human interaction with computers has advanced significantly, thanks to Natural Language Processing (NLP). NLP plays a crucial role in many areas, including text generation, opinion mining, machine translation, named entity recognition (NER), speech recognition, and text summarization [4]. One of its most impactful applications is sentiment analysis, which predicts people's thoughts and emotions based on reviews, social media posts, and online forums. As businesses increasingly move online, unstructured data has become a treasure trove for insights, market research, and competitive analysis. Automated sentiment analysis is now vital for managing customer service, monitoring social media, and analyzing customer feedback. It simplifies processes by categorizing posts, survey responses, and scanned emails or documents [5].

In the real world, customers often share their opinions on product reviews. Many reviews are found with ratings that differ from their descriptions. For example, in the text of a review, a person gives a positive review while also giving a one-star rating, which means negative. A machine learning approach using NLP techniques is used to solve cases like this. The ability of NLP to perform product sentiment analysis has been studied in recent years. The use of several algorithms and comparing them with each other to see which algorithm can give optimal results in the study of product sentiment analysis.

Yadav, Vermar, and Katiyar [6] conducted a study by presenting a sentiment analysis method designed for Hindi e-commerce product reviews, utilizing a combination of the Long Short-Term Memory (LSTM) network and the Continuous Bag of Words (CBoW) model. Using five datasets, including Hindi SentiWordNet (HSWN), ABSA, and Twitter reviews, this research addresses the challenges posed by noisy data, word order dependency, and unbalanced datasets. Pre-processing steps such as tokenization, lemmatization, stopword removal, and max-pooling for dimensionality reduction significantly improve model performance. The proposed approach achieves an average accuracy of 87.71% across datasets, outperforming traditional methods such as SVM and state-of-the-art CNN-SVM models. In addition, the approach also exhibits superior precision, recall, and F1-score, with an F-score of 0.88. While the model demonstrates robustness in handling large data sets, it highlights the limitations of Hindi-specific text challenges, including polysemy and limited linguistic resources. The authors suggest further exploring solutions to these challenges in future work to refine the model's effectiveness in Hindi sentiment analysis.

Another study conducted by Yang, Li, Wang, and Sherratt [7] proposed a novel sentiment analysis model (SLCABG) for Chinese e-commerce product reviews, which combines sentiment lexicons with deep learning techniques, specifically CNN, BiGRU, and attention mechanisms. The model enhances sentiment features by using a lexicon to weight word vectors, followed by CNN to extract important features, BiGRU for contextual features, and attention mechanism to prioritize influential words. Experiments on a dataset of 100,000 reviews from the Chinese e-commerce site Dangdang show that the model outperforms traditional and deep learning models with higher classification accuracy, precision, recall, and F1-score. Key findings show that lexicon weighting improves sentiment feature extraction, optimal performance is achieved with a specific thesaurus size and number of iterations, and dropout regularization improves generalization. The paper concludes that the SLCABG model effectively classifies sentiment as positive or negative but lacks granularity for subtle sentiment differences, thus suggesting future research to refine sentiment categorization.

Another study that explored the application of a Bidirectional Long Short-Term Memory (Bi-LSTM) network for sentiment analysis on product reviews was conducted by Mahadevaswami and Swathi [8]. They used the Amazon Product Reviews dataset, explicitly focusing on the Mobile Electronics category, which contained 104,975 reviews. The main objective was to classify user reviews into positive or negative sentiments, utilizing deep learning techniques. The study employed several pre-processing steps, such as tokenization

and case-folding, to clean and standardize the text data. A text encoder transformed the text into word vectors, and padding ensured uniform input length for the Bi-LSTM model. The proposed architecture included embedding layers, two Bi-LSTM layers with 128 units each, and dense layers with ReLU activation to improve classification performance. Dropout layers were integrated to mitigate overfitting, and binary cross-entropy was used as the loss function. The model achieved significant accuracy improvements over baseline methods. During training, the model reached a final accuracy of 90.14% with a validation accuracy of 88.08% after five epochs. The test set achieved an overall accuracy of 91.4%, outperforming other models such as CNN and standard LSTM, which attained 87.62% and 86.56%, respectively. The study concluded that the Bi-LSTM model is highly effective for sentiment analysis due to its ability to process long-term dependencies in sequential data. Future work suggested incorporating more nuanced sentiment classifications (e.g., multi-tier emotions) and employing ensemble methods to enhance performance further.

