Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 8, No. 3, November 2025, pp. 569 – 578
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Big Data Analytics for Predicting Depression Risk in Generation Z: Integrating Self-Organizing Maps and Long Short-Term Memory

[1]Joy Nasten Sinaga, [2]Nuraina, [3*]Frans Mikael Sinaga, [4]Kelvin, [5]Nurhayati
[1,3]Departement of Information System, Universitas Mikroskil
[2]Departement of Informatics Engineering, Universitas Mikroskil
Email: [1]joy.sinaga@mikroskil.ac.id, [2]nuraina.nasution@mikroskil.ac.id [3]frans.sinaga@mikroskil.ac.id, [4]kelvin.chen@mikroskil.ac.id, [5]nurhayati@mikroskil.ac.id

Article Info
Article history:
Received Aug 07th, 2025
Revised Sep 03th, 2025
Accepted Sep 30th, 2025

Keyword:
Big Data Analytics
Depression Risk Prediction
Generation Z
Long Short-Term Memory
Mental Health
Self-Organizing Maps

ABSTRACT
Mental health issues among Generation Z are rising, with depression being one of the most significant challenges. Leveraging the capabilities of big data analytics and artificial intelligence, this study proposes a hybrid method combining Self-Organizing Maps (SOM) and Long Short-Term Memory (LSTM) networks to predict depression risk based on behavioral data. The SOM algorithm is utilized for clustering high-dimensional input data to uncover hidden patterns, while the LSTM network is employed to capture sequential dependencies over time. Data were collected from various digital platforms, processed, and analyzed to train and validate the proposed model. Results show that the SOM-LSTM framework significantly improves the accuracy and reliability of early depression risk detection compared to conventional models. This study contributes a scalable and adaptable model for mental health prediction that can assist in timely interventions for Generation Z

Corresponding Author:
Frans Mikael Sinaga,
Department of Information System,
Universitas Mikroskil
Jl. M.H Thamrin No.140, Pusat Ps., Kec. Medan Kota, Kota Medan, Sumatera Utara, Indonesia
Email: frans.sinaga@mikroskil.ac.id
DOI: http://dx.doi.org/10.24014/ijaidm.v8i3.38011

1. INTRODUCTION

Depression is an increasing mental health problem, especially among Generation Z (Gen Z). This generation has unique characteristics in social patterns, technology use, and psychological pressures that are different from previous generations. Changes in digital lifestyles, high interaction through social media, and increasingly complex academic and social pressures make Gen Z more vulnerable to mental health disorders, including depression. A recent phenomenon shows that Gen Z in Indonesia faces major challenges in the world of work. With around 9.9 million unemployed people or 22.5% of the population not pursuing higher education by August 2023, many of them are applying for jobs in the small business sector, such as seblak stalls and mobile phone shops. Although they are often perceived as lazy because they embrace the soft life, as many as 67% of them are actually willing to work hard, including overtime. However, the negative stigma regarding their work ethic remains high, with 40% of employers feeling that Gen Z is not work-ready due to a lack of adequate work ethic and communication skills [1], [2], [3], [4], [5], [6], [7], [8], [9], [10].

With the development of machine learning and artificial intelligence technologies, the use of Self-Organizing Map (SOM) and Long Short-Term Memory (LSTM) based prediction models offers a more sophisticated solution in analyzing multidimensional data and temporal patterns related to depression risk factors. Through the combination of these methods, research can identify hidden patterns in big data and provide more accurate predictions regarding individuals or groups at risk of depression [4], [5], [11].

This research is very important because until now there has been no prediction model specifically designed to identify the risk of depression in Gen Z with a big data approach. Generation Z has unique characteristics in social patterns, technology use, and psychological stress that are different from previous generations. With the increasing cases of depression among Gen Z, there is a need for a prediction model that is able to analyse risk factors comprehensively and adaptively. The results of this study are expected to help health workers, government, and related organisations in designing more effective prevention strategies and more targeted interventions.

SOM method is used to cluster multidimensional data that includes various depression risk factors, enabling explicit identification of unstructured patterns. After the clustering process, the LSTM model is applied to analyse the temporal patterns of these risk factors, enabling depression risk prediction based on historical data. The data used in this study includes information from social media, mental health surveys, as well as relevant psychological data. With this approach, the model will be able to identify individuals or groups at risk of depression as well as provide in-depth insights into the key factors contributing to the condition.

