Sentiment Analysis of Learning from Home During Pandemic Covid-19 in Indonesia

Bambang Widoyono[1,*] Indra Budi[1] Prabu Kresna Putra[1] Aris Budi Santoso[1]

[1]*Master of Information Technology, University of Indonesia, Jakarta, Indonesia*
[*]*Corresponding author. Email: bambangwidoyono@ui.ac.id*

ABSTRACT
Covid-19 pandemic in Indonesia impacting education sector issuance policy for learning from home. This policy has many responses from the public on social media like Twitter. There are many types of methods using machine learning approach to classify and predict sentiment analysis about this topic on Twitter. A good method with the best performance is required to increase chances of getting correct prediction and classification. To answer these problems and solve these needs, this research aims to find the best performing classification and prediction methods for analysis. This research objective is also expected to provide an overview for practitioners as a reference for researchers in conducting research. This research used a dataset consisting of 71.70% negative sentiment, 11.78% positive sentiment, and 16.52% neutral from 27708 response data on Twitter. Sentiment grouping by 3 annotators has an average annotation agreement level of 80.57%. Logistic regression produced the best performance compared to 6 other methods with performance accuracy 98.03% and f1-score 92.69%. Information extraction that causes negative sentiment mostly comes from the process of self-adaptation to the learning from home policy. Meanwhile, the provision of the Ministry of Education and Culture's quota assistance has a major influence on positive sentiment. Information extraction that affects to this sentiment can be used by the government to improve the application of learning from home policies such as providing internet quotas which have been proven to affect positive sentiment.

Keywords: *Twitter, sentiment analysis, Covid-19, learning from home, machine learning, classification*

1. INTRODUCTION

The Covid-19 pandemic began to occur in Indonesia in March 2020 impacting all areas of life [1]. The impact in the field of education is issuance policy to learning from home, to reduce the spread of the Covid-19 virus [2]. So that efforts to improve the quality of education which are being intensified will receive additional new challenges with this policy [3]. The policy of implementing learning from home has received many responses and various opinions from the public. Responses and opinions from part of the community are channelled and conveyed through posts and comments on social media such as Twitter. Post data and public comments on Twitter can be used to analyse the types of public sentiment regarding the learning process from home [4][5][6]. Posts on Twitter were chosen for analysis because the number of users in Indonesia is very large, namely 166 million users in May 2020 [7]. Not only because of the number of users, but also because the number of posts per user in a day can be more than once [8]. So that it is easier to obtain, more effective and efficient than spreading out surveys and questionnaires [9]. Public sentiment analysis is useful for predicting and classifying the tendency of opinions or responses to learning from home to be positive, negative or neutral. Prediction and classification of sentiment on posts on Twitter can be identified by implementing and analyzing machine learning through pattern recognition and predicting the next type of pattern that will appear based on the input parameter values that influence it [4][5][6]. There are many types of methods with a machine learning approach to the application of classification and prediction to sentiment analysis. A good method with the best performance when comparison method for evaluation required to get prediction and classification. Correct prediction and classification are also useful to help related parties accelerate the filtering of positive/negative sentiments in tracing the information contained therein to make improvements to fix that sentiment factor.

Based on those reasons for achieving the objective, this research formulated this problem into several research question (RQ):

1 RQ1: What is the best classified method to predict and analysis sentiment from posting on Twitter?
2 RQ2: What are information that indicates the factors causing sentiment about learning from home?

To answer this problem and solve this need this research aim is to search and find the best performing classification and prediction method for analysis. With this method, information that indicates the factors causing sentiment about learning from home can be mapped into positive or negative. This research objective is also expected to provide

an overview for practitioners as a reference for researchers in conducting research.

