Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 379−385
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2157

Predicting Depressive Disorder Based on DASS-42 on Twitter Using XLNet's Pretrained Model Classification Text

Intan Ramadhani, Warih Maharani*

School of Computing, Telkom University. Bandung, Indonesia
Email: [1]tanermdhni@student.telkomuniversity.ac.id, [2*]wmaharani@telkomuniversity.ac.id
Submitted: **20/08/2022**; Accepted: **29/08/2022**; Published: **30/08/2022**

Abstract−Twitter is a free social media site that is not only a place to share posts and multimedia content but also offers its users to express their feelings, emotions, and sentiments about an issue. So with this, it is often found that Twitter users make posts that show how the user's behavior includes mental problems experienced users such as symptoms of depression, anxiety, and stress disorders. Only about half of depression cases can be detected by doctors or other experts, this is because until now, the diagnosis of depression starts from reports of patients, family, or close friends of patients, or also starts from the results of certain tests such as questionnaires. So this research builds a model to predict depression by building a model that predicts whether someone is depressed through tweets on Twitter using the XLNet pre-trained text classification model. Testing is done by removing stemming from the preprocessing stage. Testing is also done by adding hyperparameters for fine-tuning the XLNet model. Testing is also carried out using a dataset that filters out foreign words where foreign data is translated into Indonesian. The data stored is data that uses words based on the KBBI dictionary. Based on the results of model testing that has been carried out using confusion matrix, the model can predict tweets that indicate depression and get an accuracy value of 78.57%.

Keywords: Tweet; Depression; DASS-42; Predict; XLNet

1. INTRODUCTION

During this period of rapid information development, many platforms or containers have emerged that support human interaction, social media itself is a media platform that aims to facilitate humans in activities and collaboration [1]. Twitter is a free social media site that allows registered users to communicate with others in real-time using 140-character statements. Applications like Twitter are not only a place to share writings and multimedia content but also offer users to express their feelings, emotions, and sentiments about an issue. So with this, it is often found that Twitter users make posts that show how the user's behavior includes mental problems experienced users such as symptoms of depression, anxiety, and stress disorders.

Depression is a mental disorder with conceptual and theoretical similarities and sometimes overlaps, especially in the younger generation [2]. Only about half of depression cases can be detected by doctors or other experts, this is because until now, the diagnosis of depression starts from reports of patients, family, or close friends of patients, or also starts from the results of certain tests such as questionnaires. Therefore, further research is needed to determine the level of people's depressive disorders. The Beck Anxiety Inventory (BAI) and Beck Depression Inventory (BDI) have developed methods to assess levels of anxiety and depression, but they do not address the component of stress as a bodily reaction. To address this, Lovibond PF and Lovibond SH developed the Depression, Anxiety, and Stress scale (DASS) to define, understand and measure the magnitude of the three negative emotional states. The original DASS is a 42-item questionnaire in English that contains 14 questions each to assess depression, anxiety, and stress [3].

Several studies have been conducted to predict depression using various algorithms. One of them is the research conducted by Rodrigo Martınez Castano and his team using the BERT algorithm which is one of the representations of Unsupervised Learning Text to predict depression from Reddit social media. The research shows that the content created by Reddit users can predict the user's depression symptoms. In the classification process, it has applied the performance of the BERT algorithm (XLM-RoBERTa-base language model). Because the research uses the BERT algorithm which creates a separate model to predict the results of each question, there are weaknesses in the research where the BERT algorithm cannot get context from both sides so the prediction results are not as expected [4].

Unsupervised Learning Text representation besides the BERT algorithm is the XLNet algorithm. With the ability to model bidirectional context, BERT achieves better performance than pretrain approaches based on autoregressive language modeling. The BERT algorithm works by corrupting the input using artificial symbols and ignoring the dependency between the position of artificial symbols and the pretrain-fine tune difference [5]. XLNet, on the other hand, is a general autoregressive pretrain method that enables bidirectional context learning by maximizing the expected probability of all permutations of a factorization sequence. XLNet integrates ideas from Transformer-XL, a sophisticated autoregressive model, into the pretrain thus overcoming the limitations of BERT thanks to its autoregressive[6].

