Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 446−452
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1857

Depression Detection on Social Media Twitter Using Hierarchical Attention Network Method

Raihan Nugraha Setiawan*, Warih Maharani

School of Computing, Informatics, Telkom University, Bandung, Indonesia
Jl. Telekomunikasi No. 1, Terusan Buahbatu - Bojongsoang, Sukapura, Kec. Dayeuhkolot, Kabupaten Bandung, Jawa Barat, Indonesia
Email: [1]rnsetiawan@student.telkomuniversity.ac.id, [2]wmaharani@telkomuniversity.ac.id
Email Penulis Korespondensi: raihannsetiawan@gmail.com
Submitted: **99/99/999**; Accepted: **99/99/999**; Published: **99/99/999**

Abstrak–Penyakit mental, termasuk depresi, bukanlah penyakit ringan yang hanya dialami oleh beberapa orang yang lemah secara mental. Teknologi berkembang begitu pesat khususnya teknologi komunikasi melalui media sosial. Media sosial yang populer saat ini yaitu Twitter. Melalui tweet, penggunanya dapat meluapkan segala perasaan yang sedang dialami dengan mudah sehingga memungkinkan kita menemukan informasi mengenai perasaan emosi hingga tingkat depresi pengguna. Analisis otomatis media sosial dapat memberikan solusi untuk deteksi dini. Dengan menggunakan data dari media sosial Twitter, penelitian ini memiliki tujuan untuk memprediksi tanda-tanda awal depresi. Metode yang diterapkan pada penelitian ini yaitu klasifikasi dengan analisis sentimen media sosial menggunakan Hierarchical Attention Network. Klasifikasi dengan metode Hierarchical Attention Network dipilih karena metode tersebut menunjukkan hasil yang sangat baik untuk melakukan klasifikasi pada teks pada penelitian sebelumnya. Model klasifikasi dalam penelitian ini memperoleh akurasi terbaik sebesar 74%, dilakukan dengan menerapkan metode Hierarchical Attention Network.

Kata Kunci: Penyakit Mental; Depresi; Twitter; Analisis Sentimen; Hierarchical Attention Network

Abstract–Mental illness, including depression, is not a mild condition that only some mentally weak people experience. Technology is developing so rapidly, especially communication technology through social media. Twitter is a very popular social media today. Users can easily quickly and simply communicate all the feelings they are experiencing through tweets, which allows us to find information about emotional feelings to the level of user depression. Auto-mated analysis of social media has the potential to provide a method for early detection. This study aims to predict early signs of depression using data from social media Twitter. The method used in this research is classification by analyzing social media sentiment using the Hierarchical Attention Network. Classification using the Hierarchical Attention Network method was chosen because the method showed outstanding results for classifying texts in previous studies. The classification model in this study that represents the best accuracy, 74%, was performed by applying the Hierarchical Attention Network.

Keywords: Mental Illness; Depression; Twitter; Sentiment Analysis; Hierarchical Attention Network

1. INTRODUCTION

Mental illness, including depression, is not a mild condition that only some mentally weak people experience. According to the World Health Organization (WHO), the number of people suffering from this type of mental illness exceeds 264 million [1]. Depression can affect anyone, regardless of gender, age, race, or social status. Some people with mental disorders are reluctant to admit their condition, while others are sick and don't realize they need help [2]. In addition, sufferers consciously deny that they are sick and need help because they are ashamed or afraid of being ostracized by those around them. This condition is exacerbated by the lack of social support, because people with mental disorders are often isolated in many communities. As a result, it is unknowable and will only be detected when it is too late. Currently, technology is developing so rapidly, especially communication technology through social media. With all the features of social media, it makes it easier for users to carry out all their activities.

Twitter is a platform that enables people from all backgrounds to express their opinions about life, business, products, brands or services today [3], since it can be used to spread information in real-time quickly [4]. Twitter is used by various groups of users, ranging from celebrities, political figures, actors, business people, and various leaders, to convey facts or opinions [5]. According to the Kamus Besar Bahasa Indonesia (KBBI), facts are things that resemble reality [6], while opinions are expressions that describe the thoughts or emotions of the author [7]. Through tweets, Twitter users can express all feelings ranging from sadness, anger, confusion, and experiences experienced with ease. This phenomenon provides an opportunity for psychologists to obtain additional data through social media Twitter [2]. Social media automated analysis has the potential to provide a method for the early detection [8].

