**Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)**
Vol. 5, No. 2, September 2022, pp. 76 – 82
p-ISSN: 2614-3372 | e-ISSN: 2614-6150 

# Classification Between Suicidal Ideation and Depression Through Natural Language Processing Using Recurrent Neural Network

**[1]Rhenaldy, [2]Ladysa Stella Karenza, [3]Hidayaturrahman, [4]Muhamad Keenan Ario**
[1,2,3,4]Computer Science Department, Bina Nusantara University
Email: [1]rhenaldy@binus.ac.id, [2]ladysa.karenza@binus.ac.id,
[3]hidayaturrahman@binus.ac.id, [4]muhamad.ario@binus.ac.id

**Article Info**
***Article history:***
Received Jun 16th, 2022
Revised Jul 27th, 2022
Accepted Aug 30th, 2022

***Keyword:***
Dimensionality Reduction
Clustering
NLP
RNN
Suicide
Text

**ABSTRACT**
The use of machine learning has been implemented in various ways, including to detect depression in individuals. However, there is hardly any research done regarding classification between suicidal ideations and depression among individuals through text analysis. Differentiating between depression and suicidal ideation is crucial, considering the difference in treatment between the two mental illness. In this paper, we propose a detection model using Recurrent Neural Network (RNN) in the hopes to improve previous models made by other researchers. By comparing the proposed model with the previous works as the baseline model, we discovered that the proposed model (RNN) performed better than the baseline models, with the accuracy of 86.81%, precision of 97.13%, recall score of 94.69%, f1 score of 95.90%, and area under the curve (AUC) score of 92.84%.


***Corresponding Author:***
Rhenaldy,
Departement of School of Computer Science,
Bina Nusantara University,
9 K. H. Syahdan Road, Kemanggisan, West Jakarta 11480, Indonesia.
Email: rhenaldy@binus.ac.id

**DOI: http://dx.doi.org/10.24014/ijaidm.v5i2.17485**

## 1. INTRODUCTION

According to the World Health Organization (WHO), the number of suicide deaths occurred worldwide reached an estimate of 804000 deaths in 2012 [1]. While according to Our World in Data, the number of people dying from suicide worldwide reaches a number close to 800000 deaths every year. The number of suicides is around twice as many as the number of homicides, leading to the fact that suicide is more frequent than homicide in most countries [2].

It is estimated that 3.8% of population worldwide or approximately 280 million people is diagnosed and affected by depression, while 75% of population in low to middle income countries receive little to no treatments for mental disorders such as depression [3].

Mental health affects every stage of human life, including psychological, emotional, and social states. Mental health state is a measure of how mentally healthy a human being. In the course of life, humans will eventually experience mental health issues, ranging from mild – such as simple panic attacks, to severe – such as suicidal behaviors. Mental illness sufferers are not only limited to adults, but it can also be experienced by children and adolescents. There are 4 types of common mental illness diagnosis – anxiety problems, Attention-Deficit/Hyperactivity Disorder (ADHD), behavioral problems, and depression. Among them, we tend to focus on depression, since it is the main cause of suicidal behavior.

Suicidal behavior is the actions that tend to exhibit suicidal thoughts or suicidal ideations. Suicidal behavior occurs due to depression, post-traumatic stress disorder, abuse, etc. Depression or major depressive disorders is a medical illness that alter one's feeling, thoughts, and actions negatively. It is known that people suffering from depression has problems in concentrating and performing activities.

With the growing numbers of suicide and depression, various methods of classifying suicidal ideation and depressive behavior have been developed. However, the implementation of those methods specifically in

individuals are rarely developed. Differentiating depression and suicidal ideation has been one of the crucial problems that the community face since both mostly have different treatment methods. By classifying and successfully differentiate suicidal ideation and depression from textual/written form, medical treatments can be applied more effectively.