This research aims to develop a prediction model for depression risk in Gen Z through big data analysis by combining the SOM and LSTM methods. The model is designed to identify risk patterns based on social media activity, mental health history, sleep patterns, and other environmental factors. With the SOM-LSTM approach, it is expected to create an evidence-based early detection system that can be utilised by health workers, educational institutions, and the government for more appropriate and effective interventions [10], [12], [13].

2. RESEARCH METHOD

This research is based on a systematic framework that includes several key stages, viz, data preprocessing, risk feature extraction, data clustering using SOM, and temporal classification with LSTM. The data used includes information from social media, psychological surveys, and environmental factors relevant to Gen Z mental health. The clustering process with SOM aims to group individuals based on depression risk levels, which are then further analyzed temporally using LSTM to form a predictive model. Figure 1 illustrates the flow of the problem-solving approach applied in this study.

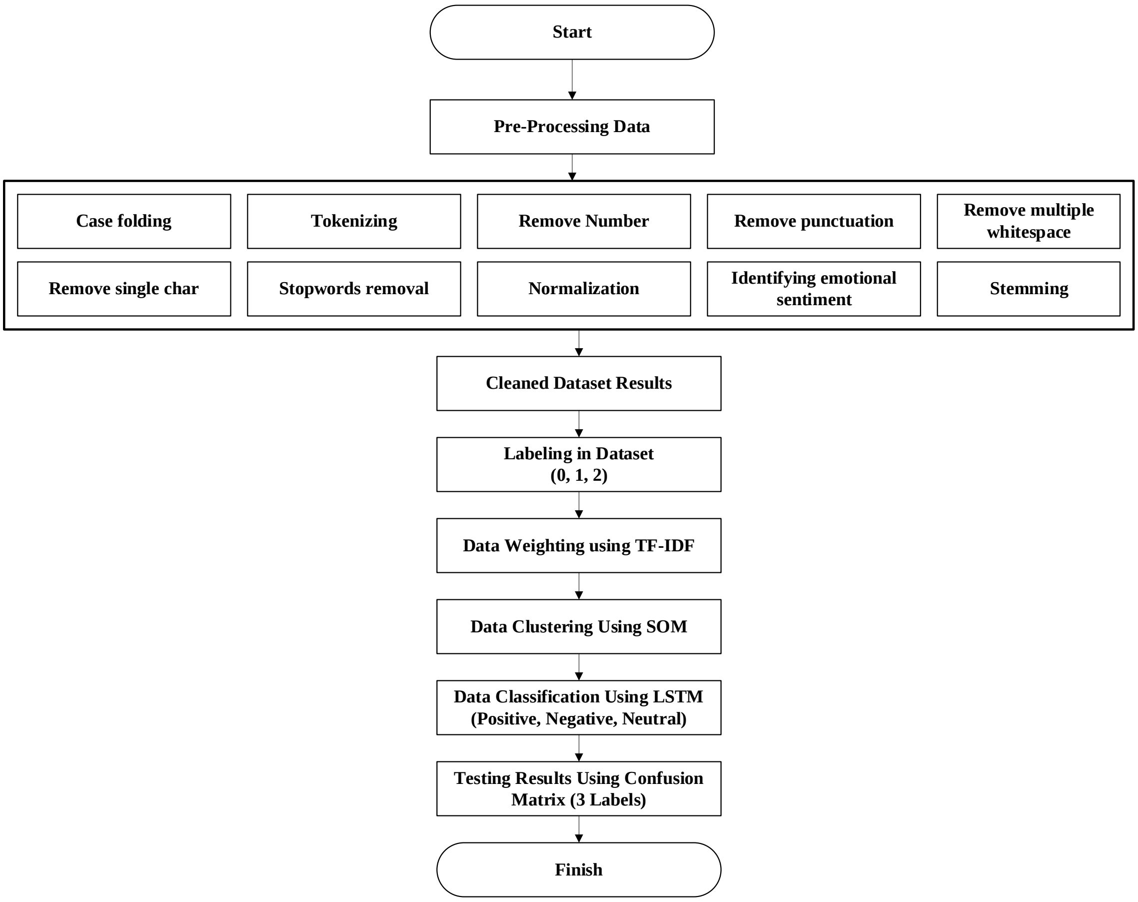

Figure 1. Research Methods

2.1 Data Collection

Data preparation involves crawling text data from multiple platforms, including Kaggle, Medium, Quora, TikTok, Twitter, and YouTube, using relevant keywords for this research. Table 1 lists the keywords used to collect text data from multiple platforms, as mentioned before. Additional datasets were incorporated, incorporating new keywords to expand on previous research.