According to a study by Semiocast, a social media research institute based in the city of Paris, France, Indonesia has the fifth-highest number of Twitter account users in the world. It is in the third-highest position in the country with the most active tweets per day [4]. When examined further, the emotions expressed through these tweets can be associated with mental illness, especially depression. Depression is a person's negative emotional state and has gone through quite a long time [9].

Several studies on this have been carried out, using social media as a source to identify depressive disorders through sentiment and emotion analysis approaches and techniques [10]. For example, research by Ivan Sekulic and Michael Strube in 2019 used Logistic Regression and Linear SVM algorithm to assess mental health on social

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 446−452
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1857

media. The Reddit data used in this study included thousands of users who had been diagnosed as having about one or more mental disorders. The study tested several methods such as Logistic Regression, Linear SVM, Supervised FastText, and HAN. As a result, the HAN method proved to be better in predicting mental health than other methods [11].

According to a study by Hasan et al [12] using the SVM and Naïve Bayes methods. There is a weakness of these methods; namely, the linear classification does not share parameters between features and classes. This may limit their generalization when the output is large. One solution to this problem is to factorize linear classification into multilayer neural networks [13]. The research that has been done shows that from the user's posts, the category of mental health can be known. This has shown that media can be a useful resource for mental health professionals preparing for the possibility of mental diseases such as depression.

In a study conducted by Zichao Yang et al [14] under the title Hierarchical Attention Networks for Document Classification. In this study, the HAN method was tested and compared with several other methods. The other classification method consists of Logistic Regression, SVM, LSTM, CNN, Conv-GRNN, and LSTM-GRNN. This study uses data from other studies, such as Yelp reviews, IMDB reviews, Yahoo answers, and Amazon reviews. In this study, the resulting HAN classification model can significantly outperform CNN up to 7.3%, 8.8%, 8.5%, and 10.2%. The research concludes that the Hierarchical Attention Network is a model that provides the best performance from all datasets.

According to a study by Hayatin N [2] with the title Implementation of Nave Bayes Multinomial for Data Classification of Tweets Containing the Term of Depression. The dataset used in this study was taken from the Sentiment140 dataset from Kaggle, which contains 1,600,000 tweets extracted using the Twitter API. From the experimental results, the accuracy value is 70%, the precision and recall values are 72% and 65%, and the f-measure value is 68%. From these tests, it can be concluded that the algorithm used produces poor accuracy.

By referring to the problems above and a number of relevant studies, researchers were motivated to conduct a similar study by detecting depression in Twitter users using a deep learning algorithm with the Hierarchical Attention Network method. This method is one of the multilayer neural network classification methods since it is aimed to capture two main insights about document structure. HAN builds a sentence representation first and then combines it into a document representation [14]. It was proven in research by Ivan Sekulic et al [11], that using the HAN method produces fairly excellent results for classifying text compared to several other methods such as Logistic Regression, Linear SVM, and Supervised FastText.

2. RESEARCH METHODOLOGY

2.1. Research Flow

In order to develop a classification system to identify depression, this research is divided into several steps. There are several steps involved in creating this classification system, including collecting the dataset, preprocessing, splitting the data, data training, data testing, modeling, and evaluating the classification model. Figure 1 shows a description of the system to be built.

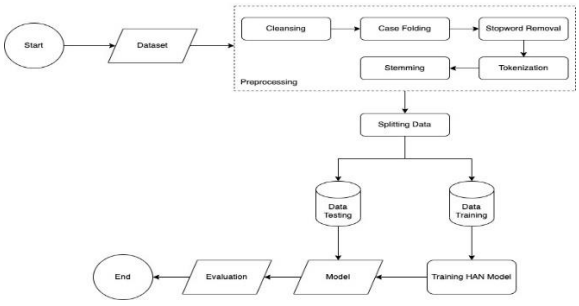

Figure 1. Flowchart of the system to be built

The dataset has an imbalanced class distribution where the positive label has more tweets than the negative label as shown in Figure 2.