Other research on NLP has also been done using customer reviews on women's clothing. Several deep learning algorithms were applied and compared to obtain optimal results. There are various neural network (NN) models, including CNN, RNN, Bi-LSTM, and ensemble models, and word embedding techniques such as Word2Vec, FastText, and BERT variants (RoBERTa and ALBERT). Data augmentation enhanced the dataset, and experiments were conducted with two sentiment ranking settings: 5-class (detailed) and 3-class (compressed) classification. Findings show that Bi-LSTM with Word2Vec embedding consistently outperforms other NN models, especially when using the augmented dataset and 3-class setting. RoBERTa performed best among the BERT variants, although ensemble NN models, specifically CNN-RNN-Bi-LSTM, achieved the highest accuracy (96%) and F1-score (91.1%). The deep learning models significantly outperformed traditional machine learning models such as SVM, Naïve Bayes, and Random Forest, which showed at least 20% lower accuracy. This study concludes that deep learning is more effective for sentiment rating prediction, especially with enhanced classes and augmented datasets. However, limitations include the potential impact of spam or fake reviews in the dataset and the exclusive focus on English reviews. Thus, future research should combine multilingual datasets and lexicon-based approaches to improve sentiment analysis further [9].

In our research, we will use review description data because the description can provide more context and depth about the product's user experience. Shopee e-commerce customer review data about SKINTIFIC cosmetic products is used to understand, analyze, and describe the opinions and feelings contained in product reviews given by consumers, both positive and negative. By categorizing customer reviews of SKINTIFIC products into two groups, namely positive and negative reviews, it is hoped that it can provide consideration to people who want to buy SKINTIFIC products. To get optimal results, two deep learning algorithms, namely LSTM and bi-LSTM, are used with various supporting parameters so that, in the end, it can be seen which algorithm with what parameters will be the most optimal.

2 Research Methodologies

The research phase begins with data collection; then, pre-processing is carried out to clean and prepare it for further analysis. Then, the process of labelling, tokenizing, and dividing training and testing data is carried out. This is followed by creating a testing scenario that aims to test the model with different situations to measure the extent to which the model can function in various contexts. Then, the model will be created according to the designed scenario, and the model will be trained using the data that has been prepared previously. In the last stage, the model will be evaluated to measure how much it performs according to the research objectives.

In general, the research is carried out in several stages, as seen in Fig. 1.
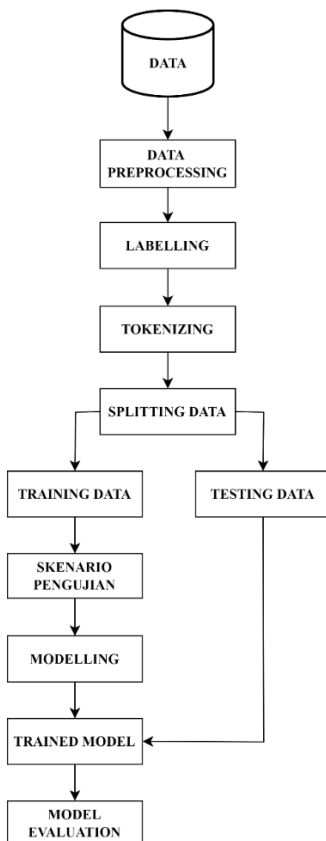
Fig. 1 Research workflow

2.1 Data

The data used are SKINTIFIC skin care product review data. The data was taken from the Shopee online store platform API on October 8, 2023. The data features are item ID, product name, comment, and rating. 20 SKINTIFIC skin care products were randomly selected and analyzed for reviews. The API used is the product API, which can be accessed via the following URL:

https://shopee.co.id/api/v2/item/get_ratings?filter=0&flag=0&itemid={item_id}&offset={offset}&shopid={shop_id}&type=0

Where item_id and shop_id was obtained from each product link, offset and limit were defined at the beginning.