Processing unlabeled data in Natural Language Processing (NLP) using unsupervised learning is considered more challenging than supervised learning. Logically, without labels in the data model, there will be no labels in the output data. This causes some performance measurements such as accuracy or F-Score difficult to implement. Unsupervised learning is rarely used and studied further in NLP due to some limitations in unlabeled learning models and the difficulty of labeled gold standard based evaluation that has been set.

However, unsupervised learning has advantages over supervised learning, where textual unlabeled data is available in very large quantities and can be run without the need for human supervision. Despite knowing the challenges that researchers will face in implementing unsupervised learning in NLP, this learning model has great potential in studying a large amount of available textual unlabeled data.

In this paper, we will be performing the following tasks:
1. Implement label correction through unsupervised learning.
2. Train an NLP model using Recurrent Neural Network based architecture.
3. Compare the performance of the proposed method with the baseline method.

## 2. LITERATURE REVIEW

Mental illness affects millions of people worldwide [4], making it a quite common phenomenon in society. Suicide is one of the highest causes of death that is difficult to predict and trying to minimize it is a challenge in itself [5]. In addition to suicide, depression is also a type of mental illness that is suffered by many, both by children, adolescents, and adults. In the current era of technological development, there are so many ways that can be done to minimize depression and suicide. Machine Learning (ML) is a method that is often talked about when it comes to the process of detecting mental illness.

In creating a predictive model, a publicly accessible dataset is needed. However, datasets on suicidal behaviour and depression are difficult to find and access due to concerns about the privacy and anonymity of participants who helped build the dataset [6]. Therefore, most publicly accessible datasets are the result of data retrieval via the internet, such as the results of previous studies that had taken data from health forums such as psychforums.com and depressionforums.org [7]. In addition, several studies use survey datasets which tend to be conducted by health institutions, universities or countries where the use of personal data is permitted and approved by participants who contribute to data collection [8], [9], [10].

There are many studies that have applied various ML algorithms to various sources with the aim of predicting depression and suicidal tendencies. Among them are using a combination of Machine Learning (ML) with Natural Language Processing (NLP) [11] to detect texts that indicates a tendency to depression and suicide in social media platforms [4], [12], [13], [14] and Online Support Forums, considering that the digital footprint left by users can help in diagnosing depression [15]. There are also several studies using data collected via brain signal called Electroencephalography (EEG) signals with the aim of measuring changes in brain activity associated with depression [16]. In addition, previous researchers have also tried to identify symptoms of suicide by means of Electronic Health Records (EHR) in children and adolescents who are hospitalized in mental hospitals [17], [18].

In conducting research, one of the other goals shared by several researchers who conducted similar studies is to find and prove the most suitable ML method that can detect early depressive symptoms [5], [11], [12], [15], [19] and has better accuracy than medical diagnoses [4], [13], with the aim of being able to predict the short and long term of people who may suffer from it [18]. In general, machine learning models designed to predict depression can only produce classification model or severity level detection [20]. However, based on depth literature, the possibility of ML method application in detecting real-time symptoms is quite high, one example is through a smart chatbot system [13].

It requires several procedures/stages to conduct research on the application of ML algorithms in detecting depression and suicide. First, the data used must be prepared before being processed. most researchers use Natural Language Processing (NLP) to retrieve the required features/parts, perform classification, and identification [11], [17], [15]. Other methods used to prepare datasets before processing are Mann – Whitney U Test (classification) [14], One-Hot encoding (extraction) [12], Principal Component Analysis (representation) [12], Kolmogorov–Smirnov (K–S) Test (test for normality) [14], SS3 Text-Classifier (classification) [21], and other options that may be used in certain scenarios. After the data is ready to be processed, various ML algorithms are implemented.

