AN ANALYSIS OF BAYESIAN ALGORITHM IN PREDICTING THE INTERESTS OF HIGH SCHOOL GRADUATES TO STUDY FURTHER INTO UNIVERSITY USING SOCIAL MEDIA TWITTER

Suhartono Lolo[1], Stopira Ricambi[2]
[1,2]Bina Nusantara University; Jl. Kebon Jeruk Raya No. 27, Jakarta 11530, Indonesia
[1,2]Master of Information System Management, Binus Graduate Program
e-mail: [1]suhartono.lolo001@binus.ac.id, [2]stopira.ricambi@binus.ac.id

Abstract

Sentiment analysis is a very necessary feature in collecting information about the user preferences about a topic on social media. Analyzing the user sentiments is important to find out negative comments and positive comments. This study aims to classify tweet data into 2 sentiments namely positive and negative. In this study, we collect the Indonesian texts dataset from big data of Twitter social media. The collected dataset from tweeter is called as tweets bag of word is used in this study to predict the student preferences about university study. After extracting the datasets, we get 1978 example of tweets data containing both positive and negative opinions which will be classified using a text mining approach of Naïve Bayes algorithm. Before classification, several stages of text processing are carried out such as case folding, normalization, tokenization and stopwords removal. There are 113 negative tweets, 1744 neutral tweets and 116positive tweets. We also test the accuracy of the Naïve Bayes algorithm and get 85% accuracy rate. We implied that the tweet analysis can be used by university decision maker as information for decision maker to build strategy and interest of high school graduates to continue studying to university.

***Keywords**:* Sentiment analysis, Text Mining, Naïve Bayes, classification, twitter.

1. INTRODUCTION

This time youth generation which so-called millennial youth has wide knowledge about social media which help them to get and share knowledge through online platform. They also have wider access to social media application and share their opinion which important for both them and related stakeholders. Their accumulative opinions lead to a trend of sentiments which can be positive or negative toward certain topics. One of the implementations of the sentiments is toward their preferences to study further to university. In Indonesia, university is a higher education institution after the student graduated from general school or college [1]. The university is a level of education after secondary education of Senior High School (SMA).

Based on data from the Central Statistics Agency the number of graduates of high school at the age of 19-24 years has increases every year. However, the number of the graduates to continue study further is slowing down.

It means that the participation rate of the graduates to continue study to university is difficult to predict with manual approach. Even though big university has tendency to get increased number of students, however, for small university, it sometimes has difficulty to get new students which causing the university to lack of student's numbers. The university decision makers need a tool to know the cause of the decreased number of new students entering their university.

As marketing strategy, the public interest in continuing to college needs to be cultivated in every high school graduate. The graduates who have an interest in continuing to a higher level of education will have a sense of interest and are motivated to study harder, so they can compete with other students.

One of the most popular social media services today is Twitter. Twitter has produced 110 million tweets every day and has more than 200 million users [2]. Lots of research on text mining makes social media as a medium to get sentiment or poll information. In addition, Twitter is often used by users as a medium to publicize daily activities or places to express what users feel [3]. Many Twitter users unconsciously provide information about their personality through tweets or posts they make in natural language [4]

In previous studies the Naive Bayes method was also used in predicting neglected dermatological diseases but could even cause death where the Naive Bayes method was used to recognize data patterns to reveal the possibility of dermatological diseases [5]. The Naive Bayes method is also considered to have good potential in classifying documents compared to other classification methods in terms of computational accuracy and efficiency [6]. Previous research has been conducted relating to the problems faced by the author, namely the title "Sentiment Analysis on Twitter for Regarding the Use of Public Transportation in Cities with the Support Vector Machine Method". The study was conducted by Novantirani, Anita., et al in 2015. In this study the results of testing and analysis conducted by the author with the support vector machine method, with accuracy reached 78.12%. Accuracy results on the use of the support vector machine method are influenced by several things, namely the composition of the amount of training and testing data, the number of datasets used and the composition of the amount of positive and negative data. [7]

The purpose of this research is to classify sentiments in tweet data and use text mining with the Naïve Bayes method for sentiment analysis on Twitter social media.