Figure 2. Distribution of tweets based on labels in the dataset

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 446−452
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1857

2.1.1. Dataset

The study collects tweets from Twitter users who responded to a 42-question depression test based on the Depression Anxiety Stress Scale (DASS-42). DASS-42 is a measuring instrument used in this study to measure the severity of disorders such as depression, anxiety, and stress. DASS-42 questionnaire developed by Lovibond, S.H, and Lovibond, P.F has been tested for reliability validity and has been declared valid and reliable [15].

For each user, we have gathered data on their tweets, mentions, and replies using the Twitter API. The data have been categorized into two different labels, a positive label implies that the user's tweet has the potential for depression, while the negative label is the opposite. A negative label consists of 1856 tweets and a positive label consists of 2013 tweets.

2.1.2. Preprocessing

Preprocessing is a crucial and initial step in the sentiment analysis and depression detection processes. It changes raw data into a format that can be analyzed. The core of this procedure is the cleaning and transformation of the required data. Preprocessing methods used in this study include stemming, tokenization, stopword removal, case folding, and cleansing. Cleansing is a stage in preprocessing to clean existing data by removing symbols, numbers, punctuation marks, redundant spaces, and characters that are not in the alphabet. Case folding is the process of uniforming all existing letters into lower case letters. Stopword removal is the process of removing words that are unnecessary, meaningless, and less influential for the upcoming process. Tokenization is the process of splitting a sentence into a series of words. Stemming is the process of finding the basic word of a word. This process removes affixes so that the word becomes a basic word.

Table 1 demonstrates an example of the preprocessing process, starting with the data input and finale with data that have built high quality.

Table 1. Preprocessing Process

Process Name	Input	Output
Cleansing	Isu mau offline nih, enakan online bisa sambil kerja HHHHH!!	Isu mau offline nih enakan online bisa sambil kerja HHHHH
Case Folding	Isu mau offline nih enakan online bisa sambil kerja HHHHH	isu mau offline nih enakan online bisa sambil kerja hhhhh
Stopword Removal	isu mau offline nih enakan online bisa sambil kerja hhhhh	isu mau offline enakan online bisa sambil kerja
Tokenization	isu mau offline enakan online bisa sambil kerja	“isu”, “mau”, “offline”, “enakan”, “online”, “bisa”, “sambil”, “kerja”
Stemming	isu mau offline enakan online bisa sambil kerja	isu mau offline enak online bisa sambil kerja

2.1.3. Hierarchical Attention Network

The method used in this research is the Hierarchical Attention Network method. The sentence encoder, sentence-level attention layer, word sequence encoder, and word-level attention layer are some of the components of HAN method [14]. Figure 3 shows the overall architecture of the Hierarchical Attention Network (HAN).

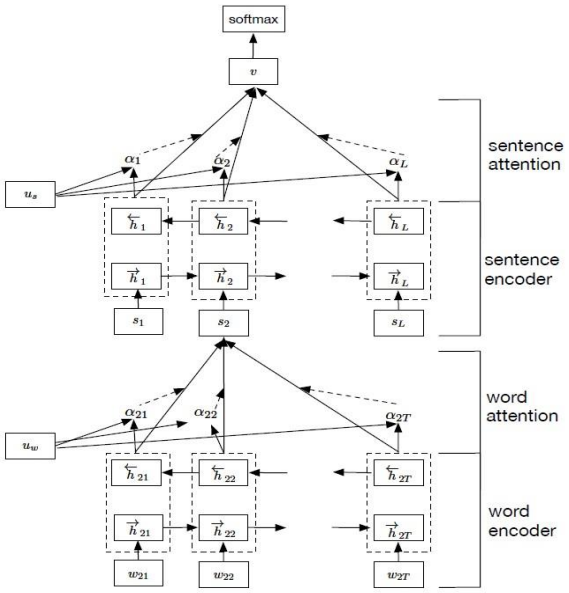

Figure 3. Hierarchical Attention Networks

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 446−452
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1857

It uses GRU-based sequence encoders on the sentence and document level, producing a document representation in the process. A representation of a given sentence is created by the word sequence encoder and delivered to the sentence sequence encoder, which, given a list of encoded sentences, produces a document representation. Both word sequence encoders and sentence sequence encoders use attention mechanisms to enhance the representation of the input sequence [11].