The most widely used ML model for detecting depression and suicidal behaviour is supervised learning [15], followed by unsupervised learning and combined learning [5]. Algorithms used in supervised

learning include Random Forest (RF), Naïve Bayes (NB), Decision Trees (DT), Least Squares Regression (LSR), Support Vector Machine (SVM), Minimum Description Length (MDL), Multi-Layer Perceptron (MLP), K-Nearest Neighbor (KNN), Linear Regression (LR) [5], [12], [22]. Meanwhile, the unsupervised learning algorithm used in a similar case is the Clustering Algorithm, Neural Network, and Self-Organizing Maps (SOM) [5]. The combination of the two methods is called combined learning, for example, cross-validation and data separation [5]. In addition to Machine Learning methods, many researchers also use Deep Learning methods, such as Long-Short-Term-Memory (LSTM), Bidirectional Long Short-Term Memory (BiLSTM) [20], [23] and Recurrent Neural Network (RNN) [4], [11], [12], [13], [15], [19].

The results obtained from various studies regarding the detection and act of suicide using the ML method are varied. A study at the end of various factors that can produce different results, such as the amount of data, test subjects, algorithms, media/hardware implementation, assistive devices, data cleanliness, data uniqueness, and so on. To get optimal results, researchers need to constantly try various methods and fix every failure. Regardless, the implementation of Machine Learning in predicting mental disorders, depression and suicide has enormous potential in helping the diagnosis process in real-time which is more accurate and faster than traditional methods.

In performing natural language processing (NLP) tasks, data in the form of texts needs to be converted into numerical data before being processed by computers. Recent studies introduced BERT, a new language representation model designed with the function of pre-training deep bidirectional representations from the form of unlabeled text [24].

## 3. DATASET

The dataset we use is the same dataset from the baseline of our study entitled “Deep Learning for Suicide and Depression Identification with Unsupervised Label Correction” [25]. The dataset consists of 1895 anonymous Reddit posts, where the original posts are used as input, while subreddits are used as the labels. Each scraped reddit post is labeled as 'suicidal' or 'depressed'. The anonymity of the data enables the availability of the dataset to the public.

## 4. RESEARCH METHOD

In preprocessing the dataset, we use the SDCNL method from the baseline of our study entitled “Deep Learning for Suicide and Depression Identification with Unsupervised Label Correction” [25], which starts with processing the dataset with word embedding models. The resulting embeddings are then processed by implementing an unsupervised dimensionality reduction to reduce the number of variables or features in a dataset. The output of the reducer is then clustered using a clustering-based algorithm to predict new labels unsupervised, which are then compared against the ground-truth labels by implementing a confidence-based threshold algorithm to correct the ground-truth labels. Hence, the results of the confidence-based threshold algorithm and the word embedding models are utilized to train the proposed Recurrent Neural Network (RNN) model through supervised learning.

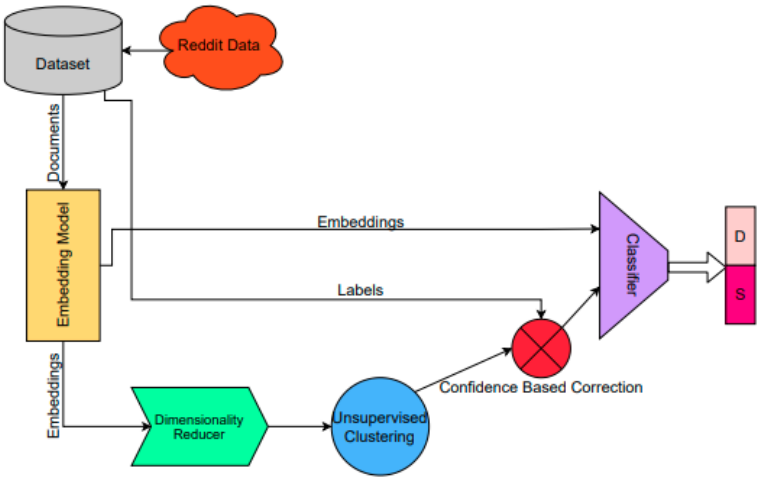


**Figure 1.** The schematics of SDCNL pipeline [25].

### 4.1. Word Embedding

Word embedding is the process of converting raw text data into numerical data that can be read and understood by computers. In this paper, we utilized Bidirectional Encoder Representations from Transformers

(BERT) [24] as the word embedding models, which are mostly known for outperforming similar architectures in various NLP tasks. The output generated after the implementation of BERT is a vector measuring dimensional vector in the form of multi-dimensional word-level and document-level embeddings.