The benefits of this research are as follows:

a. This research is expected to provide additional knowledge about the implementation of the NB (Naïve Bayes) algorithm and can provide an overview of how sentiment analysis on Twitter social networks to classify opinions for various interests and optimize social network information for the public interest.
b. It is also useful to find information about a product or brand to find out the market response to the product or brand whether it has a positive or negative response from the market and is also beneficial for public figures to measure the public response to themselves.
c. This study is also useful to find out how successful the NB (Naive Bayes) algorithm is in classifying Indonesian texts.

2. THEORITICAL FRAMEWORK

2.1. Naive Bayes Definition

Naive Bayes is a simple probabilistic classification that calculates a set of probabilities by adding up the frequency and combination of values from a given dataset. The algorithm uses the Bayes theorem and assumes all the independent or non-interdependent attributes given by values to class variables [8]. Another definition says that Naive Bayes is a classification with probability and statistical methods presented by the British scientist Thomas Bayes, which predicts future opportunities based on experience in the past [10].

Naive Bayes is based on the simplification assumption that attribute values are conditionally mutually independent if output values are given. In other words, given the value of output, the probability of observing together is a product of individual probabilities [8]. The advantage of using Naive Bayes is that this method only requires a small amount of training data

to determine the estimated parameters needed in the classification process. Naive Bayes often works far better in most complex real-world situations than expected [9].

2.2. The Naive Bayes Method Equation

The equation of the Bayes theorem is [10]:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \text{ .....(1)}$$

where:
X: Data with unknown classes
H: The data hypothesis is a specific class
P (H | X): H hypothesis probability based on condition X (posteriori probability)
P (H): Hypothesis probability H (prior probability)
P (X | H): Probability of X based on conditions on the hypothesis H
P (X): Probability of X
To explain the Naive Bayes method, it is important to know that the classification process requires a number of clues to determine what class is suitable for the analyzed sample. Therefore, the Naive Bayes method above is adjusted as follows:

$$P(C|XF1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)} \text{ .....(2)}$$

To explain the Naive Bayes theorem, it is important to know that the classification process requires a number of clues to determine what class is suitable for the analyzed sample. Therefore, the Bayes theorem above is adjusted as follows. The above formula can also be written simply as follows:

$$Posterior = \frac{Prior\ x\ likehood}{evidence} \text{ .....(3)}$$

Evidence values are always fixed for each class in one sample. The value of the posterior will later be compared with the value of the posterior grades of other classes to determine to which class a sample will be classified. Further elaboration of the Bayes formula is carried out by describing (C | F1 ... Fn) using the following multiplication rules:
P (C | F1 ... Fn) = P (C) P (F1 ... Fn | C)
= P (C) P (F1 | C) P (F2, ... Fn | C, F1)
= P (C) P (F1 | C) P (F2 | C, F1) P (F3, ... Fn | C, F1, F2)
= P (C) P (F1 | C) P (F2 | C, F1) P (F3, ... Fn | C, F1, F2) P (F4, ... Fn | C, F1, F2, F3)
= P (C) P (F1 | C) P (F2 | C, F1) P (F3, ... Fn | C, F1, F2) P (Fn | C, F1, F2, F3, ..., Fn-1)……..(4)

It can be seen that the results of the elaboration cause more and more complexity of the condition factors that affect the probability value, which is almost impossible to analyze one by one. As a result, the calculation becomes difficult to do. Here is used a very high assumption of independence (naive), that each of the instructions (F1, F2, ... Fn) are independent of each other. With these assumptions, the following similarities apply:

For I ≠ j, such that $P(F_i|C, F_j) = P(F_i|C)$ .....(5)

The equation above is a model of the Naive Bayes theorem which will then be used in the classification process. For classification with continuous data the Gauss Density formula is used:

$$P(X_i = x_i|Y = y_{ij}) = \frac{1}{\sqrt{2\pi\sigma ij}} e^{\frac{(X_i - \mu_{ij})^2}{2\sigma^2 ij}} \text{ .....(6)}$$

where:
P: Opportunity
Xi: Attribute to i
xi: Attribute value to i
Y: Class sought
yi: Sub class Y is sought
μ: mean, represents the average of all attributes
σ: Standard deviation, denotes variants of all attributes.