Hans Christian et al, used three classifications for personality prediction from various social media data sources, namely BERT, RoBERTa, and XLNet language models to predict personality based on text. During the steps of work, data collection from social media then data preprocessing is carried out after that model development and model evaluation. From this research, the best accuracy value of BERT is 83%, RoBERTa is 81%, and XLNet is 81% [7].

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 379−385
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2157

Based on the explanation above, where Twitter social media becomes a place to vent people's heart and the lack of knowledge and assistance to find out people's mental disorders/depression, it is necessary to conduct research that aims to determine the level of people's mental disorders/depression from seeing people's outpouring on Twitter social media. This research aims to predict depressive disorders based on DASS-42 on Twitter using the XLNet pre-trained model text classification method. By using the tweet data that has been collected, it can be seen that the model works well by looking at the accuracy. This research only focused on tweets from Twitter users who had filled out an Indonesian-language questionnaire with more than 100 respondents. The questionnaire is in the form of questions asked for the general public based on the DASS-42 questionnaire which contains 42 questions. The tweets used are only focused on tweets in Indonesian, for other languages such as English and so on will be translated.

2. RESEARCH METHODOLOGY

2.1 System Design

In this research process to perform detection, there are several steps, such as data collection, data preprocessing, model building, and evaluation. Figure 1. Below shows an overview of the system design that will be made.

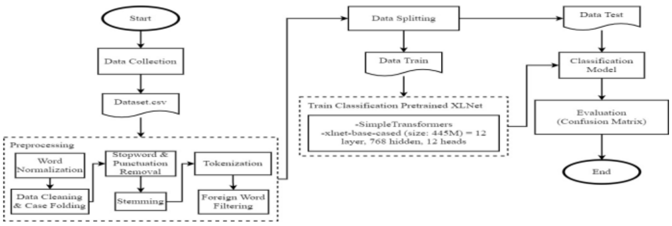

Figure 1. Research Method

2.2 Data Collection

The first stage of data collection carried out in this research was to collect Twitter user data from respondents who filled out the DASS-42 (Depression Anxiety And Stress-42) questionnaire. This questionnaire is for labeling the data. DASS-42 is one of the measurement tools that can be used to measure the severity of 3 scales, such as depression, anxiety, and stress. In the DASS-42 questionnaire, there are 42 questions, which for each scale has 14 questions. Table 1 shows the distribution items of the scales [8].

Table 1. DASS-42 Questionnaire Items of Scales

Scales	Items
Depression	3, 5, 10, 13, 16, 17, 21, 24, 26, 31, 34, 37, 38, 42
Anxiety	2, 4, 7, 9, 15, 19, 20, 23, 25, 28, 30, 36, 40, 41
Stress	1, 6, 8, 11, 12, 14, 18, 22, 27, 29, 32, 33, 35, 39

The 42-item questionnaire contains questions about self-assessment, which is done by filling in a value of 0 to 3 for each item. With 0 beings 'does not occur', 1 being 'rarely occurs', 2 being 'sometimes occurs', and 3 being 'frequently occur'. Using the DASS-42 as a measure of depression, anxiety, and stress, the total calculated score for each scale is 3 x 14 which is 42. Table 2 shows the DASS-42 scoring guide[9].

Table 2. DASS-42 Scoring Guide

The Severity of The Disorder	Scales		
	Depression	Anxiety	Stress
Normal	0-9	0-7	0-14
Light	10-13	8-9	15-18
Moderate	14-20	10-14	19-25
Heavy	21-27	15-19	26-33
Extremely Heavy	28-42	20-42	34-42

In this research, only using the depression scale to label respondents as indicating depression if the person has a score above 9 (10 to 42) or whose severity of the disorder is above 'light' will be labeled as indicating depression. Then crawling the tweets of respondents' usernames who had filled out the questionnaire for the dataset. The crawling of tweets is done without date and keyword restrictions.

2.3 Preprocessing

Datasets that have been labeled will enter the preprocessing. The process of preparing the raw dataset is also called preprocessing. Preprocessing aims to transform unstructured data into more structured data and stored it in a database of rules, so that the dataset will be easier for the system to accept or process [10]. In this research, the

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 379−385
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2157

preprocessing stages carried out are word normalization (turning slang and short words into normal words), data cleaning & case folding (cleaning data from unnecessary characters such as mention, URL, etc., and converting capital letter data into normal letters), stopword & punctuation removal (removing stopwords and symbols, which have no meaning), and stemming (converting words into their initial or root words]). Table 3 shows the preprocessing process performed.