### 4.2. Dimensionality Reduction

To compress a high-dimensional features resulted from word embedding process into a lower-dimensional form, we implement a dimensionality reduction algorithm. There are several commonly used algorithms for this task, namely Principal Component Analysis (PCA), Uniform Manifold Approximation and Projection (UMAP), and Deep Neural Autoencoders.

In this paper, we use UMAP to reduce the dimensionality of the embeddings created by the BERT model. UMAP works by generating high-dimensional data graph and converting it into a low-dimensional data graph while maintaining the similarity to the input [26].

### 4.3. Label Correction

After the reduction, we proceed to implement clustering algorithms in order to create two different clusters and assign their individual labels. We use the Gaussian Mixture Model (GMM), a parametric probability density function. It is utilized with the means to cluster a dataset based on probabilities.

The implementation of Gaussian Mixture produces two different labels: the original ground-truth labels and the new unsupervised clustering labels. The new labels will be utilized to correct the ground-truth labels or the original labels. When the produced label has a probability above the tuned threshold, the original label will be replaced by the new label produced by Gaussian Mixture. Otherwise, the original label will not be replaced. By implementing a threshold-based label correction, we avoid the elimination of noisy labels.

### 4.4. Baseline Classifier

The baselines used in this study consist of Convolutional Neural Network (CNN), Dense Neural Network (DNN), and Bidirectional LSTM (BiLSTM) based architectures, taken from previous study that proposed the SDCNL method. The CNN, DNN, and BiLSTM based architectures utilized BERT as the word embedding model, UMAP as the dimensionality reduction algorithm and GMM as the threshold label correction [25].