1. Word Encoder

A two-way GRU is used to annotate words from both directions and combine contextual information.

$$x_{it} = W_e w_{it} t \in [1, T], \quad (1)$$

$$\underset{h}{\rightarrow}_{it} = \underset{GRU}{\longrightarrow}(x_{it}), t \in [1, T], \quad (2)$$

$$\underset{h}{\leftarrow}_{it} = \underset{GRU}{\longrightarrow}(x_{it}), t \in [T, 1]. \quad (3)$$

x_{it} is a vector of words corresponding to w_{it}.

2. Word Attention

A mechanism is needed to extract important or informative words and combine the representations of these words to form vectors.

$$u_{it} = \tanh(W_w h_{it} + b_w \quad (4)$$

$$\alpha it = \frac{\exp(u_{it} u_w)}{\sum_t \exp(u_{it} u_w)} \quad (5)$$

$$s_i = \sum_t a_{it} h_{it} \quad (6)$$

3. Sentence Encoder

The Sentence Encoder also uses a two-way GRU to encode sentences like the Word Encoder.

$$\underset{h}{\rightarrow}_{i} = \underset{GRU}{\longrightarrow}(s_i), i \in [1, L], \quad (7)$$

$$\underset{h}{\leftarrow}_{i} = \underset{GRU}{\longleftarrow}(s_i), i \in [1, L]. \quad (8)$$

4. Sentence Attention

The attention mechanism at the sentence level measures how important the sentence is.

$$u_i = \tanh(W_s h_i + b_s \quad (9)$$

$$\alpha_{it} = \frac{\exp(u_i u_s)}{\sum_i \exp(u_i u_s)} \quad (10)$$

$$v_i = \sum_i \alpha_i h_i \quad (11)$$

In this study, we classify a user as a document, providing the HAN to be simply customized. Just as a document is a sequence of sentences, we consider modeling a Twitter user as a sequence of postings. This study also classifies tweets as sentences because they both consist of a sequence of tokens. This interpretation makes it suitable for this study to successfully apply the HAN, which was quite successful in classifying documents, to Twitter users.

2.1.4. Evaluation

Evaluation is conducted to determine the performance quality of the developed classification model. In this paper, the accuracy of the model has been found out by the confusion matrix. There are four values used to represent the results of the classification process such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Based on these values, the accuracy, precision, recall, and f1-score values can be obtained by the formula given below:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (12)$$

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 446−452
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1857

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$f1 = \frac{2 * (Recall * Precision)}{Recall + Precision} \quad (15)$$

3. RESULTS AND DISCUSSION

The testing in these steps are the test tuning of the parameters used in the process of calculating the Hierarchical Attention Network (HAN) method. The tuning parameters attempt to obtain the most optimal parameters that will provide the best accuracy of each kernels, so the models will be good to use on classifying the depression. These parameters include partitioning training and test data, changing the number of epochs, changing the batch size value, and eliminating some components in preprocessing. To determine the impact of splitting the test data and training data, the first test was carried out. Then the second test was carried out to determine the impact of batch size. And the last is a test to determine the effect of doing the preprocessing process without stemming, without stopword removal, without both, and with complete preprocessing. The data used is from Twitter users who have filled out the questionnaire. There are about 3800 data which are divided into two categories or classes, including positive and negative. A positive label indicates that the user's tweet may be depressed. Meanwhile, a negative label indicates the opposite. 1856 tweets are included in a negative label, while 2013 tweets are included in a positive label.

3.1 Result and discussion of the effect of data partitioning

Data partition testing is done by changing the size of the testing data and training data, Performance data partition is calculated by using formulas, we calculated the precision, recall, F1 scores, and accuracy using the formulas as shown in Table 2.