Table 1. List of Dataset Keywords Used

No	Keyword Crawling Dataset	Total number of datasets
1	Merasa Lelah	November 2022 – 2023 25,000 datasets
2	Tidak ada semangat	
3	Hidup hampa	
4	Tidak Berguna	
5	Ingin Menyerah	November 2022 – 2024 27,000 datasets
6	Sendiri Terus	
7	Tidak ada yang peduli	
8	Kehilangan Motivasi	
9	Sedih banget	May 2024 – 2025 100,000 datasets
10	Stres Berat	
11	Cemas Berlebihan	
12	Susah tidur	
13	Ingin Mati	

2.2. Standard Text Data Preprocessing

1. Data Cleaning
   The data cleaning process involves removing unnecessary elements that could introduce noise into the analysis, including eliminating punctuation marks, emojis, special characters, and other non-textual symbols that do not contribute to the sentiment classification task. By standardizing the input format, the model is directed to focus on linguistically meaningful patterns.

2. Removal of URLs, Hashtags, and Usernames
   URLs, hashtags (#), and usernames (@) are removed from the dataset as they typically do not carry semantic information relevant to sentiment analysis. These components may act as noise and hinder the effective operation of these features.

3. Translation
   Text originally written in English is translated into Indonesian to ensure linguistic consistency across the dataset. This step enhances the model's ability to interpret context accurately by eliminating potential issues arising from code-mixed data.

4. Lowercasing
   All text is converted to lowercase to reduce case-sensitivity issues, ensuring that identical words in different cases are treated uniformly, thus improving token matching and model performance.

5. Truncation and Padding
   Long sequences are truncated, and short sequences are padded to achieve uniform input length. This step ensures compatibility with deep learning models that require fixed.

6. Lemmatization and Stemming
   To standardize word forms and improve text consistency, words are processed using lemmatization and stemming. Lemmatization converts words to their base forms based on linguistic context (e.g., "running" → "run"), while stemming removes affixes to obtain the root form (e.g., "running" → "run", but “better” still remained as “better”). By applying both techniques, the dataset retains meaningful word structures while reducing variations.

2.3. Initial Labeling with Sentiment Lexicon

In this step, sentiment values are combined using heuristic rules[14], [15], as described in Equation 1, to generate sentiment labels. This process ensures that sentiment aggregation aligns with contextual meaning, refining the dataset for improved classification performance.

$$S_{\text{very positive}} \sum_{i \in t}^{n} \text{very positive score}_i \quad (1)$$

$$S_{positive} \sum_{i \in t}^{n} \text{positive score}_i \tag{2}$$

$$S_{neutral} \sum_{i \in t}^{n} \text{neutral score}_i \tag{3}$$

$$S_{negative} \sum_{i \in t}^{n} \text{negative score}_i \tag{4}$$

$$S_{very\ negative} \sum_{i \in t}^{n} \text{very negative score}_i \tag{5}$$

The sentiment lexicon-based initial labeling process assigns sentiment scores based on predefined word polarities and contextual modifiers [15]. Each sentence is evaluated using a lexicon that captures sentiment intensity, allowing classification into five fine-grained sentiment categories: very positive, positive, neutral, negative, and very negative, as summarized in Table 2. This fine-grained labeling provides a more detailed sentiment distribution, enabling deeper insights into sentiment intensity and emotional variations across multi-platform data[16]. These labels also serve as training data, enabling SOM+LSTM to capture nuanced sentiment expressions with higher accuracy.

The sentiment labeling process begins by analyzing each word within the text and assigning a score based on its presence in predefined sentiment lexicons. If a word appears in the positive lexicon, its score is incremented by one, whereas if it is found in the negative lexicon, the score is decremented by one. Words that do not belong to either category do not contribute to the overall sentiment score. This word-level scoring serves as the foundation for determining the sentiment at the sentence level. The pseudo-algorithm for this process is presented in Algorithm 1.