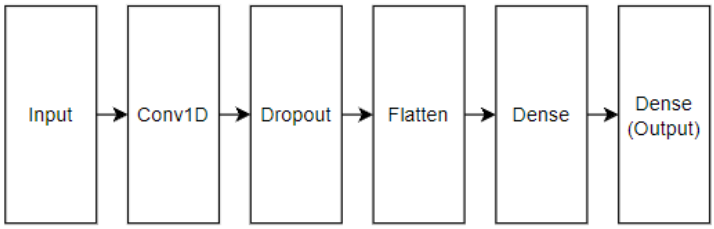


**Figure 2.** Illustration of the CNN model architecture [25].

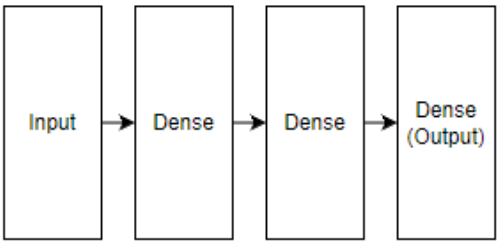


**Figure 3.** Illustration of the DNN model architecture [25].

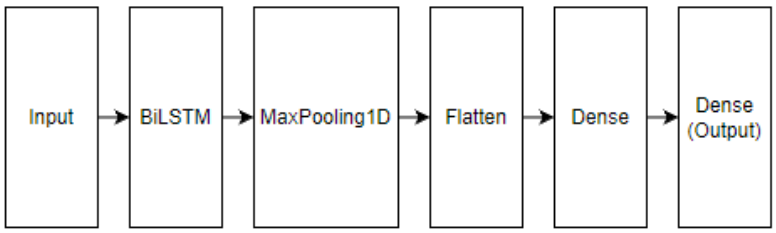


**Figure 4.** Illustration of the BiLSTM model architecture [25].

### 4.5. Proposed Classifier

The model in our study is constructed using Recurrent Neural Network (RNN). Our RNN model implemented a 5-unit Gated Recurrent Unit (GRU) layer as the first layer, utilizing rectified linear unit (ReLU) as the activation function, and he uniform as the kernel initializer. The first layer is then directly flattened, followed by a 64-unit dense layer, utilizing rectified linear unit (ReLU) as the activation function, and he uniform as the kernel initializer. Last, a sigmoid function was applied to calculate the final output.

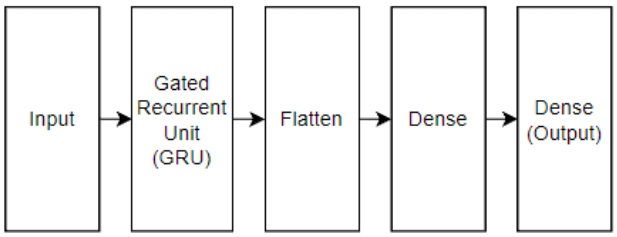


**Figure 5.** Illustration of the proposed RNN model architecture.

## 5. RESULTS AND ANALYSIS

Our proposed model was trained and tested using anonymous reddit posts which predicted the label of the post. The dataset was preprocessed without cleaning out noise in the data. The dataset is first converted using BERT, which then outputs 768 columns of dimensional vector of embeddings. The embeddings are then used to correct ground-truth labels, train and test the model.

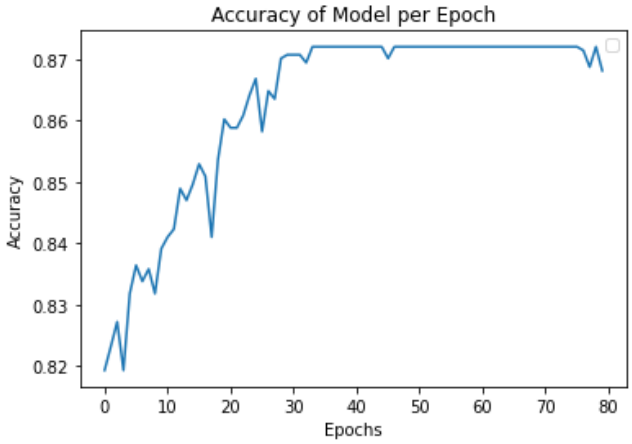


**Figure 6.** Illustration of the model's accuracy at every epoch.

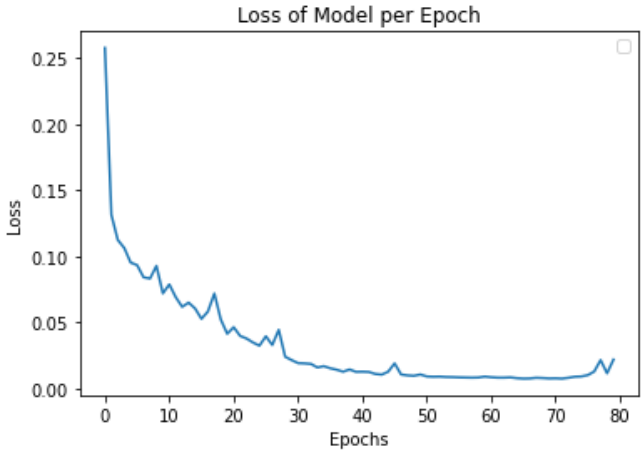


**Figure 7.** Illustration of the model's loss at every epoch.

The training was done with the total of 80 epochs, while the number of batch size being 32. The accuracy of the model increased to its peak within 35 epochs, while the loss of the model decreased drastically only within a few epochs, with the lowest loss being around 60 to 70 epochs.