Table 3. Preprocessing Steps

Preprocessing	Tweet
Raw Tweet	kdg aku gk tau kabar mereka sm sekali, aku jadi khawatir gmn mereka di sana ya AKU MAU PULANG BUTUH LIBURAN!! [URL]
Word Normalization	kadang aku tidak tahu kabar mereka sama sekali, aku jadi khawatir bagaimana mereka di sana iya AKU MAU PULANG BUTUH LIBURAN!! [URL]
Data Cleaning & Case Folding	kadang aku tidak tahu kabar mereka sama sekali aku jadi khawatir bagaimana mereka di sana iya aku mau pulang butuh liburan
Stopword & Punctuation Removal	kadang tidak kabar khawatir kabar iya pulang butuh liburan
Stemming	kadang tidak kabar khawatir kabar iya pulang butuh libur

2.4 Text Classification using pre-trained XLNet

XLNet is a generalized autoregressive language model that learns unsupervised text sequence representation. It incorporates modeling techniques from Auto-encoder (AE) models (BERT) into Auto-regressive (AR) models while avoiding the limitations of AE. XLNet focuses on the pre-trained phase. In the pre-trained phase, XLNet proposes a new goal called Permutation Language Modeling (PLM). Using permutation operations during training time, bidirectional context information can be captured and make it a general order-aware autoregressive language model.

XLNet uses the Two-Stream Self-Attention Architecture is used to overcome the problems posed by traditional Transformers. One of the two-stream self-attention is content stream representation, similar to the standard self-attention in Transformers that considers both content and positional information. The other one is the query representation, this essentially replaces the MASK of BERT, learned by the query stream attention to predict content only with positional information but not the content. Only the position information of the target token and context information before the token is available [14]. Figure 2 below shows an illustration of the Two-Stream Self-Attention Architecture that shows (a) the illustration of the content stream attention, which is the same as the standard self-attention, and (b) the illustration of the query stream attention, which does not have access information about the content variable x, also (c) the illustration of the overview of the permutation language modeling objective for predicting x given the same input sequence x but with different factorization orders.

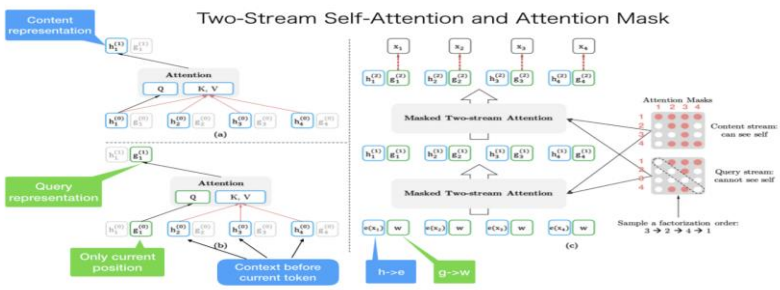

Figure 3. Illustration of the Two-Stream Self-Attention Architecture and Permutation Language Modeling

The process carried out before modeling is to divide the data into train data, test data, and data validation using the K-Fold Cross Validation library. By using K-Fold Cross Validation, each iteration of the dataset will alternate into training data and testing data [15]. Then, classify the text using the Simple Transformer library by initializing the task-specific Classification Model. Then train the model with the train_model() function. The model is trained by reading the train data and starting to perform text classification [16]. In this research, the XLNet method was tested. Testing is done by finding the best data sharing ratio and removing stemming from the preprocessing stage. In addition, testing was also carried out by adding hyperparameters for fine-tuning the XLNet model in Table 4. In addition, testing was also carried out using a dataset that filters out foreign words where

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 379−385
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2157

foreign data is translated into Indonesian. The data stored is data that uses words based on the KBBI dictionary. When the model is trained, for each step, validation is performed using data validation. Then, test data will be used to evaluate the performance of the learned model.