1.1. Related Work

There are many interesting topics on Twitter that can be obtained and developed by sentiment analysis, one of which is about the Covid-19 pandemic. Data support that is easily accessible and available has attracted many researchers to conduct sentiment analysis from Twitter. Some of the research including:

1. Country Image in COVID-19 Pandemic. Huimin Chen et al. in their research [4] conducted a sentiment analysis regarding the image of China which experienced the first effects of the Covid-19 pandemic on the response of Twitter users based on news from international media such as the UK and USA. Huimen Chen grouped positive, neutral and negative sentiment analysis categories on politics, economy, foreign, culture, situation, measures and racism. The classification methods used are SVM and Ours. The results show that politics, culture and racism get negative opinion trends from the foreign community. This research provides suggestions for government follow-up in improving the image of Chinese state.
2. Sentiment Analysis of Tweet about Covid-19. Goran Matosevic et al. in their research [10] conducted a sentiment analysis on Twitter users in April 2020 in 6 selected countries: the US, UK, Spain, Italy, Sweden and Germany. There are two datasets formed in this research: first contains tweets about Covid-19 in certain countries, and second, contains tweets from top 10 politicians, Twitter users in certain countries. The purpose of this sentiment analysis is to determine people's emotional responses by grouping them into several categories, including anger, anticipation, disgust, fear, joy, sadness, surprise and trust. The research shows that Spain is the only country where fear is in the top rank, followed by sadness. Based on only a sample of 6 countries, the researchers recommend this technique to see the perspective of people's emotions in various countries.
3. Sentiment Identification in COVID-19 Specific Tweets. Manoj Sethi et al. in their research [11] collected data via the Twitter API with data containing the hashtags # Covid-19 and #coronavirus to conduct sentiment analysis and create predictive models. There are 3 datasets that are processed, dataset 1 and 2 are 10000 data, while dataset 3 is a combination of dataset 1 and 2. Sentiment analysis is carried out by comparing 6 different methods through a combination of 3 datasets. The results of [11] show that the decision tree and svm have a good performance in classifying positive and negative sentiments of 91% -93% and tend to be stable against the combination of the 3 datasets. These researchers recommend processing larger data with the addition of the n-gram approach for more accurate results.

1.2. Our Contribution

There are many literatures that can be found in online journal databases about classify method to prediction and analysis sentiment from posting on Twitter. But, when researchers search sentiment analysis of learning from home during pandemic covid-19 in Indonesia from 5 online database journals such as ACM, IEEE, Scopus, Sciencedirect and Proquest, the database shows no available journal. Based on this result condition, researchers want to provide sentiment analysis of learning from home during pandemic covid-19 in Indonesia by finding the best classified method to prediction and analysis sentiment from posting on Twitter. Best correct prediction and classification is also useful to help related parties accelerate the filtering of positive/negative sentiments in tracing the information contained therein to make improvements. This research project can be used to overview for practitioners as a reference for researchers in conducting research.

1.3. Paper Structure

The rest of the paper is organized as follows. The first section introduces background, identifies problem, formulation research question and the aim of the research. Section 2 introduces the preliminaries used in this paper, which include machine learning, classification method, dataset and pre-processing. Section 3 presents a methodology that will be applied in this research. Then, the methodology and result will be discussed and analysis to answer the research question in Section 4. Section 5 contains conclusion of the paper. Finally, Section 6 contains limitation of the paper and presents directions for future research.

2. BACKGROUND

2.1. Machine Learning and Classification Method

According to Thomas W. Edgar (2017) in his research [12], machine learning is a field of research that looks at the use of computational algorithms to convert empirical data into useful models. According to Wittek (2014) in his research [13], machine learning is a broad research field that is mainly concerned with finding patterns in empirical data. According to Kavakiotis et al., (2017) in his research [14], machine learning is a scientific field that discusses ways how machines adapt and learn from experience to recognize patterns. Based on the opinion of the 3 references, it can be concluded that machine learning is a science in developing computer science used for learning and adaptation processes to recognize patterns formed from empirical data. Machine learning is classified into 3 categories:

1. Supervised learning, where the system makes functions by recognizing patterns of training data labeled to predict.
2. Unsupervised learning, in which the system tries to deduce unlabelled data structures.
3. Reinforcement learning, where the system interacts with a dynamic environment.