The first process in data collection is to collect product links in a list and iterate over each product link [10]. In the iteration, it will extract the item_id and shop_id using regular expressions and compile the product's API URL. After that, it will request the Shopee API using the URL. After getting a response in the form of data, the data will be processed to be more structured, and the data retrieved includes item ID, product name, link, comment, and rating.

The total data obtained is 9,192, and the data that will be used for training is data whose word length is less than or equal to 100, so the remaining data is 9,184. Furthermore, the data will be stored in the form of data frames.

2.2 Pre-processing

Pre-processing steps include lowercasing, removing unnecessary characters, removing non-alphanumeric characters, normalizing slang words, stemming, and stopping word removal.

Lowercasing: Changing all letters into lowercase so that the text becomes insensitive to the difference in uppercase and lowercase letters. Lowercasing can improve classification performance without having to worry about inconsistencies in the text [11].

Remove Unnecessary Characters: Removing unnecessary characters is the process of cleaning the text from irrelevant elements such as punctuation marks, special symbols, and excessive spaces. This process improves the quality of text data, making it easier to process and analyze.

Remove Non-Alphanumeric: The remove non-alphanumeric process removes characters or symbols that are not letters or numbers in the tweet. This is done because these symbols have no information or meaning in the text. In the remove non-alphanumeric process, the program will separate words by connecting characters such as punctuation marks and remove all characters that are not letters or numbers. Non-alphanumeric characters can also represent regular expression patterns (Regex Pattern) but must not contain spaces [12].

Slang Word Normalization: Slang words refer to highly informal uses of language and expressions, which tend to be more figurative, playful, brief, lively, and short-lived than everyday language. In addition, slang words are not an official language listed in dictionaries. They are only a language style certain groups use [13].

Stemming: Stemming is a step where words are mapped and separated into their basic form [14]. The purpose is to facilitate the analysis and understanding of the language and to group words that have the same meaning into basic forms so that they can be considered a single unit in the text. For example, stemming can convert varied words such as "play," "game," and "play" into their simpler base form, namely "play."

Stop word Removal: Stop word removal is the process of removing words that have no meaning in a particular context, such as removing links, hash marks, and words that do not play an important role in sentiment. These stop words are words that do not affect sentiment. For example, words like "at", "by", "of", "a", and the like [15].

2.3 Labelling

The data collected do not yet have a label, so they need to be labelled. Labelling is the process of labelling a text based on the polarity of the sentiment contained in it. This research will do labelling with the output of 2 positive and negative labels. This research does not use neutral labels because neutral labels are between positive and negative, so there is a possibility that data in neutral labels can be identified as positive or negative [16].

Labelling is done using the Indonesian RoBERTa Base Sentiment Classifier. RoBERTa (Robustly optimized BERT approach) is a model that thoroughly understands text context and is used to handle various natural language processing tasks [17]. The Indonesian RoBERTa Base Sentiment Classifier is an optimized version of the RoBERTa model for classifying sentiment in text. The model is initially based on the Indonesian RoBERTa Base, which has gone through a previous training process, and then adapted to the indole SmSA dataset, consisting of Indonesian comments and reviews [18].

Tokenizing is the process of dividing text into smaller units, or "tokens," so that it is easy to process further.