Table 2. The Results of Labeling The Dataset Using Lexicon

Lexicon Scores	Lexicon Result	Tweet	Date	Username
-0.6	very negative	Aku merasa Lelah dan tidak ada semangat	Sun Jul 16 03:54:19 +0000 2023	JusDooIt
-0.2	negative	Hidup ini terasa hampa dan membosankan	Thu Jun 29 22:52:50 +0000 2023	leniptri__
0.3	positive	Senang sekali bisa main lagi dengan teman lama	Thu May 11 03:17:43 +0000 2023	bagussptyno
0.1	neutral	Tidak ada yang peduli padaku aku sendiri terus	Sat Jan 07 10:00:08 +0000 2023	morasaki12
0.4	Sangat positive	Alhamdulillah project kuliah selesai tepat waktu	Tue Jun 06 05:53:39 +0000 2023	xxbrightvcxx

Algorithm 1. Pseudo-Algorithm to Label Each Word [17]

```
For each word in the text do
  if the word is in the positive list, then
    sum = sum + 1
  else
    if the word is in the negative list, then
      sum = sum - 1
    End if
  End if
End for
```

Following the word-level sentiment scoring, the overall sentiment of a sentence is determined using a predefined scoring formula, as referenced in Equation 7. The sentiment initial label is determined based on predefined score thresholds. If the score is 0.6 or higher, the sentence is labeled as very positive (4). Scores between 0.4 and 0.6 are classified as positive (3), while scores ranging from -0.4 to 0.4 are labeled as neutral (2). Sentences with scores between -0.6 and -0.4 are categorized as negative (1), and those with scores of -0.6 or lower are assigned a very negative (0) label.

This fine-grained labeling approach enhances sentiment analysis precision by distinguishing varying degrees of sentiment intensity, as outlined in Algorithm 2. Where each label number represents the level of the sentiment label: 0 = very negative; 1 = negative; 2 = neutral; 3 = positive; and 4 = very positive.

2.4. Conversion of Text into Vector Data

Text is transformed into numerical vector representations using Word2Vec, a technique that captures semantic relationships between words by mapping them into a continuous vector space. This process enables the model to better understand word similarities and contextual meanings, improving the effectiveness of sentiment classification [17], [18] The resulting numerical embeddings facilitate further analysis and classification by allowing machine learning algorithms to process text-based data more efficiently. The outcomes of this transformation are presented in Table 3.

Table 3. Vector Data Result

Before Vectorized	Vector Data Result
Aku sering merasa lelah dan overthinking setiap malam, rasanya semua beban hidup terlalu berat	[-0.45, 0.12, 0.08, -0.31, -0.67, 0.29, -0.14, 0.05, 0.22]

Experimental results indicate that the inclusion of time-aware tokens improves the classification performance across various sentiment categories. The model demonstrates an increased ability to distinguish between sentiment fluctuations occurring at different times, particularly in social media data where user opinions may change based on external influences such as current events, holidays, or market trends. Comparative evaluations between SOM+LSTM with and without time-aware embedding reveal a notable improvement in sentiment classification accuracy, especially for categories with high temporal sensitivity, such as political discussions or real-time product reviews.

Algorithm 2. Pseudo-Algorithm to Fine-Grained Labeling Each Sentence [17]

```
If score >= 0.6 then
  label = 4
else
  if 0.4 <= score < 0.6 then
    label = 3
  else
    if -0.4 >= score > -0.6 then
      label = 2
    else
      if -0.4 >= score > -0.6 then
        label = 1
      else
        if score <= -0.6 then
          label = 0
        End if
      End if
    End if
  End if
End if
```

2.5. Conversion of Text into Vector Data

Text is tokenized according to SOM+LSTM linguistic structure, ensuring proper segmentation for effective processing. The updated SOM+LSTM tokenizer recognizes time-aware tokens as distinct entities, preventing them from being broken into subwords. This modification preserves the temporal context embedded in the input.

During encoding, each token, including the new time-aware markers, is mapped to an embedding vector. Since these tokens were not in the original SOM+LSTM vocabulary, their embeddings are either initialized randomly or derived from semantically related words. This enables SOM+LSTM to incorporate temporal context dynamically, refining sentiment predictions based on time-sensitive factors. The outcomes of this tokenization process are presented in Table 4.