**Table 1.** Comparison between the proposed model and the baseline models, with the best score in each metrics bolded

| Metrics (%) | RNN | CNN | DNN | BiLSTM |
|---|---|---|---|---|
| Accuracy | **86.81** | 84.59 | 83.74 | 84.16 |
| Precision | **97.13** | 95.38 | 95.51 | 93.10 |
| Recall | **94.69** | 86.05 | 85.08 | 87.09 |
| F1 Score | **95.90** | 90.45 | 89.99 | 89.99 |
| AUC | **92.84** | 82.91 | 81.97 | 85.08 |

From the table, our proposed model performed better than all three baseline architectures. The proposed model resulted in a higher accuracy (86.81%), higher precision (97.13%), higher recall score (94.69%), higher F1 score (95.90%), and a higher Area Under the Curve (AUC) score (92.84%). It is likely that the proposed model performed better due to the nature of RNN being better suited in analyzing sequential data such as texts and performing sentiment analysis.

The result of this study is aimed to provide wider options for implementing machine learning as a diagnostic and supplementary tool in identifying suicidal ideations and depression through text analysis. Based on the result, artificial intelligence (AI) and machine learning alone cannot provide the utmost proper screening. Therefore, further research of the topic needs proper clinical support and awareness of the ethical concerns regarding suicidal ideations.

## 6. CONCLUSION

This study constructed a RNN based architecture that predicted suicidal ideations and depression in anonymous reddit posts, with the implementation of unsupervised clustering and threshold-based label correction. This study showed that the proposed RNN architecture resulted in higher evaluation score compared to all three baseline architectures (CNN, DNN, and BiLSTM) in all five metrics (accuracy, precision, recall, f1, and AUC score). This study also concluded that the use of RNN based architecture fits tasks that perform sentiment analysis and sequential data analysis such as texts.

## REFERENCES

[1] World health Organization, "Preventing suicide - A global imperative," 2014. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/131056/9789241564779_eng.pdf.

[2] H. Ritchie, M. Roser and E. Ortiz-Ospina, "Suicide," Our World in Data, 2015. [Online]. Available: https://ourworldindata.org/suicide.

[3] World Health Organization, "Depression," World Health Organization, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression.

[4] A. Amanat, M. Rizwan, R. A. Javed, M. Abdelhaq, R. Alsaqour, S. Pandya and M. Uddin, "Deep Learning for Depression Detection from Textual Data," Electronics, 2022.

[5] A. Fatima, Y. Li, T. T. Hills and M. Stella, "DASentimental: Detecting Depression, Anxiety, and Stress in Texts via Emotional Recall, Cognitive Networks, and Machine Learning," Big Data and Cognitive Computing.

[6] R. Sawhney, P. Manchanda, P. Mathur, R. Shah and R. Singh, "Exploring and Learning Suicidal Ideation Connotations on Social Media with Deep Learning," Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, p. 167–175, 2018.

[7] R. Németh, D. Sik and F. Máté, "Machine Learning of Concepts Hard Even for Humans: The Case of Online Depression Forums," vol. 19, 2020.

[8] S. Ryu, H. Lee, D.-K. Lee, S.-W. Kim and C.-E. Kim, "Detection of Suicide Attempters among Suicide Ideators Using Machine Learning," 2019.

[9] D. Shin, W. I. Cho, C. H. K. Park, S. J. Rhee, M. J. Kim, H. Lee, N. S. Kim and Y. M. Ahn, "Detection of Minor and Major Depression through Voice as a Biomarker Using Machine Learning," Journal of Clinical Medicine, vol. 10, 2021.

[10] J.-C. Weng, T.-Y. Lin, Y.-H. Tsai, M. T. Cheok, Y.-P. E. Chang and V. C.-H. Chen, "An Autoencoder and Machine Learning Model to Predict Suicidal Ideation with Brain Structural Imaging," Journal of Clinical Medicine, vol. 9, 2020.