2.4 Data splitting.

Data splitting will be done using the k-fold cross-validation method. The data obtained will be divided into training and test data using the k-fold approach. The dataset will be divided into k equal subsets. Each subset will be used alternately as test data, while the other subset will be used as training data. This process will be repeated k times, where each subset will be used as test data exactly once.

2.5 Modelling.

Modelling will be done using the Keras framework in Python. The model uses Sequential objects, a simple and linear model type. The models that will be created are LSTM and Bi-LSTM models. Then, create a parameter_search function to run the model using the specified parameter combination. First, it divides the data into five folds. After that, iterate each combination of bidirectional parameters, learning rate, num units, and dropout. For each parameter combination, it will be entered into the create_model function. Then, the model will be trained using training data from each fold with five epochs and a batch size of 64. Next, we will examine the model's performance by calculating Loss, accuracy, and f1-score.

2.6 Experiments

The experiment will compare the use of LSTM and Bi-LSTM layers, learning rate, num_units, and dropout parameters. Table 1 shows the parameter combinations to be tested, and Table 2 shows the model architecture.

Table 1. Parameter of experiments

Parameters	Values
bidirectional	[True, False]
learning_rate	[0,001, 0,01]
num_units	[64, 128, 256]
dropout	[0,2, 0,5]

Table 2. Model Architectures

Model	*Layer*
LSTM	Embedding (100) LSTM (num_units/ activation) GlobalMaxPooling1D () Dense(num_units/activation) Dropout () Dense(2/softmax)
Bi-LSTM	Embedding (100) Bi-LSTM (num_units/ activation) GlobalMaxPooling1D () Dense(num_units/activation) Dropout () Dense(2/softmax)

The dataset will be divided into five folds, each of which will take the performance results of the model evaluation in the form of Loss, accuracy, and f1-score. The model is trained

using several combinations of parameters determined above. The model evaluation results from each fold will then be averaged to obtain the average Loss, accuracy, and f1-score. The average results of the three evaluations metrics will be analyzed.

3 Results and discussions

The results are analyzed based on the five highest average accuracy values because it wants to know how precise the model is in classifying. The results of the five highest values can be seen in Table 3. The table highlights the performance of different configurations of LSTM and Bi-LSTM models based on average Loss, accuracy, and F1-score. These metrics are crucial for evaluating the models' ability to generalize and perform well on a given task.

Table 3. The best five results of the implementation of LSTM and Bi-LSTM

No.	Layer	Learning rate	Num Units	Dropout	Average Loss	Average accuracy	Average F1-score
1	Bi-LSTM	0,01	64	0,2	14,17%	95,91%	95,82%
2	Bi-LSTM	0,01	64	0,5	14,34%	95,79%	95,74%
3	LSTM	0,01	128	0,2	13,67%	95,78%	95,82%
4	Bi-LSTM	0,01	128	0,2	13,79%	95,57%	95,60%
5	Bi-LSTM	0,01	128	0,5	17,34%	95,54%	95,52%

A detailed analysis of the results reveals important trends and trade-offs between model complexity, dropout rates, and performance. Average Loss is a key metric for understanding how well the model fits the training data. The lowest Loss, 13.67%, is achieved by the LSTM configuration with 128 units and a dropout rate of 0.2 (Row 3). In contrast, the highest Loss, 17.34%, occurs in the Bi-LSTM configuration with 128 units and a higher dropout rate of 0.5 (Row 5). This trend suggests that increasing dropout, while helpful for reducing overfitting, can lead to worse model performance if over-applied.

When evaluating Average Accuracy, the Bi-LSTM configuration with 64 units and a 0.2 dropout rate (Row 1) stands out with the highest accuracy of 95.91%. Accuracy slightly decreases when dropout is increased, as seen in the Bi-LSTM configurations in Rows 1 and 2 (0.2 vs. 0.5 dropout) or Rows 4 and 5. This indicates that a moderate dropout value (e.g., 0.2) is generally better at balancing overfitting and performance.