Table 4. Tokenizing Result

Before tokenizing text	After tokenizing text
Aku sering merasa lelah dan overthinking setiap malam, rasanya semua beban hidup terlalu berat	['aku', 'sering', 'merasa', 'lelah', 'dan', 'overthinking', 'setiap', 'malam', 'rasanya', 'semua', 'beban', 'hidup', 'terlalu', 'berat']

2.6. Classification using SOM+LSTM

SOM+LSTM model trained using masked language modeling with an Indonesian language dataset, is used for sentiment analysis [17], [18], [19], [20], [21]. Fine-tuning on fine-grained labeled data enables SOM+LSTM to classify sentiment into five categories with high accuracy. Its deep contextual understanding allows it to capture subtle sentiment variations across different languages and domains. The stages of the SOM+LSTM model include:

1. Data Processing for SOM+LSTM
   Each token needs to be tokenized or transformed into vector representations using embedding techniques.

$$V_t = Trans\ (W_t X_a + b_t) \tag{6}$$

2. Initial Fine-Grained Labeling with Sentiment Lexicon
   Sentiment lexicon-based labeling is a widely used approach for sentiment classification, where words and phrases are assigned sentiment scores based on predefined polarity values [34].

$$\sum_{i=1}^{k} \frac{(p-n)_i}{k} \tag{7}$$

   Where k represents the total number of words in the text that are listed in the lexicon, and p-n denotes the sentiment polarity score associated with each word in the lexicon, which is determined based on its occurrence in positive and negative categories [37].

3. SOM+LSTM-specific Tokenization
   The tokenized text, including standard tokens and newly introduced time-aware tokens, is processed using the SOM+LSTM tokenizer.

4. Feeding Data into the SOM+LSTM Model
   The SOM+LSTM model processes text and generates vector representations for each token in the text.

$$\mathrm{V_b} = \mathrm{Bert}(\mathrm{W_b X_b} + \mathrm{b_b}) \tag{8}$$

5. Self-Attention
   Self-attention to produce better representations of each word within the context of a sentence or text.

$$\mathrm{Attention}(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{9}$$

   Where Q, K, V are matrices of query, key, and value associated with tokens in the input.

6. Fine-tuning SOM+LSTM
   Fine-tuning SOM+LSTM involves retraining the model on domain-specific sentiment data, optimizing it for better classification performance while maintaining contextual integrity.

7. Sentiment Classification
   The SOM+LSTM model will process the text and classify it into 5 fine-grained sentiment categories based on the generated vector representations.

2.7. Data Testing

The model's performance is evaluated using a confusion matrix, which assesses accuracy based on four components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TN represents correctly classified negative data, FP indicates negatives misclassified as positives, TP denotes correctly identified positives, and FN reflects positives misclassified as negatives. Accuracy testing, described in Equation 10, is utilized to measure the effectiveness of the classification method.

3. RESULTS AND ANALYSIS

The findings of this study include data collection, preprocessing, sentiment labeling using a Sentiment Lexicon, embedding time-aware token, and fine-grained sentiment classification using the SOM+LSTM model. The preprocessing phase involves text normalization, partial data translation of English terms into Indonesian, tokenization, truncation, and padding to optimize input representation for SOM+LSTM.

3.1 Fine-Tuned SOM+LSTM Model

Fine-tuning SOM+LSTM optimizes its parameters using task-specific training data, enhancing its ability to capture complex linguistic patterns. This process improves accuracy, precision, and recall by aligning the model's performance with task-specific nuances.

The dataset consists of approximately 24,000 samples with a distribution across fine-grained sentiment categories, as detailed in Table 5. Understanding the label distribution, as visualized in Figure 2, is crucial for evaluating SOM+LSTM performance in handling imbalanced sentiment classes and addressing potential impacts on model training.

Table 5. Fine-Grained Sentiment Label Distribution

Label	Training	Validation	Testing
Very Positive	1,153	346	149
Positive	920	276	118
Neutral	5,684	1,705	731
Negative	2,213	663	285
Very Negative	1,419	426	182
Total Data	11,389	3,416	1,465

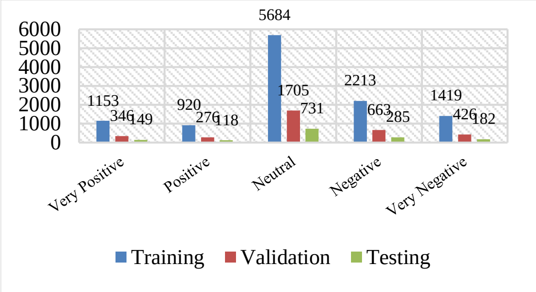

Figure 2. Distribution of Fine-Grained Sentiment Labels

Figure 3 further illustrate the model's convergence and learning progress over training epochs.