3. RESEARCH METHOD

In this study, the data used is tweet specialization of high school graduates to continue to tertiary institutions found on Twitter. Tweets used are tweets containing opinions about interests in college. Total tweets used as data is 1978 tweets. The selection of data manually is to choose tweet sentences that are in Indonesian and do not contain pictures. Tweets that have been selected are then saved to an Excel file. The data needed in this study consists of two types, namely training data and test data. For training needs, the data collected will be categorized manually by the author and assess the sentiments contained in the Tweet and mark the Tweet into 2 sentiment categories, namely Tweets that contain negative and positive sentiments.

4. DISCUSSION

Preparing the data to be processed is applied by RStudio, in this study the data to be processed is text data from tweets that are on social media Twitter related to the interests of high school graduates to continue studying taken in August 2019. The data is entered into an excel file that is named kuliah.csv. Then the data is loaded into the RStudio application for further processing.

4.1. Data Extraction as Sampling

This study found a total of 10 variables using the 'userTimeline' function, a snapshot of the sample data shown in Table 1 below.

tweets.df

Filter

	text	favorited	favoriteCount	replyToSN	created	truncated	replyToSID	id	replyToUID	statusSource
1	the way how i always feel sick di minggu terakhir libu…	FALSE	0	NA	2019-08-28 16:44:51	TRUE	NA	1166753521070292992	NA	<a href="http://tw
2	Please give to me . I need a lot of tudung to go kuliah…	FALSE	0	NA	2019-08-28 16:12:03	FALSE	NA	1166745265438482432	NA	<a href="http://tw
3	± 2 tahun lalu. dah pisah jauh karena merantau kuliah…	FALSE	1	NA	2019-08-28 15:48:03	TRUE	NA	1166739227473633280	NA	<a href="http://tw
4	@myouimina Kuliah Kerja Nyata (final project that req…	FALSE	0	myouimina	2019-08-28 15:22:52	FALSE	1166732081033932800	1166732888382959616	225800682	<a href="http://tw
5	RT @mollycecils: Kuliah gonna start soon :((	FALSE	0	NA	2019-08-28 15:15:20	FALSE	NA	1166730994738532354	NA	<a href="http://tw
6	@askmenfess Sharing hal 18+ itu apa? Kerjaan? Kuliah…	FALSE	0	askmenfess	2019-08-28 15:12:33	TRUE	1166729872049197056	1166730294159990784	810423824034299904	<a href="http://tw
7	Untuk anak bapaknya yg lagi kuliah di unbraw, please…	FALSE	0	NA	2019-08-28 14:43:15	TRUE	NA	1166722920082591744	NA	<a href="http://tw
8	Come and see <U+24E2><U+1D40><U+1D2E><U+00…	FALSE	0	NA	2019-08-28 14:37:55	FALSE	NA	1166721578563858433	NA	<a href="http://ww
9	Listen to the most recent episode of my podcast: Ep …	FALSE	2	NA	2019-08-28 14:08:52	FALSE	NA	1166714266834984967	NA	<a href="http://tw
10	no offense but dapat beasiswa kuliah actually sucks.…	FALSE	0	NA	2019-08-28 13:58:34	FALSE	NA	1166711673886924800	NA	<a href="http://tw
11	I dont feel like going back to kuching and start kuliah…	FALSE	1	NA	2019-08-28 13:57:49	FALSE	NA	1166711485436837888	NA	<a href="http://tw
12	RT @mollycecils: Kuliah gonna start soon :((	FALSE	0	NA	2019-08-28 13:48:00	FALSE	NA	1166709017944608768	NA	<a href="http://tw
13	RT @mollycecils: Kuliah gonna start soon :((	FALSE	0	NA	2019-08-28 13:43:5[		NA	1166707985143353344	NA	<a href="http://tw
14	background: most indo unis require students to und…	FALSE	0	crowleygoblok	2019-08-28 13:42:1[		1166706538364334083	1166707578954321920	1107205717163806720	<a href="http://tw
15	Kuliah gonna start soon :((	FALSE	7	NA	2019-08-28 13:38:17	FALSE	NA	1166706569272184833	NA	<a href="http://tw
16	Yo, Dan! Mungkin kita gak terlalu deket pas kuliah, bu…	FALSE	0	NA	2019-08-28 13:28:54	TRUE	NA	1166704211142119430	NA	<a href="http://tw
17	Most of my friend asked me "ngapain udh mau 20 bar…	FALSE	0	NA	2019-08-28 13:22:33	TRUE	NA	1166702609421369345	NA	<a href="http://tw
18	h-5 kuliah yaampun welkam back to my hectic lyf<U+…	FALSE	0	NA	2019-08-28 13:22:28	FALSE	NA	1166702591520075776	NA	<a href="http://tw
19	KULIAH IN LESS THAN A WEEK<U+0001F62D>	FALSE	0	NA	2019-08-28 13:18:46	FALSE	NA	1166701657922535424	NA	<a href="http://tw
20	@radenrauf day6: congratulation, first time, colors s…	FALSE	0	radenrauf	2019-08-28 13:14:43	FALSE	1166317680602669057	1166700639226740736	26970305	<a href="http://tw
21	dapet kuliah sound gitar pedal digital vs analog, tube…	FALSE	0	NA	2019-08-28 13:01:55	TRUE	NA	1166697419654029312	NA	<a href="http://ins
22	no offense but kuliah fig di luar bali/negeri and getti…	FALSE	2	NA	2019-08-28 12:54:08	FALSE	NA	1166695458703470592	NA	<a href="http://tw
23	RT @bucinhayunseong: BARU GUE TINGGAL KULIAH …	FALSE	0	NA	2019-08-28 12:22:47	FALSE	NA	1166687569423167488	NA	<a href="http://tw
24	tadi kuliah matkul prodi pertama and it turns out fine, …	FALSE	0	NA	2019-08-28 12:11:21	FALSE	NA	1166684691564355584	NA	<a href="http://tw