In terms of the Average F1 Score, which balances precision and recall, two configurations achieve the highest value of 95.82%: Bi-LSTM with 64 units and 0.2 dropouts (Row 1) and LSTM with 128 units and 0.2 dropouts (Row 3). However, Row 1 also has a slightly higher accuracy than Row 3, making it the more reliable configuration overall.

Comparing the configurations, the best overall model is Row 1 (Bi-LSTM, 64 units, 0.2 dropout). This configuration offers the highest accuracy, a very high F1 Score, and a reasonably low average loss. While Row 3 (LSTM, 128 units, 0.2 dropouts) achieves the lowest Loss, it slightly lags in accuracy compared to Row 1. Since accuracy and F1 Score are often prioritized in classification tasks, Row 1 provides the best trade-off between performance metrics.

Of the five highest results, more models use the Bi-LSTM layer than the LSTM layer. According to the existing theory, the Bi-LSTM layer has a better understanding because it

can receive information from both directions. Based on average accuracy, the combination of a learning rate of 0.01 dominates the top five results; in other words, the use of a learning rate of 0.01 gives better results than the learning rate of 0.001.

The results of comparing the Num-unit parameter can be seen in the following table.

Table 4. The Num-unit comparison results where other parameters are equal

No.	Layer	Learning rate	Num Units	Dropout	Average Loss	Average accuracy	Average F1-score
1	Bi-LSTM	0,01	64	0,2	14,17%	95,91%	95,82%
4	Bi-LSTM	0,01	128	0,2	13,79%	95,57%	95,60%

From the two comparisons of the Num-unit 64 and 128 with the same Bi-LSTM layer, learning rate, and dropout, it can be seen that the model with the Num-unit 64 is superior in terms of average accuracy and average f1-score. In terms of average Loss, the model with num-units 128 is superior. In terms of model performance and accuracy, the model with the Num-unit 64 is better because it has superior accuracy.

Next, the Bi-LSTM model will be compared if the dropout parameter is changed. The following table shows the results of the bi-LSTM model if the dropout parameter is changed. The results of comparing the dropout parameters can be seen in Table 5.

Table 5. Dropout comparison results where other parameters are equal

No.	Layer	Learning rate	Num Units	Dropout	Average Loss	Average accuracy	Average F1-score
4	Bi-LSTM	0,01	128	0,2	13,79%	95,57%	95,60%
5	Bi-LSTM	0,01	128	0,5	17,34%	95,54%	95,52%

The table above shows that using dropout with a value of 0.2 gives better performance for all metrics. This shows that a 0.2 dropout performs better in suppressing Loss and improving average accuracy and f1-score than a 0.5 dropout for the same parameters. When the dropout is increased to 0.5, the number of disabled neurons becomes more significant, making it harder for the model to learn patterns from the data [13]. When dropout is applied, the neurons in the neural network are deactivated randomly, which can affect the calculation of the weight and bias of the neural network in performing classification.

In conclusion, the Bi-LSTM model with 64 units and a 0.2 dropout rate (Row 1) is the optimal configuration. Its high accuracy and F1 score make it particularly well-suited for tasks requiring reliable classification. Increasing dropout or model complexity (e.g., the number of units) can degrade performance, highlighting the importance of tuning these parameters for optimal results.

4 Conclusions

Based on the results of this study, it can be concluded that the Bi-LSTM model with a combination of learning rate parameters 0.01, num units 64, and dropout 0.2 provides the best results with a relatively small average loss (14.17%), high average accuracy (95.91%), and high average f1-score (95.82%). The average Loss of 14.17% indicates that the model has a relatively low prediction error rate. The average accuracy of 95.91% indicates that the

model performs accurately and is 95.91% correct in predicting sentiment in a sentence. An average f1-score of 95.82% indicates that the model has an excellent balance between precision (the model's ability to avoid false positives) and recall (the model's ability to find all true positives). The best model has been proven to use the optimal parameters according to the parameter comparison that has been done.