Table 4. Hyperparameters Used for Fine-tuning XLNet

Hyperparameter	Values
adam epsilon	1e-8
eval batch size	56
gradient accumulation steps	2
learning rate	1e-5, 5e-5
num train epoch	5, 10, 15
train batch size	56
scheduler	constant schedule with warmup linear schedule with warmup cosine schedule warmup
warmup ratio	0.06
warmup steps	50
weight decay	0.1

2.5 Performance Evaluation (Confusion Matrix)

Performance evaluation is the process of calculating the accuracy value. Calculation of performance evaluation using the Confusion Matrix method. Accuracy calculation is carried out to determine the level of closeness between the predicted value and the observed value. The formula performed has four outputs, namely recall, precision, accuracy, and error rate [17]. Confusion matrix is used in performance calculations as in Table 5. In this research, only the accuracy value is used because it will determine how well the system model can predict the data correctly. After measuring the performance, the system model that has been built is stored and will later be used to predict the percentage of Twitter users' depression levels taken from the dataset.

Table 5. Confusion Matrix

	Actual Positive	Negative Actual
Positive Prediction	TP (True Positive)	FP (False Positive)
Negative Prediction	FN (False Negative)	TN (True Negative)

With these 4 terms, an accuracy calculation is performed to determine the level of closeness between the predicted value and the actual or observed value. If the prediction is positive and the actual value is positive is True Positive (TP), if the prediction is positive and the actual value is negative is False Positive (FP), if the prediction is negative and the actual value is positive is False Negative (FP), if the prediction is negative and the actual value is negative is True Negative (TN). The accuracy value is obtained from formula 2.1, the precision value is obtained from formula 2.2, the recall value is obtained from formula 2.3 and the f1-score value is obtained from formula 2.4 [18].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

$$Precision = \frac{TP}{FP+TP} \tag{2}$$

$$Recall = \frac{TP}{FN+TP} \tag{3}$$

$$F1-Score = \frac{2\times(Precision\times Recall}{Presisi + Recall} \tag{4}$$

3. RESULTS AND DISCUSSIONS

3.1 Dataset

In this research, we distributed questionnaires to the general public where the respondents who filled out the questionnaires were used to label the datasets. Table 6 shows the top four rows of the results of the dataset labeling that has been done.

Table 6. Top Four Rows Dataset

Account Name	Tweet	Total DASS-42 Score	Level of Depressive Disorder	Label

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 379−385
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2157

Akun A	[@mention_name] Ya beda-beda atuh, ga bisa disamaratakan [@mention_name] Coba minta daftar tempat yg pernah di-KP-in ke angakatan-angkatan sebelumnya Awkward banget untuk introvert sepertiku [@mention_name]	31	Extremely Heavy	1
Akun B	ya allah mau punya cewe muka bad bitch:((gagitu konsepnyaa, we still can change this. hold on plisss [@mention_name] pengencptkaya Yg di tiktok bukan?	6	Normal	0
Akun C	[@mention_name] Org sukabumi pada tilil gini kenapa ya? [@mention_name] [@mention_name]	10	Light	1
Akun D	"coba cache sama cookiesnya dihapus dulu kang" [@mention_name][@mention_name] kirain setiap orang dikirim satu", hebat" wkwkwk [testing] multiple image no card [URL] [testing]	22	Heavy	1

3.2 Experimental Result

In this research, two scenarios are carried out to find the best performance results by finding the best data sharing ratio, looking at the comparison of model accuracy using data that removes stemming from the preprocessing stage and those that do not. In addition, tests were also conducted by adding hyperparameters for fine-tuning the XLNet model. In addition, testing was also carried out using a dataset that filters out foreign words where foreign data is translated into Indonesian. The data stored is data that uses words based on the KBBI dictionary. When the model is trained, for each step, validation is performed using data validation. From the classification is performed, and accuracy is obtained, which can be said to be the percentage of success in classifying data against sentiment labels.

3.2.1 First Scenario Test (Best Data Split Ratio and Eliminating Stemming in Preprocessing Stage)

The first scenario stage is to find the best data sharing ratio and determine whether to use a dataset by removing stemming in the preprocessing stage or not. The performance results obtained are expressed by the accuracy value. Comparison of data sharing ratios tested were 90:10 with data validation taking 10% of the train data, 80:20 with data validation taking 20% of the train data, and 67:33 with data validation taking 33% of the train data. A comparison of model performance accuracy values can be seen in Table 7 below.