This research uses a supervised learning solution approach to predict and conduct sentiment analysis. In supervised learning, the system must inductively "learn" a function called the target function, which is an expression of a model that describes data. To infer the best target function, the learning system, given the training set, considers alternative functions, which are called hypotheses and are denoted by h. In supervised learning there are two kinds of learning tasks: classification and regression. Classification models try to predict class differences, such as blood groups, whereas regression models predict numerical values.
According to Ge et al. (2017), classification is a method in supervised machine learning that is used to recognize and predict variables based on input parameters [15]. According to Kavakiotis et al. (2017), classification is one of the supervised machine learning categories which is used to predict results based on the functions that have been training, usually, prediction targets only consist of 2 categories [14]. Based on these 2 papers, it can be concluded that classification is one of the methods in machine learning which is supervised to predict the output results based on pattern recognition when training on the dataset. Some of the most common techniques are Naïve Bayes (NB), Decision Trees (DT), k-Nearest Neighbour (k-NN), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN).. In evaluating the performance results of the classification method, several ways are used to form a confusion matrix as shown below:

Table 1 Confusion matrix

		Prediction	
		YES	**NO**
Actual	**YES**	TP	FN
	NO	FP	TN

Based on table 1 explained that True Positive (TP) indicates that the model correctly predicts Positive cases as Positive, False positive (FP): indicates that the model incorrectly predicts Negative cases as Positive, False Negative: (FN): indicates that the model is wrong predicts Positive cases as Negative, True Negative (TN): shows that the model correctly predicts Negative cases as Negative. By using confusion matrix we can measure performance aspects including:

1. Measuring accuracy: (TP + TN) / (TP + FN + FP + TN), but accuracy is not suitable for class imbalance. So we need another evaluation category.
2. Sensitivity = TP / P
3. Specificity = TN / N
4. Precision = TP / (TP + FP)
5. Recall = TP / (TP + FN)
6. F1 Score = 2 * (Precision * Recall) / (Precision + Recall), the goal is to find average performance on recall and precision.

2.2. Dataset and Pre-Processing

According to Jason Brownlee (2017), a dataset is a set of data consisting of several row and column data that describes an example of data that will be observed using machine learning [16]. The dataset is needed to create a predictive model during training. The use of datasets in machine learning for classification is divided into, namely training and testing. In general, the proportion of training data is greater, namely 70% -80% of the total dataset. Machine learning performance is highly dependent on the quality of the dataset [11]. Dataset in this research is the result of crawling posts on Twitter so that it is included in text mining. The quality of a good dataset in machine learning is determined by the accuracy of annotations when labeling the target column on the dataset and pre-processing to reduce irrelevant data [13]. According to Andrew Kurbatov (2015), pre-processing is a feature selection process on a dataset to reduce the use of irrelevant data during the computation process of training models in machine learning so that it is faster [17]. According to Carlos Vladimiro (2019), pre-processing is a process to reduce redundant and irrelevant data in datasets before being used by machine learning [18]. Based on these two opinions, it can be concluded that pre-processing is a process of reducing irrelevant data and redundant data to optimize the computational dataset at the training stage of the machine learning model to be faster [14] [15] [16] [17]. In-text mining, there are several examples of pre-processing to reduce data, namely tokenizing is the separation of word features with spaces, normalizing words, for example being transformed into lowercase letters, removing stopwords to eliminate unnecessary words [14] [15] [16] [17].

3. METHODS

This section describes the methodology that has been made to be applied in this research. The process used consists of seven main processes, including:

1. Data collection
2. Annotation dataset
3. Data pre-processing & feature selection
4. Machine learning classification methods
5. Comparison of machine learning classification methods
6. Selected and determine classification methods
7. Describe sentiment analysis results

The flow of the main all processes is described in the following figure:

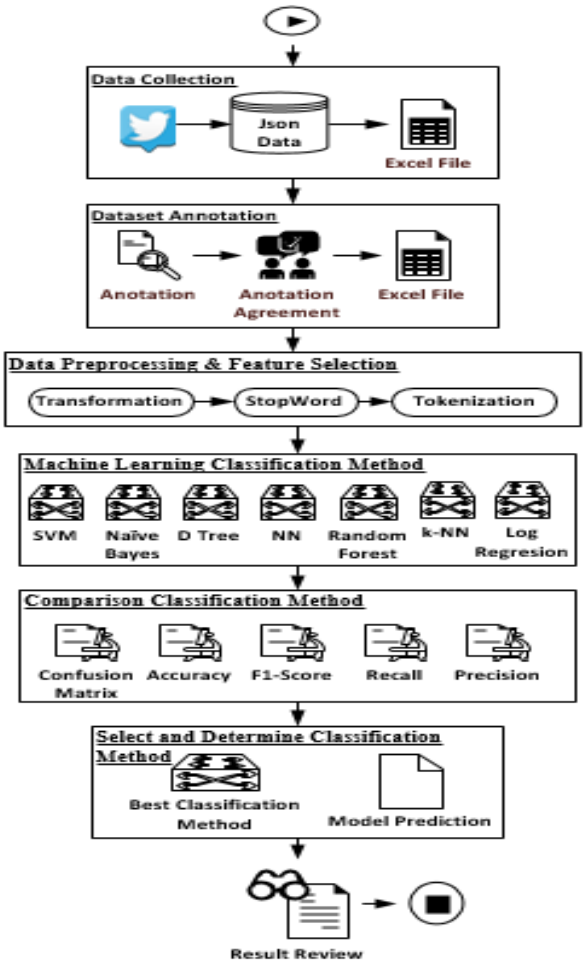

Figure 1 Flowchart diagram methodology

3.1. Dataset Collection and Annotation

This research uses data from posts on Twitter in November 2020 based on responses to learning from home policies during the pandemic in Indonesia. Data obtained through crawling using the Twitter API and then stored in the form of a json file. Because the Twitter policy is only allowed to be interesting for the previous 7 days, it is crawled periodically for 4 times. Any data that has been stored in json form is also transformed into an excel file to facilitate further processing. Data stored in Excel consists of only a few column values, including the data id, the date it was created, full_text which is the content of the post, display the name of the posting account owner and the location of the account owner. The annotation dataset is formed from a data collection labeled as the target for each data row. The target labels in the data collection are positive, negative and neutral. The target labeling of the data collection was done manually annotating using 3 annotators. The number of annotators is more than 1 and an odd number to get objective labeling targets. The chosen annotator has a criterion: a person who has used Twitter for at least 1 year and has a minimum education degree. Researchers assume that these criteria are indicators that the annotator understands how to group categories. In the annotation process, if there are labels with different values, the majority value of the 3 annotators is chosen. If all label values are different, a prediction analysis is carried out from the previous data. The final result at this stage is a dataset in excel format which is used for the training and testing process.

3.2. Pre-processing and Feature Selection

After obtaining the dataset, the next step is to pre-processing data. Data pre-processing aims to reduce irrelevant data so that the subsequent computation process is faster and more efficient. The pre-processing data used in this research included:

1. Transformation, which aims to homogenize content with lowercase, remove url, remove word accents and remove html formatting.
2. Stopword, which aims to eliminate unnecessary words to reduce the number of words. Like this word, that, that, and others.
3. Tokenization, which aims to make each sentence of the post content on Twitter broken down by word using delimited spaces and the character '_'.

The results of the pre-processing data will make it easier to calculate the number of words that are important for the classification process, namely the bag of words.

3.3. Comparison and Selecting Classification Method

Machine learning classification method will run after pre-processing data. This research carried out the classification method selection stage by applying the classification process to 7 classification methods including SVM, Naïve Bayes, Decision Tree, Neural Network, Random Forest, k-NN and Logistic Regression. The performance of these seven methods will be sought by looking at the resulting confusion matrix. All of these methods use training data as much as 75% of the total dataset and testing data using 25% of the total dataset.

At this stage, the results of the performance of each classification method are reduced with an accuracy threshold of at least 80%. After reduction during training, performance comparisons were made against the remaining classification methods that were not reduced during the testing stage. The performance that is compared between classification methods is seen from the aspects that can be calculated with confusion matrix, such as accuracy, F1-score, precision and recall.

Researchers choose the best classification method from the 7 methods used based on the highest performance results in the previous stage. The method selection is done by ranking the performance by prioritizing the accuracy value, F1-Score, and the recall value. The classification method with

the first ranking order will then be used to create a continuous prediction model of the commentary pattern regarding 'learning from home'.

3.4. Review and Describe Sentiment Analysis Result

At this stage, the researcher conducts a sentiment analysis using the resulting predictive model to categorize sentiments and find the information contained in them. Information that can be obtained among others:

1. Word or sentence structure as information that indicates the factors causing negative sentiment from the public towards applied learning from home.
2. Word or sentence structure as information that indicates the factors causing positive sentiment from the public towards applied learning from home.