References

1. R. C. Indonesia, "Bukan harga, ini alasan orang Indonesia belanja di e-commerce," CNBC Indonesia, 16 Feb. 2023. [Online]. Available: https://www.cnbcindonesia.com/tech/20230216095033-37-414241/bukan-harga-ini-alasan-orang-indonesia-belanja-di-ecommerce

2. Shopee, "Shopee," [Online]. Available: https://shopee.co.id/

3. SKINTIFIC, "SKINTIFIC," [Online]. Available: https://skintific.com/id/pages/about-us

4. J. Eisenstein, Introduction to Natural Language Processing, MIT Press, 2019.

5. V. Raina and S. Krishnamurthy, "Natural language processing," in Building an Effective Data Science Practice, Springer, 2022, pp. 63–73.

6. V. Yadav, P. Verma, and V. Katiyar, "Long short-term memory (LSTM) model for sentiment analysis in social data for e-commerce products reviews in Hindi languages," Int. J. Inf. Technol., vol. **15**, no. 2, pp. 759–772, 2023. https://doi.org/10.1007/s41870-022-00867-4

7. L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning," IEEE Access, vol. **8**, pp. 23522–23530, 2020. https://doi.org/10.1109/ACCESS.2020.2969858

8. U. B. Mahadevaswamy and P. Swathi, "Sentiment analysis using bidirectional LSTM network," Procedia Comput. Sci., vol. **218**, pp. 45–56, 2023. https://doi.org/10.1016/j.procs.2023.02.006

9. V. Balakrishnan, Z. Shi, C. L. Law, R. Lim, L. L. Teh, and Y. Fan, "A deep learning approach in predicting products' sentiment ratings: a comparative analysis," J. Supercomput., vol. **78**, no. 5, pp. 7206–7226, 2022. https://doi.org/10.1007/s11227-021-03903-4

10. A. Kesely, "How to scrape Shopee user review with bs4," Stack Overflow, 2020. [Online]. Available: https://stackoverflow.com/questions/62485799/how-to-scrape-shopee-user-review-with-bs4

11. M. Işik and H. Dağ, "The impact of text pre-processing on the prediction of review ratings," Electr. Eng. Comput. Sci., vol. **28**, no. 3, pp. 1405–1421, 2020.

12. S. A. H. Bahtiar, "Perbandingan Naïve Bayes dan Logistic Regression dalam Sentiment Analysis pada Review Marketplace Menggunakan Rating Based Labelling," Master's Thesis, Universitas Islam Indonesia, Yogyakarta, 2023.

13. L. Saputra and L. Marlina, "An analysis of slang word used by Instagram account Plesbol," E-Journal of English Language and Literature, 2020.

14. Y. Pratama, L. D. Sianturi, R. D. Manalu, and D. F. Pangaribuan, "Implementation of sentiment analysis on Twitter using Naïve Bayes algorithm to know the people responses to debate of DKI Jakarta governor election," Physics: Conference Series, vol. **1337**, 2019. https://doi.org/10.1088/1742-6596/1337/1/012081

15. M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving text pre-processing for student complaint document classification using Sastrawi," in IOP Conference Series: Materials Science and Engineering, vol. **909**, p. 012123, 2020. https://doi.org/10.1088/1757-899X/909/1/012123

16. "RoBERTa: An Efficient Dating Method of Ancient Chinese Texts," ACM Digital Library, p. 293, 2023.

17. W. Wongso, "Indonesian RoBERTa-base sentiment classifier," Hugging Face, 2023. [Online]. Available: https://huggingface.co/w11wo/indonesian-roberta-base-sentiment-classifier

18. P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment analysis using Word2vec and Long Short-Term Memory (LSTM) for Indonesian hotel reviews," in Procedia Computer Science, vol. **179**, pp. 45–55, 2021. https://doi.org/10.1016/j.procs.2021.01.045