Fig. 1. Result of Crawling data from Twitter Big Data

In Fig.1, it showed that the text column contains the opinion of tweeter users. In favorite column, we can get the information that they are not prefer to discuss about university. Whereas, in truncated column, which means, after we implemented the Naïve Bayes approach, we get there

are true-false sentiments which represents their negative and positive opinions about studying further to the university.

4.2. Data Cleaning

The 'text' field contains the tweet, hashtag, and URL sections. We need to remove the hashtag and URL from the text field so we only have the main tweet section to run our sentiment analysis. It contains many URLs, hashtags, and other Twitter handles. We will delete all of this using the gsub function. Our current text fields look like below:

Table 1. Data that has been deleted hashtag and URL

No	Tweet data
1	RT
2	Untung ntar mo kuliah onlen. gaperlu dah tuh ketemu org2 yg permasalahin outfit cuz u wont see me bitch. i be lying…
3	kenapa mau jualan? Ga ribet berbisnis sambil kuliah? \n<U+0001F469> : bcs i enjoyed it, i like to share something that some…
4	nsenin masuk kuliah auto makan nasi + garem
5	whenever ada kuliah tamu, i feel stupid lol

4.3. Scoring of Sentiment of Each Tweet

This research needs to try to get an emotional score for each tweet. By breaking down emotions into 10 different emotions - Anger, anticipation, Lov, Fear, Joy, sadness, surprise, trust, negative and positive.

Table 2. Sentiment score for each tweet

Anger	Anticipation	Lov	Fear	Joy	Sadness	Surprise	Trust	Negative	Positive
0	0	0	0	0	0	0	0	0	0
2	2	1	2	1	0	1	0	0	2
3	0	1	0	0	0	1	0	0	1
4	0	0	0	0	0	0	0	0	0
5	0	1	0	0	0	0	1	1	0
6	0	0	0	0	0	0	0	0	1

The output above shows the various emotions that exist in each tweet.
Now, we will use the get_sentiment function to extract the sentiment score for each tweet.

Table 3. Grouping based on positive tweets

No	Tweet data
1	kenapa mau jualan? Ga ribet berbisnis sambil kuliah? \n\U0001f469 : bcs i enjoyed it, i like to share something that some… "
2	nsenin masuk kuliah auto makan nasi + garem"
3	"gue lulus kuliah mau balik indo and make a 24 hour boba place not because i think it'll make me a lot of money but bc I Want That"
4	"At certain point, I should agree that kuliah is a comfort zone."
5	"i know this won't change anything but i just hope they come to indo on feb 2020 so i can watch it klo sept tuh kuli… "
6	"Kuliah di Seoul Absence of\nComprehensive Art School."