With that information, we can create graph conversation about learning from home. Account Twitter as node and relation between account using retweet as an edge.

4. RESULT AND DISCUSSIONS

4.1. Dataset Result

The dataset is obtained by crawling via Twitter api using post filters from Indonesia and using a search query: 'learning from home' or a search query with the same meaning. Information describing the condition of the dataset is described in the following table:

Table 2 Dataset information

Attribute	Value	Attribute	Value
Total	27708	**Data Training**	25%
Annotator	3	**Data Testing**	75%
Annotator 1-2 Agreement	78.45%	**Sentiment Negative**	19867
Annotator 1-3 Agreement	80.85%	**Sentiment Positive**	3263
Annotator 2-3 Agreement	82..42%	**Sentiment Neutral**	4578
Source	Twitter	**Total Option**	23130

Based on the information in table 2, it is explained that the amount of raw data in the dataset is 27708. Based on the labelling of the dataset by 3 annotators, there are 19867 negative sentiments, 3263 positive sentiments, and 4578 neutral. The percentage level of confidence between annotators in labelling the dataset was 78.45% between the 1st annotators and the 2nd annotators, 80.85% between the 1st and 3rd annotators and 82.42% between the 2nd and 3rd annotators. The percentage level of confidence between 2 annotators is 100% if all labels given to each raw dataset have the same target label category. The following is an example of categorizing sentiment analysis by annotators:

1. An example of negative sentiment is "@schfess setelah diteliti lagi ternyata gua semakin goblok karena sekolah online, jadi gua di tim sekolah offline".
2. An example of positive sentiment is "@schfess HIDUP SEKOLAH ONLINE!".
3. An example of neutral sentiment is "lebih asik belajar online atau belajar offline? sudah tentu lebih asik bersamamu till the end dong".

The proportion of this labelled dataset is divided for training and testing, namely 75% for machine learning model training and 25% for testing. The total option was chosen to only predict 2 categories of sentiment, positive or negative.

4.2. Pre-Processing Result

Pre-processing aims to select data and reduce irrelevant and unimportant data so that the subsequent computation process is faster and more efficient. This research emphasizes the importance of pre-processing before the dataset is computed by machine learning. If you don't use pre-processing, the number of characters or irrelevant words can dominate computing during the training process in machine learning as shown below.

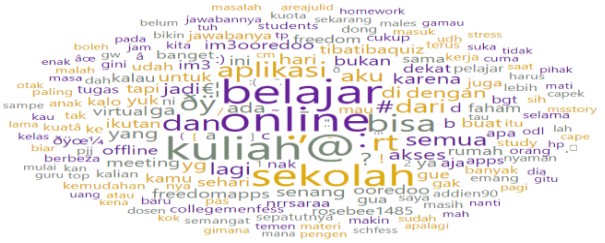

Figure 2 Word cloud before pre-processing.

Before pre-processing the dataset, the number of words or characters that appeared was dominated by punctuation and conjunctions. This can reduce machine learning accuracy performance and increase compute time, since the amount of data is not reduced. To solve this problem, research optimizes the process by using pre-processing. So the words that influence sentiment analysis really seem to dominate. After pre-processing the word count changes as shown below.

Figure 3 Word cloud after pre-processing

After pre-processing the dataset, the number of words or characters that appeared was no longer dominated by punctuation and conjunctions but was dominated by important words such as adverbs, object words and predicate words. So that it can improve machine learning

performance and speed up computing, because the amount of data has been reduced.

4.3. Comparison Performance

This research uses 7 classification methods on the same training and testing data. The seven classification methods were compared not only on accuracy performance but also on recall, precision and f1-score performance. The comparison was carried out by researchers in 2 stages, training and testing. Researchers use a minimum threshold of 80% accuracy. This was chosen because the proportion between categories is not balanced. The amount of negative sentiment is 72% of the total data, whereas if you use target data with 2 categories, then the negative sentiment has a proportion of 85.89%. With this threshold, if one of the classification methods does not have the accuracy, it will be reduced and not reused for performance comparisons during the testing process. Performance testing was carried out on a database consisting of 2 categories, namely negative and positive, so that the number of data used was 23130 with a training proportion of 75% as much as 17348 and 25% testing for 5782. The performance results during the training phase were for 2 categories, positive sentiment and negative sentiment. During test on training phase there is no accuracy below the threshold of 80%, so all classification methods used during training will be reused for comparisons during the testing process. The highest to lowest performance of the 7 sequential classification methods during process test on testing data are Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), k-Nearest Neighbour (k-NN), Neural Network (NN), Naïve Bayes (NB) and Support Vector Machine (SVM). Performance results during testing phase for 2 categories, positive sentiment and negative sentiment for each classification method are as shown in the following table:

Table 3 Test on testing data comparison performance

Model	Accuracy	Precision	Recall	F1-Score
LR	98.03%	96.92%	88.82%	92.69%
RF	97.87 %	96.25%	88.33%	92.12%
DT	97.72 %	96.21%	87.22%	91.49%
k-NN	97.42 %	98.54%	82.92%	90.06%
NN	85.39%	85.39%	89.07%	87.19%
NB	88.40%	55.41%	89.93%	68.57%
SVM	85.92%	-	0%	-

Based on the results of table 3 above, Logistic regression has accuracy value 98.03% and f1-score 92.69%. While SVM has accuracy value 85.92% but value on precision, recall and f1-score is zero. This condition on SVM performance because SVM does not has true positive rate (TP). Detail of differences performance between logistic regression (LR) and support vector machine (SVM) shown in the following table:

Table 4 Confusion matrix LR vs SVM

Model	TP	TN	FP	FN
LR	723	4945	23	91
SVM	0	4968	0	814

Based on table 4, using SVM method in this research, all data testing were predicted as negative sentiment. SVM never predict positive sentiment on 5782 data testing. This research proves the previous theory [17] [16] that accuracy is not suitable as a measure of performance on imbalanced data. With two result analysis, logistic regression which ranks first in performance compared to the other 6 methods, was chosen as the classification method chosen for predictive sentiment analysis regarding continuous learning from home.

4.4. Sentiment Analysis

Logistic regression as the best classification method that has been found used to predict sentiment and divided again to two categories, such as negative sentiment and positive sentiment. Extraction of information using bigram, trigram, 4-gram, 5-gram and 6-gram, then counted the number of occurrences of each extracted information. By using logistic regression, negative sentiment from Twitter about learning from home can be predicted and founded, such as post by @1Widha: "Pelajaran banyak gak masuk di otak karna kuliah online…" on 14th Nov 2020. The dominating information on negative sentiment can be seen in the following figure.

Figure 4 Word cloud negative sentiment

In Figure 4 above, it is explained that the amount of information affects the size of that information. The more similar information appears, the larger the word size is displayed. The author takes 10 samples to provide examples of information that affects the causes of negative sentiment and the number of occurrences. Ten samples of information on the causes of negative sentiment are listed in the following table.

Table 5 Information in negative sentiment

No	Information	Number
1	"…ga masuk…" "...masuk otak…" (couldn't catch a lesson)	304
2	"…ga ngerti…" (cannot be understood)	264
3	"…capek…" "…capek banget…" (tired)	248
4	"…stress..." "…susah…" (stressful and complicated)	236
5	"…tugas numpuk…" "…tugas…" (pile of tasks)	125
6	"…males…" "…bosen…" (lazy and bored)	193
7	"…nilai turun…" (grade score decrease)	148
8	"…ketemu teman…" (want to meet friends)	36
9	"…uang jajan..." (pocket money)	25
10	"…suka offline…" (like offline school)	22

Based on table 5 above, all information items can be grouped as factor complaints that come from the individual's adaptation process to condition of learning from home pandemic effect. In fact, this post data was taken in November 2020 while the learning from home policy in Indonesia has been in place since March 2020. This shows, even though it has been going on for more than 8 months there are still many negative sentiments towards this policy originating from the individual adaptation process every time. There is no issue about infrastructure in top 10 information that indicator causing negative sentiment of process learning from home. Meanwhile, using logistic regression positive sentiment from Twitter about learning from home can be predicted and founded, such as post by @chocholimbis: "makin kesini udah makin nyaman buat sekolah online" on 17th Nov 2020. The information that dominates positive sentiment can be seen in the following figure.