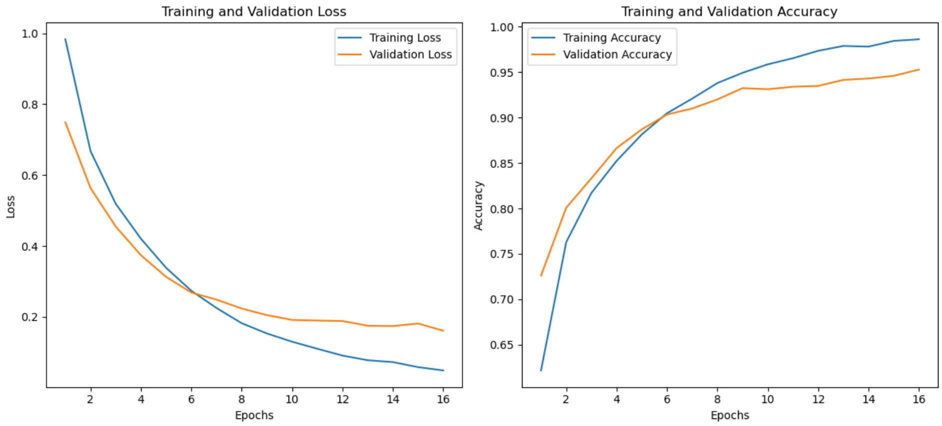

Figure 3. Validation Data Graph of the SOM+LSTM Model

Table 6 presents the classification performance metrics from confusion matrix in Figure 4, including accuracy, precision, recall and f1-score during testing mode, demonstrating the effectiveness of fine-tuned SOM+LSTM.

Table 6. Fine-Tuned SOM+LSTM Model Performance Metrics on Testing Mode

Label	Precision	Recall	F1-Score	Total Data
Very Positive	0.99	0.94	0.96	149
Positive	0.91	0.97	0.94	118
Neutral	0.98	0.98	0.98	731
Negative	0.94	0.95	0.95	285
Very Negative	0.94	0.95	0.95	182
	Accuracy			96.38%

3.2 Analysis of Dataset Size on Model Performance

To evaluate the impact of training data size on model performance, experiments were conducted using different subsets of the dataset for training, validation, and testing with varying dataset sizes and label distributions. The results, summarized in Table 7 indicate that increasing the training dataset size generally improves classification accuracy.

Table 7. Performance Comparison Across Different Dataset Sizes

Raw Data	Data Size			
	850	2,000	25,000	36,000
Training	577	980	17,091	21,867
Validation	159	294	5,128	6,747
Testing	88	126	2,198	2,624
Accuracy	100%	75%	96%	89%

The perfect accuracy (100%) on the smallest dataset (850 raw data samples) suggests overfitting, as the model memorizes patterns instead of generalizing. Limited data reduces variation, making classification easier but less robust. However, beyond a certain threshold, additional data showed diminishing performance gains. These findings highlight the importance of striking a balance between dataset size, label distribution, and computational efficiency when fine-tuning SOM+LSTM for sentiment analysis.

The conclusion of this study is that integrating a Sentiment Lexicon with fine-grained and dynamic sentiment labeling significantly enhances sentiment classification performance. By categorizing text into five sentiment classes very positive, positive, neutral, negative, and very negative this approach provides a more detailed representation of sentiment compared to traditional classification methods.

Additionally, the SOM+LSTM contextual learning capabilities, combined with lexicon-based labeling, improve sentiment analysis accuracy, enabling better detection of nuanced sentiment expressions across diverse Indonesian social media data. The preprocessing techniques employed in this study further facilitate text normalization, ensuring cleaner inputs for the SOM+LSTM model. Furthermore, the incorporation of time-aware tokenization enables dynamic sentiment analysis, allowing the model to account for sentiment evolution based on temporal context.

The findings demonstrate that leveraging fine-grained and dynamic sentiment analysis, rather than relying solely on binary or ternary sentiment categories, enhances the robustness of sentiment classification. Future work may explore further improvements in lexicon-based sentiment scoring and the potential integration of multi-modal sentiment analysis to expand classification performance, particularly in tracking sentiment trends over time.