Table 4. Grouping based on negative tweets

No	Tweet data
1	"bsk masuk kuliah and guess what me got a massive diarrhea AGAIN fuck "
2	"kuliah sore kills me slowly:("
3	"Run .... Never back !!! You are fucking COWARD !!\n "

The next step is to make a classification from the score of each tweet. In total, 1978 tweets were evaluated, so there must be 1978 positive, negative or neutral scores, one for each tweet. so, we get a total number of tweets scores for the interest of high school graduates to continue their studies, the classification is depicted as a histogram of tweets containing three histogram bars of negative, neutral and positive sentiment in the is as follows:

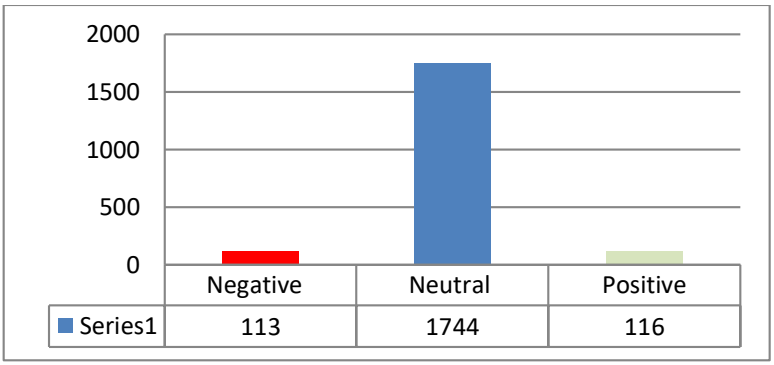

Fig. 2. Tweet Sentiment result

After classification phase, we get 113 negative text processing stage, the data will enter the stage of classification to find out of the *1978* tweets data into the category of positive or negative sentiment. The classification process for the tweets data are grouped using the Naïve Bayes with the sample is given in Table 5.

Table 5. The classification phase of 100 tweet data

[1] 0.00 -1.50 2.75 0.00 1.10 -1.00 -0.40 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[16] 0.00 0.00 0.60 0.00 2.05 0.00 0.50 0.00 0.00 0.00 0.00 0.00 0.00 0.85 0.00
[31] 1.40 0.50 0.00 0.00 0.00 0.00 0.00 0.00 1.60 0.00 0.00 0.40 0.00 -0.50 0.00
[46] 0.00 0.00 -1.00 0.00 -1.30 0.00 0.50 0.00 0.00 0.50 -0.25 -0.65 0.50 0.00 1.60
[61] 0.00 0.00 0.75 0.00 0.50 -1.50 0.00 -0.60 -0.25 0.00 0.00 0.80 0.00 0.00 0.00
[76] 0.00 -1.00 -1.75 0.00 0.00 0.00 0.00 -1.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[91] 0.00 0.00 0.00 0.00 0.00 1.75 0.00 0.00 0.00 0.00 0.00 0.80 0.00 0.00 0.00
[106] 1.75 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.80 0.00 -1.00 0.00 0.00 0.00 0.00
[121] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[136] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 -0.20 0.00
[151] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[166] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[181] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[196] 0.00 0.00 0.60 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00

Table 6. The classification into negative, positive and neutral sentiment

[1] "Neutral" "Negative" "Positive" "Neutral" "Positive" "Negative" "Negative" "Neutral"
[9] "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral"
[17] "Neutral" "Positive" "Neutral" "Positive" "Neutral" "Positive" "Neutral" "Neutral"
[25] "Neutral" "Neutral" "Neutral" "Neutral" "Positive" "Neutral" "Positive" "Positive"
[33] "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Positive" "Neutral"
[41] "Neutral" "Positive" "Neutral" "Negative" "Neutral" "Neutral" "Neutral" "Negative"
[49] "Neutral" "Negative" "Neutral" "Positive" "Neutral" "Neutral" "Positive" "Negative"
[57] "Negative" "Positive" "Neutral" "Positive" "Neutral" "Neutral" "Positive" "Neutral"
[65] "Positive" "Negative" "Neutral" "Negative" "Negative" "Neutral" "Neutral" "Positive"
[73] "Neutral" "Neutral" "Neutral" "Neutral" "Negative" "Negative" "Neutral" "Neutral"
[81] "Neutral" "Neutral" "Negative" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral"
[89] "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Positive"
[97] "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Positive" "Neutral" "Neutral"
[105] "Neutral" "Positive" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral"
[113] "Neutral" "Positive" "Neutral" "Negative" "Neutral" "Neutral" "Neutral" "Neutral"
[121] "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral"
[129] "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral"
[137] "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral" "Neutral"