Figure 5 Word cloud positive sentiment

In Figure 5 above, it is also explained in figure 4 that the amount of information affects the size of the information. The more similar information appears, the larger the word size is displayed. The author takes 5 samples to provide examples of information that affects positive sentiment and the number of occurrences. Five samples of information that affect positive sentiment are listed in the following table.

Table 6 Information in positive sentiment

No	Information	Number
1	"…bantuan kuota…" "…kuota…" "…manfaatkan kuota…" (internet package support)	673
2	"…nyaman…" "…udah nyaman…" "…pw…" (comfortable)	458
3	"…sehat…" "…kondisi sehat…" (health support)	316
4	"…enak..." "…suka…" "…senang…" (good, like, happy)	149
5	"… optimis…" "…semangat" (optimism)	125

Based on table 6 above, information items 2-5 can be grouped as responses that come from the individual's adaptation process to conditions. This shows, after 8 months of the issuance of the learning from home policy, there are several stakeholders who have adapted well. Table 6 also shows that there is one item that affects positive sentiment originating from external factors, namely the provision of quotas by the ministry of education to stakeholders who participate when learning from home. The graph regarding the learning from home conversation from Twitter according to the dataset processed in this research is described as follows.

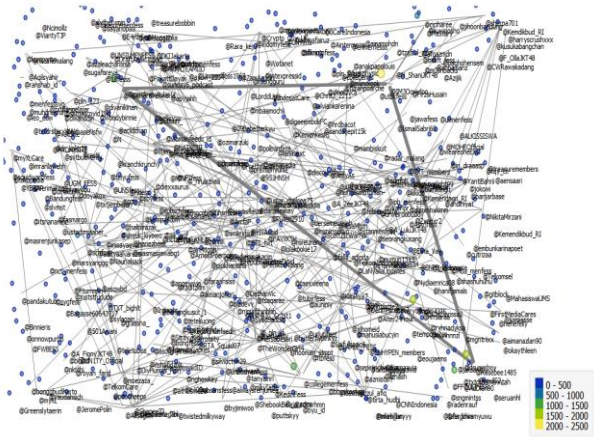

Figure 6 Graph learning from home

5. CONCLUSIONS

In this research, information was obtained about the proportion of sentiment towards the learning from home

policy, namely 71.70% negative sentiment, 11.78% positive sentiment, and 16.52% neutral from 27708 response data on Twitter. Sentiment grouping by 3 annotators has an average annotation agreement rate percentage value of 80.57%. The results show that logistic regression has the best performance compared to 6 other machine learning methods for classification, such as SVM, Naïve Bayes, Decision Tree, Neural Network, Random Forest, and k-NN on data testing. Logistic regression resulted in performance with accuracy value 98.03% and f1-score 92.69%. Meanwhile, SVM is ranked 7th with an accuracy value of 85.92%. This research also shows evidence such as the results on the SVM that measuring the performance of the classification method only with accuracy is not suitable because it is possible that high accuracy only detects one of the true negatives. Information extraction that causes negative sentiment mostly comes from the process of self-adaptation to the learning from home policy. There is no issue about infrastructure in top 10 information that indicator causing negative sentiment of process learning from home. Meanwhile, the provision of Ministry of Education and Culture quota assistance has a major influence on positive sentiment. Information extraction that affects this sentiment can be used by the government to improve the application of learning from home policies such as providing internet quotas which have been proven to affect positive sentiment.

6. LIMITATION AND FUTURE WORKS

This research was conducted with the limitation of taking the dataset only in November 2020. The number of annotators in this research was only 3 people with an average agreement percentage rate of 80.57%. Extraction of information on positive and negative sentiments using only n-grams and counting the number of words that often appear in each category.

Based on the limitations of this research, similar research needs to be developed by adding to expanding the dataset. To improve the quality of the annotation and labeling of the target dataset, researchers suggest that annotators who are more skilled so that the percentage agreement of annotations is higher than the present. This research has not applied normalization during pre-processing to homogenize words, so the researchers suggest word normalization to reduce word diversity and thus speed up computation. Researchers also suggest the use of better information extraction methods. So that the extracted information can be used easily to mitigate negative sentiment.