Table 2. Test Results of Data Train and Data Test Partitioning

Test No.	Data Train %	Data Test %	Precision %	Recall %	F1-Score %	Accuracy %
1	60	40	71.52	79.96	75.47	70.98
2	70	30	69.10	74.54	75.93	69.37
3	80	20	73.59	78.79	76.09	73.92
4	90	10	74.15	79.56	79.15	**75.44**

If visualized through a graph, the results of the data partition test can be seen in the following figure:

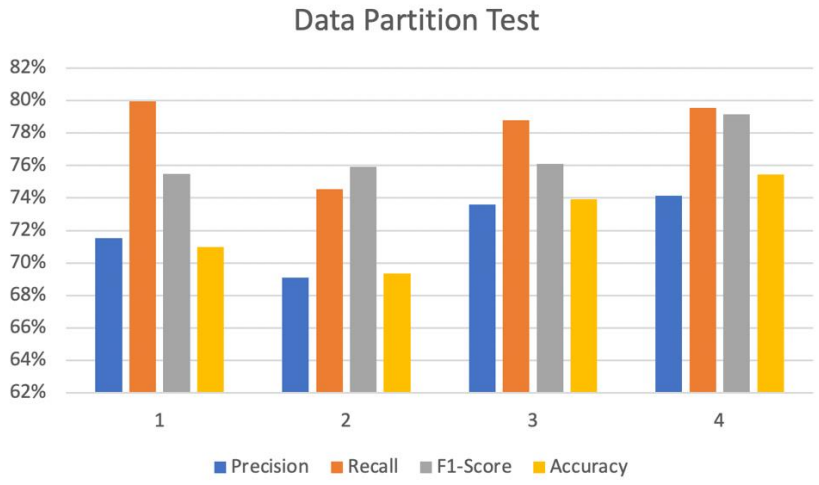

Figure 4. Data Partition Test Graph

In Figure 4, it can be seen that the fourth test is the test with the highest value for each parameter, especially on accuracy with a value of 75.44%. The test will be used as a dataset in the next test.

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 446−452
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1857

3.2 Result and discussion of the effect of batch size and epoch

Batch size and epoch testing are done by changing the batch size and epoch values according to the provisions in Table 3. This test will use a data partition with the best result from the previous test, which is 90:10.

Table 3. Test Results of Batch Size and Epoch

Test No	Batch Size	Epoch	Accuracy (%)	Loss
1	8	10	70.99	0.76
2	16	20	69.87	0.69
3	32	20	**74.82**	0.81
3	32	30	71.78	0.72
4	64	40	73.87	0.74
5	128	50	69.56	1.85

Table 3 shows the result of testing batch size and epoch. The third test is the highest accuracy, with an accuracy value of 74,82%. However, the accuracy obtained in each test did not show a significant difference except for the training time.

3.3 Result and discussion of the effect of preprocessing

In this scenario, testing is carried out by applying different preprocessing techniques. At this stage, four tests were carried out by applying various combinations of preprocessing techniques, namely without stemming, without stopword removal, without stemming and stopword removal, and using full preprocessing. The results of the accuracy of this scenario can be seen in Table 4.

Table 4 Effect of preprocessing

Preprocessing	Accuracy
No Stemming	74.13%
No Stopword Removal	69.94%
No Stemming and Stopword Removal	68.37%
Full Preprocessing	72.3%

Based on the results of the tests that have been carried out, it can be concluded that tests involving complete preprocessing techniques and with stopwords but without stemming have a fairly good effect on improving the performance of the classification system when compared to the results of tests that do not involve both stemming and stopword removal processes which only achieve an accuracy of 68.37 percent.