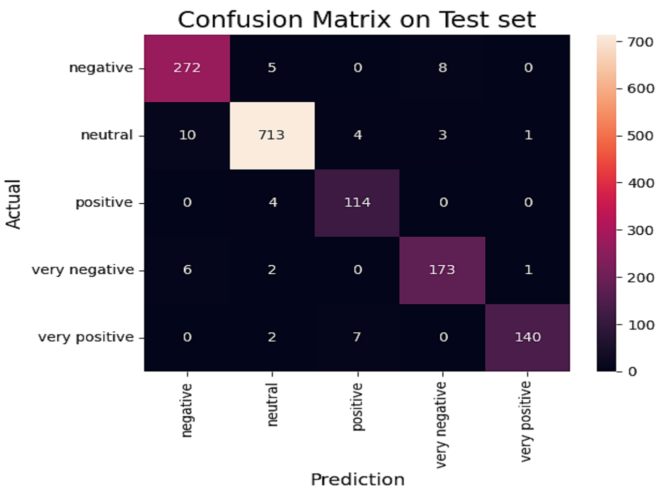

Figure 4. Confusion Matrix on Testing Set

4. CONCLUSION

This study demonstrates that integrating Sentiment Lexicon-based fine-grained labeling with the SOM+LSTM framework enhances sentiment classification accuracy for Indonesian social media data, achieving a testing accuracy of 96.38% with balanced precision, recall, and F1-scores across five sentiment categories. The combination of SOM clustering, LSTM temporal modeling, and time-aware tokenization effectively captures nuanced emotional variations and sentiment evolution over time. The proposed approach outperforms conventional binary or ternary classifications, offering a more detailed sentiment representation. Future work will focus on enhancing lexicon scoring and multimodal sentiment analysis to further strengthen robustness and trend tracking capabilities.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the Direktorat Penelitian dan Pengabdian kepada Masyarakat, Direktorat Jenderal Riset dan Pengembangan Kementerian Pendidikan Tinggi, Sains dan

Teknologi, for providing funding support for this research through the Penelitian Dosen Pemula (PDP) grant scheme.

REFERENCE

[1] A. G. Reece and C. M. Danforth, "Instagram photos reveal predictive markers of depression," EPJ Data Sci, vol. 6, no. 1, p. 15, Dec. 2017, doi: 10.1140/epjds/s13688-017-0110-z.
[2] S. Chancellor, M. L. Birnbaum, E. D. Caine, V. M. B. Silenzio, and M. De Choudhury, "A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media," in Proceedings of the Conference on Fairness, Accountability, and Transparency, New York, NY, USA: ACM, Jan. 2019, pp. 79–88. doi: 10.1145/3287560.3287587.
[3] W. Bin Tahir, S. Khalid, S. Almutairi, M. Abohashrh, S. A. Memon, and J. Khan, "Depression Detection in Social Media: A Comprehensive Review of Machine Learning and Deep Learning Techniques," IEEE Access, vol. 13, pp. 12789–12818, 2025, doi: 10.1109/ACCESS.2025.3530862.
[4] R. A. CALVO, D. N. MILNE, M. S. HUSSAIN, and H. CHRISTENSEN, "Natural language processing in mental health applications using non-clinical texts," Nat Lang Eng, vol. 23, no. 5, pp. 649–685, Sep. 2017, doi: 10.1017/S1351324916000383.
[5] A. Husseini Orabi, P. Buddhitha, M. Husseini Orabi, and D. Inkpen, "Deep Learning for Depression Detection of Twitter Users," in Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, Stroudsburg, PA, USA: Association for Computational Linguistics, 2018, pp. 88–97. doi: 10.18653/v1/W18-0609.
[6] G. Shen et al., "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution," in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, California: International Joint Conferences on Artificial Intelligence Organization, Aug. 2017, pp. 3838–3844. doi: 10.24963/ijcai.2017/536.
[7] A. Williams, "Twenge, J. M. (2017). iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy and Completely Unprepared for Adulthood. New York, NY: Atria. ISBN: 978-1-5011-5201-6 paperback. 342 pp.," Fam Consum Sci Res J, vol. 48, no. 3, pp. 290–293, Mar. 2020, doi: 10.1111/fcsr.12345.
[8] A. Turner, "Generation Z: Technology and Social Interest," The Journal of Individual Psychology, vol. 71, no. 2, pp. 103–113, Jun. 2015, doi: 10.1353/jip.2015.0021.
[9] Michele Parmelee, "Deloitte's 2023 Gen Z and Millennial Survey reveals workplace progress despite new setbacks," www.deloitte.com.
[10] Grace Huckins, "An AI Used Facebook Data to Predict Mental Illness," https://www.wired.com/.
[11] T. Kohonen, Self-Organizing Maps, vol. 30. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. doi: 10.1007/978-3-642-56927-2.
[12] D. Muzafar, F. Y. Khan, and M. Qayoom, "Machine Learning Algorithms for Depression Detection and Their Comparison," Jan. 2023.
[13] M. Kerasiotis, L. Ilias, and D. Askounis, "Depression detection in social media posts using transformer-based models and auxiliary features," Soc Netw Anal Min, vol. 14, no. 1, p. 196, Sep. 2024, doi: 10.1007/s13278-024-01360-4.
[14] T. Kohonen, Self-Organizing Maps, vol. 30. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. doi: 10.1007/978-3-642-56927-2.
[15] J. Tervonen et al., "Personalized mental stress detection with self-organizing map: From laboratory to the field," Comput Biol Med, vol. 124, p. 103935, Sep. 2020, doi: 10.1016/j.compbiomed.2020.103935.
[16] D. Muzafar, F. Y. Khan, and M. Qayoom, "Machine Learning Algorithms for Depression Detection and Their Comparison," Jan. 2023.
[17] M. Savadkoohi, T. Oladunni, and L. A. Thompson, "Deep Neural Networks for Human's Fall-risk Prediction using Force-Plate Time Series Signal," May 21, 2021. doi: 10.1101/2021.05.19.21257466.
[18] P. Gopika, "Depression Prediction Using LSTM and SVM," International Research Journal of Innovations in Engineering and Technology (IRJIET), vol. 8, no. 2, pp. 128–132, 2024, doi: 10.47001/IRJIET/2024.802018.
[19] A. Safayari and H. Bolhasani, "Depression Diagnosis by Deep Learning Using EEG Signals: A Systematic Review," Nov. 2021, doi: 10.20944/preprints202107.0028.v2.
[20] M. Kanahuati-Ceballos and L. J. Valdivia, "Detection of depressive comments on social media using RNN, LSTM, and random forest: comparison and optimization," Soc Netw Anal Min, vol. 14, no. 1, p. 44, Feb. 2024, doi: 10.1007/s13278-024-01206-z.
[21] X. Lan, Y. Cheng, L. Sheng, C. Gao, and Y. Li, "Depression Detection on Social Media with Large Language Models," Mar. 2024.