ACKNOWLEDGMENT

This work was supported by course assignments of Analytics Social Media, Master of Information Technology, University of Indonesia, Jakarta, Indonesia.

REFERENCES

[1]B. P. S. (BPS), “Statistik Pendidikan Indonesia 2019,” pp. 73–92, 2019.

[2] K. Pendidikan, “Pelaksanaan Kebijakan Pendidikan Dalam Masa Darurat Penyebaran Corona Virus (Covid-19),” p. 300, 2020.

[3] U. Nasional, “Pelaksanaan Kebijakan Pendidikan Dalam Masa Darurat Penyebaran Corona Virus (Covid-19),” p. 300, 2020.

[4] H. Chen *et al.*, “Country Image in COVID-19 Pandemic: A Case Study of China,” *IEEE Trans. Big Data*, vol. 14, no. 8, p. 1, 2020, doi: 10.1109/tbdata.2020.3023459.

[5] S. G. Kanakaraddi, A. K. Chikaraddi, K. C. Gull, and P. S. Hiremath, “Comparison Study of Sentiment Analysis of Tweets using Various Machine Learning Algorithms,” in *Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2020*, 2020, pp. 287–292, doi: 10.1109/ICICT48043.2020.9112546.

[6] A. Samuels and J. McGonical, “Sentiment Analysis on Social Media Content,” *arXiv*, no. Iciccs, pp. 911–915, 2020.

[7] Detik, “Jumlah Pengguna Twitter Meningkat,” 2020, [Online]. Available: https://inet.detik.com/cyberlife/d-5001786/jumlah-pengguna-Twitter-meningkat-tapi.

[8] Kompas, “Pengguna Aktif Harian Twitter Indonesia Diklaim Terbanyak,” 2019, [Online]. Available: https://tekno.kompas.com/read/2019/10/30/16062477/pengguna-aktif-harian-Twitter-indonesia-diklaim-terbanyak.

[9] D. A. Kristiyanti, Normah, and A. H. Umam, “Prediction of Indonesia presidential election results for the 2019-2024 period using Twitter sentiment analysis,” *Proc. 2019 5th Int. Conf. New Media Stud. CONMEDIA 2019*, pp. 36–42, 2019, doi: 10.1109/CONMEDIA46929.2019.8981823.

[10] G. Matosevic and V. Bevanda, “Sentiment Analysis of Tweet about Covid-19,” pp. 1290–1295, 2020.

[11] M. Sethi, S. Pandey, P. Trar, and P. Soni, “Sentiment Identification in COVID-19 Specific Tweets,” *Proc. Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2020*, no. Icesc, pp. 509–516, 2020, doi: 10.1109/ICESC48915.2020.9155674.

[12] D. O. M. Thomas W. Edgar, “Machine Learning in Research Methods for Cyber Security,” 2017.

[13] P. Wittek, “Quantum Machine Learning,” *Acad. Press*, p. Pages 11-24, 2014.

[14] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, “Machine Learning and Data Mining Methods in Diabetes Research,” *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 104–116, 2017, doi: 10.1016/j.csbj.2016.12.005.

[15] Z. Ge, Z. Song, S. X. Ding, and B. Huang, “Data Mining and Analytics in the Process Industry: The Role of Machine Learning,” *IEEE Access*, vol. 5, pp. 20590–20616, 2017, doi: 10.1109/ACCESS.2017.2756872.

[16] J. Brownlee, “Deep Learning for Natural Language Processing Develop Deep Learning Models for Natural Language in Python,” *Mach. Learn. Mastery*, pp. 1–17, 2017, [Online]. Available: http://web.stanford.edu/class/cs224n/readings/cs224n-2019-notes06-NMT_seq2seq_attention.pdf.

[17] A. Kurbatow, “The research of text preprocessing effect on text documents classification efficiency,” *2015 Int. Conf. "Stability Control Process. Mem. V.I. Zubov, SCP 2015 - Proc.*, pp. 653–655, 2015, doi: 10.1109/SCP. 2015.7342234.

[18] C. V. G. Zelaya, “Towards explaining the effects of data preprocessing on machine learning,” *Proc. - Int. Conf. Data Eng.*, vol. 2019-April, pp. 2086–2090, 2019, doi: 10.1109/ICDE.2019.00245.