Table 7. First Scenario Test Results

	Data Split Ratio	**Num Train Epoch**	**Accuracy**
With Stemming	90:10	5	50.00%
		10	50.00%
		15	50.00%
	80:20	5	52.00%
		10	60.00%
		15	60.00%
	67:33	5	60.00%
		10	68.57%
		15	68.57%
Without Stemming	90:10	5	**73.33%**
		10	**73.33%**
		15	66.67%
	80:20	5	60.00%
		10	64.00%
		15	60.00%
	67:33	5	52.00%
		10	68.57%
		15	50.00%

3.2.2 Second Scenario Test (Fine-Tuning)

In the second scenario stage, fine-tuning is performed with predefined hyperparameters. During the model training, the model is validated for each step using data validation. From the validation that has been done with 14 validation data, the accuracy value of the model changes at each epoch, sometimes at the beginning of the epoch a high accuracy value is obtained but in the last epoch, the accuracy value is lower than the beginning of the epoch and vice versa. From the validation results, the stored accuracy value is the accuracy value at the last epoch as the last accuracy value of the model that has been trained with a certain number of epochs. Table 8 shows the stored accuracy value for each epoch, learning rate, and scheduler performed. In this second scenario test, the best model

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 379−385
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2157

is obtained with hyperparameters that use learning rate 5e-5, scheduler cosine schedule with warmup, and num train epoch as many as 15.

Table 8. Second Scenario Test Results

Learning Rate	Num Train Epoch	Scheduler		
		Constant Schedule with Warmup	Linear Schedule with Warmup	Cosine Schedule with Warmup
1e-5	5	50.00%	57.14%	64.29%
	10	57.14%	50.00%	57.14%
	15	50.00%	57.14%	50.00%
5e-5	5	50.00%	57.14%	50.00%
	10	50.00%	50.00%	57.14%
	15	64.29%	64.29%	**71.43%**

3.2.3 Third Scenario Test (Using Foreign Language Filters For Datasets)

In the third scenario stage, after obtaining the best hyperparameters for fine-tuning XLNet, testing is carried out using datasets that have been added with foreign language filters in the preprocessing stage. The foreign language filter stage is tokenized, then translated into Indonesian, the last one is filtering words based on the KBBI dictionary. Table 9 shows the results of the dataset which has been filtered out of foreign languages.

Table 9. Dataset which has Been Filtered Out of Foreign Languages

Preprocessing	Tweet
Tweet	capek lelah feel sleepy
Tokenization	[capek, lelah, feel, sleepy]
Translated	[capek, lelah, merasa, mengantuk]
Filtered	capek lelah merasa mengantuk

From this test, the accuracy value increased by 7.14%. Table 10 shows the results of the tests that have been carried out.

Table 10. Third Scenario Test Results

Data Split Ratio	Num Train Epoch	Learning Rate	Sceduler	Accurac y
90:10	15	5e-5	Cosine Shedule With Warmup	78.57%

3.2.4 Analysis of Experimental Results

Based on the results of three test scenarios using XLNet, the best model is a model trained with hyperparameters for fine-tuning by removing stemming in the preprocessing stage, using a data division ratio of 90:10 (144 train data, 15 test data and 14 validation data), learning rate 5e-5, scheduler cosine schedule with warmup and num train epoch as much as 15 with an accuracy value of 71.43%. By using a learning rate of 5e-5 (0.00032) where the model will learn the data or the training process runs quickly and more thoroughly than using a learning rate of 1e-5 (1) [19]. Using a scheduler will improve the final result but the increase is not guaranteed, therefore it is added with warmup where the learning rate increases and to cool the rate until the end of the optimization process [20]. In the tests that have been carried out, the scheduler that gets the best accuracy value is to use the cosine schedule with warmup. Then, by using these hyperparameters, testing is carried out using a dataset that has been filtered for foreign languages to get an accuracy value of 78.57%. The following Table 11 shows the accuracy, precision, recall, and f1-score values for the model with the best performance.