BIBLIOGRAPHY OF AUTHORS

Joy Nasten Sinaga is currently serving as a faculty member in the Information Systems Study Program at Mikroskil University. He earned his Master's degree in Information Technology at Mikroskil University. He has research interests in Artificial Intelligence, Data Science, and expertise in graphic design and system design.

Nuraina is currently serving as a faculty member in the Information Systems Study Program at Mikroskil University. She has research interests in Artificial Intelligence.

Frans Mikael Sinaga, S.Kom., M.Kom., Lecturer at the Department of Informatics Engineering, Faculty of Informatics, Mikroskil University, Medan. Born in Penggalangan village on October 24, 1993. The author is the third child out of 4 siblings of Mr. Waristo and Mrs. Linda. The author completed a Bachelor's degree (S1) in Informatics Engineering and a Master's degree (S2) in Information Technology at STMIK Mikroskil Medan. The author has written several book titles such as Introduction to Computer Networks and Data Mining. In addition to writing books, the author has also conducted several research projects in the fields of Data Science and Computer Vision.

Kelvin, S.Kom., M.Kom., The author is a Software Engineer and Lecturer in the Informatics Engineering, Faculty of Informatics, Mikroskil University, Medan. He completed his bachelor's degree in Informatics Engineering at STMIK Mikroskil in 2018. Then, in 2020, the author pursued postgraduate studies in Information Technology at Mikroskil University and successfully completed them in 2021. The courses he has taught include Introduction to Algorithms, Web Design, C Programming, Object-Oriented Programming, Back-End Web Development, Artificial Intelligence, and Natural Language Processing. In addition to his academic involvement, the author has over 5 years of experience as a software engineer, working for both domestic and international companies. For more information, visit the author's LinkedIn page at https://www.linkedin.com/in/kelvinchen96.

Nurhayati is currently serving as a faculty member at Mikroskil University in the Informatics Engineering Study Program. Her areas of expertise include Data Science, Deep Learning, Image Detection, and Artificial Intelligence.