4. CONCLUSION

In conclusion, this paper developed and analyzed the performance of Hierarchical Attention Networks to identify depressed and non-depressed participants from their tweets, which were acquired from those who filled out the questionnaire. Based on the results of various test scenarios that have been carried out, it can be concluded that the best system performance is produced when using a combination of preprocessing without stemming. The more training data used, the higher the accuracy value obtained according to the data partition test. In this case, the best data partition is with 90% training data and 10% test data. Batch size value speeds up the data training process. In addition, a high batch size will require a more considerable epoch value to get the maximum accuracy value. However, it requires more advanced computer performance. In this case, the best batch size and epoch values are 32 and 20. By changing the data partition, preprocessing, batch size, and epoch, the highest accuracy results obtained in this test is 74.13%. Based on the results of the research that has been done. Suggestions that can be applied for further research are the need to increase the number of datasets labeled by experts. Use a high-performance computer so that the training process is more optimal and gets better results. Additionally, implement feature extraction to produce more information in order to minimize misclassification.

REFERENCES

[1] P. M. Depression, “Depression 13,” *Depression*, no. September, pp. 21–24, 2021, [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/depression.

[2] N. Hayatin, “Implementasi Multinomial Naïve Bayes Untuk Klasifikasi Data Tweets Mengandung Term Depresi,” pp. 344–349, 2020, [Online]. Available: http://research-report.umm.ac.id/index.php/sentra/article/download/3921/3903.

[3] P. Arora and P. Arora, “Mining Twitter Data for Depression Detection,” *2019 Int. Conf. Signal Process. Commun. ICSC 2019*, pp. 186–189, 2019, doi: 10.1109/ICSC45622.2019.8938353.

[4] P. Antinasari, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1718–1724, 2017, [Online]. Available: http://j-ptiik.ub.ac.id.

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 446−452
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1857

[5] L. Mandloi and R. Patel, "Twitter sentiment analysis using machine learning methods," *2020 Int. Conf. Emerg. Technol. INCET 2020*, pp. 1–5, 2020, doi: 10.1109/INCET49848.2020.9154183.
[6] "Kamus Besar Bahasa Indonesia" Badan Pengembangan dan Pembinaan Bahasa, "fak.ta ⇢," [Online]. Available: https://kbbi.kemdikbud.go.id/entri/fakta.
[7] "Kamus Besar Bahasa Indonesia" Badan Pengembangan dan Pembinaan Bahasa, "opi.ni ⇢," [Online]. Available: https://kbbi.kemdikbud.go.id/entri/opini.
[8] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: an integrative review," *Curr. Opin. Behav. Sci.*, vol. 18, pp. 43–49, 2017, doi: 10.1016/j.cobeha.2017.07.005.
[9] A. Chanaa and N. eddine El Faddouli, "E-learning Text Sentiment Classification Using Hierarchical Attention Network (HAN)," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 13, pp. 157–167, 2021, doi: 10.3991/ijet.v16i13.22579.
[10] F. T. Giuntini, M. T. Cazzolato, M. de J. D. dos Reis, A. T. Campbell, A. J. M. Traina, and J. Ueyama, "A review on recognizing depression in social networks: challenges and opportunities," *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 11, pp. 4713–4729, 2020, doi: 10.1007/s12652-020-01726-4.
[11] I. Sekulic and M. Strube, "Adapting Deep Learning Methods for Mental Health Prediction on Social Media," pp. 322–327, 2019, doi: 10.18653/v1/d19-5542.
[12] A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," *Int. Conf. Inf. Commun. Technol. Converg. ICT Converg. Technol. Lead. Fourth Ind. Revolution, ICTC 2017*, vol. 2017-Decem, pp. 138–140, 2017, doi: 10.1109/ICTC.2017.8190959.
[13] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *15th Conf. Eur. Chapter Assoc. Comput. Linguist. EACL 2017 - Proc. Conf.*, vol. 2, pp. 427–431, 2017, doi: 10.18653/v1/e17-2068.
[14] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical Attention Networks for Document Classification," 2016, [Online]. Available: http://arxiv.org/abs/1606.02393.
[15] W. Noviani, "Hubungan Tingkat Stres Dengan Efikasi Diri Pada Pasien TB Paru di Wilayah Kerja Puskesmas Patrang Kabupaten Jember," *Fak. Keperawatan, Univ. Jember*, p. 9, 2018.