From the results of the classification will proceed to the stage of accuracy calculation. The accuracy phase is useful for knowing the performance of the Naïve Bayes classifier algorithm in classifying text data whether it gets high accuracy or even low accuracy. The following picture shows the accuracy of the Naïve Bayes classifier algorithm:

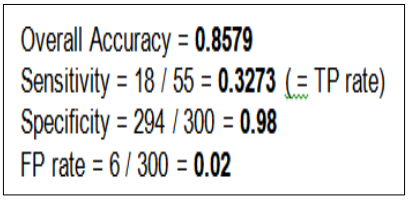

Fig. 3. Result of accuracy testing (source: RStudio)

From this picture we got 85% accuracy test results. It can be concluded that the Naïve Bayes algorithm is quite successful in predicting the correct sentiment category because the accuracy of the naïve Bayes algorithm results gets a high accuracy of 85% which means that the performance of the Naïve Bayes algorithm can classify text data very well.

5.CONCLUSION

Based on the results of the analysis and testing that has been done on, the following conclusions can be drawn:

a. Naïve Bayes algorithm is very effective to be used as a tweet classification process needed in this sentiment analysis system where the value obtained in testing is up to 85%.
b. Naïve Bayes method can be used to classify tweets quite well on sentiment analysis systems.

6. SUGGESTIONS

The author suggests developing further research into the Tweet classification system as follows:

a. Researchers suggest for further research to increase the amount of training data and test data to get better results when classifying tweets.
b. The researcher suggests that for further research the language used is not only Indonesian but can use regional or foreign languages such as English and other languages.
c. Researchers suggest for further research at the text processing stage plus the stemming feature to get better results
d. The researcher recommends that future studies not only use the Naïve Bayes algorithm but also use the support vector machine algorithm and other text classification algorithms.

REFERENCES

[1] Hendayana, S., Asep, S., & Imansyah, H. (2010). Indonesia's issues and challenges on quality improvement of mathematics and science education. Journal of International Cooperation in Education, 4(2), 41-51.

[2] McNeil, K., Brna, P. M., & Gordon, K. E. (2012). Epilepsy in the Twitter era: a need to re-tweet the way we think about seizures. Epilepsy & behavior, 23(2), 127-130.

[3] Martínez-Cámara, E., Martín-Valdivia, M. T., Urena-López, L. A., & Montejo-Ráez, A. R. (2014). Sentiment analysis in Twitter. Natural Language Engineering, 20(1), 1-28.

[4] Wald, R., Khoshgoftaar, T. M., Napolitano, A., & Sumner, C. (2012, December). Using Twitter content to predict psychopathy. In 2012 11th International Conference on Machine Learning and Applications (Vol. 2, pp. 394-401). IEEE.

[5] Manjusha, K. K., Sankaranarayanan, K., & Seena, P. (2014). Prediction of different dermatological conditions using naive bayesian classification. International Journal of Advanced Research in Computer Science and Software Engineering, 4(1).

[6] Ting, S. L., Ip, W. H., & Tsang, A. H. (2011). Is Naive Bayes a good classifier for document classification. International Journal of Software Engineering and Its Applications, 5(3), 37-46.

[7] Alsaffar, A., & Omar, N. (2014, November). Study on feature selection and machine learning algorithms for Malay sentiment classification. In Proceedings of the 6th International Conference on Information Technology and Multimedia (pp. 270-275). IEEE.

[8] Peling, I. B. A., Arnawan, I. N., Arthawan, I. P. A., & Janardana, I. G. N. (2017). Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm. International Journal of Engineering and Emerging Technology, 2(1), 53-57.

[9] Murphy, K. P. (2006). Naive bayes classifiers. University of British Columbia, 18, 60.