[11] C. Su, R. Aseltine, R. Doshi, K. Chen, S. C. Rogers and F. Wang, "Machine Learning for Suicide Risk Prediction in Children and Adolescents with Electronic Health Records," Translational Psychiatry, 2020.

[12] I. A. N. Arachchige, P. Sandanapitchai and R. Weerasinghe, "Investigating Machine Learning & Natural Language Processing Techniques Applied for Predicting Depression Disorder from Online Support Forums: A Systematic Literature Review," Information, 2021.

[13] M. J. Havigerova, J. Haviger, P. Hoffmannova and D. Kucera, "Text-Based Detection of the Risk of Depression," Front. Psychol., 2019.

[14] M. Z. Uddin, K. K. Dysthe, A. Folstad and P. B. Brandtzaeg, "Deep Learning for Prediction of Depressive Symptoms in a Large Textual Dataset," Neural Computing and Applications.

[15] R. A. Bernert, A. M. Hilberg, R. Melia, J. P. Kim, N. H. Shah and F. Abnousi, "Artificial Intelligence and Suicide Prevention: A Systematic Review of Machine Learning Investigations," International Journal of Environmental Research and Public Health, 2020.

[16] M. Cukic, D. Pokrajac, M. Stokic, s. Simic, V. Radivojevic and M. Ljubisavljevic, "EEG machine learning with Higuchi fractal dimension and Sample Entropy as features for successful detection of depression," 2018.
[17] N. J. Carson, B. Mullin, M. J. Sanchez, F. Lu, K. Yang, M. Menezes and B. L. Cook, "Identification of Suicidal Behavior Among Psychiatrically Hospitalized Adolescents Using Natural Language Processing and Machine Learning of Electronic Health Records," Plos One, 2019.
[18] P. Jain, K. R. Srinivas and A. Vichare, "Depression and Suicide Analysis Using Machine Learning and NLP," Journal of Physics: Conference Series, 2022.
[19] S. Chakraborty, H. F. Mahdi, H. A. M. Al-Abyadh, K. Pant, A. Sharma and F. Ahmadi, "Large-Scale Textual Datasets and Deep Learning for the Prediction of Depressed Symptoms," Computational Intelligence and Neuroscience, vol. 2022, 2022.
[20] H. Dinkel, M. Wu and K. Yu, "Text-based depression detection on sparse data," arXiv, 2020.
[21] S. G. Burdisso, M. Errecalde and M. Montes-y-Gómez, "A Text Classification Framework for Simple and Effective Early Depression," 2019.
[22] F. Cacheda, D. Fernández, F. J. Novoa and V. Carneiro, "Analysis and Experiments on Early Detection of Depression," 2018.
[23] V. K. Gunjan, Y. Vijayalata, S. Valli, S. Kumar, M. O. Mohamed and V. Saravanan, "Machine Learning and Cloud-Based Knowledge Graphs to Recognize Suicidal Mental Tendencies," 2022.
[24] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2019.
[25] A. Haque, V. Reddi, and T. Giallanza, “Deep Learning for Suicide and Depression Identification with Unsupervised Label Correction,” 2021.
[26] L. McInnes, J. Healy, N. Saul, and L. Großberger, “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction,” 2018.

## BIBLIOGRAPHY OF AUTHORS

Rhenaldy is currently a university student majoring in Computer Science at Bina Nusantara University, Jakarta.

Ladysa Stella Karenza is currently a university student majoring in Computer Science at Bina Nusantara University, Jakarta.

Hidayaturrahman is faculty member of Bina Nusantara University. He got his master's degree from Bandung Institute of Technology in 2018. His research fields are Computer Vision, Natural Language Processing, and Machine Learning.

Muhamad Keenan Ario, finished his master education at Bina Nusantara University and graduated in 2022. He currently serves as a lecturer at the Bina Nusantara University, Jakarta.