Table 11. Best Model Performance Results

Data Split Ratio	Num Train Epoch	Learning Rate	Sceduler	Accurac y	Presisi	Recall	F1-Score
90:10	15	5e-5	Cosine Shedule With Warmup	78.57%	75.00%	85.71%	80.00%

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 379−385
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2157

4. CONCLUSION

In this research, a prediction of depressive disorders based on DASS-42 on Twitter has been carried out using the XLNet pretrained text classification model with tweet data from Twitter which is crawled and labeled based on the results of the DASS-42 questionnaire. Based on the results of the research conducted, show how to build a model that can predict depression from Twitter tweets using XLNet classification text. The first stage is collecting data, then preprocessing, splitting data, performing XLNet classification, and evaluating the performance of the model that has been built. Preprocessing consists of 4 processes, namely word normalization, data cleaning & case folding, stopword & punctuation removal, and stemming. The model performance accuracy value is obtained from applying all test scenarios in this research, namely by removing stemming in the preprocessing stage, using a data division ratio of 90:10 (144 train data, 15 test data, and 14 validation data), and adding hyperparameters for fine-tuning. The hyperparameters obtained for the best model are learning rate 5e-5, scheduler cosine schedule with warmup, and num train epoch of 15. After that, by using these hyperparameters and using a dataset that has been filtered for foreign languages from the test results the best performance obtained in this research is the accuracy of 78.57%.

REFERENCES

[1] Islam, Md. R., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. Health Information Science and Systems, 6(1). https://doi.org/10.1007/s13755-018-0046-0
[2] Cunha, M., Ribeiro, A., & André, S. (2016). Anxiety, Depression and Stress in Patients with Rheumatoid Arthritis. Procedia - Social and Behavioral Sciences, 217. https://doi.org/10.1016/j.sbspro.2016.02.098
[3] Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behaviour Research and Therapy, 33(3). https://doi.org/10.1016/0005-7967(94)00075-U
[4] Rodrigo Martınez-Castano1, Amal Htait, Leif Azzopardi, and Yashar Moshfeghi. (2020). Early Risk Detection of Self-Harm and Depression Severity using BERT-based Transformers. http://ceur-ws.org/Vol-2696/paper_50.pdf
[5] Al-Rfou, R., Choe, D., Constant, N., Guo, M., & Jones, L. (2019). Character-level language modeling with deeper self-attention. 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019. https://doi.org/10.1609/aaai.v33i01.33013159
[6] Kevin Clark, Minh-Thang Luong, Urvashi Khandelwal, Christopher D. Manning, and Quoc V. Le. (2019). Bam! born-again multitask networks for natural language understanding. anonymous preprint under review.
[7] Christian, H., Suhartono, D., Chowanda, A., & Zamli, K. Z. (2021). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. Journal of Big
[8] S. Kusumadewi and H. Wahyuningsih, "Model Sistem Pendukung Keputusan Kelompok untuk Penilaian Gangguan Depresii, Kecemasan dan Stress Berdasarkan DASS-42," J. Teknol. Inf. dan Ilmu Komput., vol. 7, no. 2, p. 219, 2020, doi: 10.25126/jtiik.2020721052.
[9] Depression Anxiety and Stress Scale DASS (-42). https://www.healthfocuspsychology.com.au/tools/dass-42/
[10] Buani, D. C. P. (2016). Optimasi Algoritma Naive Bayes dengan Menggunakan Algoritma Genetika untuk Prediksi Kesuburan (Fertility). Jurnal Evolusi
[11] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. v. (2019). XLNet: Generalized autoregressive pretraining for language understanding. Advances in Neural Information Processing Systems, 32.
[12] Pedregosa, F. and Varoquaux, G., Gramfort, A. and Michel, V., Thirion, B. and Grisel, O., Blondel, M. and Prettenhofer, P., Weiss, R. and Dubourg, V., Vanderplas, J. and Passos, A., Cournapeau, D. and Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12(85).
[13] Singh, A., & Jain, G. (2021). Sentiment Analysis of News Headlines Using Simple Transformers. 2021 Asian Conference on Innovation in Technology, ASIANCON 2021. https://doi.org/10.1109/ASIANCON51346.2021.9544806
[14] Visa, S., Ramsay, B., Ralescu, A., & van der Knaap, E. (2011). Confusion matrix-based feature selection. CEUR Workshop Proceedings, 710.
[15] Dj Novakovi, J., Veljovi, A., Ili, S. S., Zeljko Papi, ˇ, & Tomovi, M. (2017). Evaluation of Classification Models in Machine Learning. In Theory and Applications of Mathematics & Computer Science (Vol. 7, Issue 1).
[16] Smith, S. L., Kindermans, P. J., Ying, C., & Le, Q. v. (2018). Don't decay the learning rate, increase the batch size. 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings.
[17] Loshchilov, I., & Hutter, F. (2017). SGDR: Stochastic gradient descent with warm restarts. 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings.