Journal of Computer and Information Systems Ampera
Vol. 5, No. 1, January 2023 e-ISSN: **2775-2496**

https://journal-computing.org/index.php/journal-cisa/index

Analysis of YouTube Videos as Educational Media Using the Naive Bayes Method

Phung Mulan Visaka Rani[1], Phoebe Cecilia Anggreiny[1], Yusianne Kasih Husada[1], Grasella Aldonia Pangandaheng[1], Adelia Vannissa Arwi[1], Viny Christanti Mawardi[2]

[1]Psychology Departement, Tarumanagara University, DKI Jakarta, Indonesia
[2]Information Engineering Departement, Tarumanagara University, DKI Jakarta, Indonesia
Email: [2]viny@untar.ac.id

Abstract

Social media has become a way for people to socialize without meeting face to face, seek knowledge, sell, and so on. Analysis of the video "How Old is the Universe? (And How Do You Know?" aims to find out the public's response to educational videos on YouTube social media. With this research, it is hoped that social media can develop as an educational medium. First, 611 video comments were collected but only 100 comments could be analyzed. Then Pre-processing was carried out with the following steps: parsing, tokenization, stopword removal, and stemming. After that, comments were divided into 4 clusters based on the occurrence of the same words and their synonyms. Comments were also classified using 3 scales, namely negative, neutral, and positive. The initial 50 comments were manually classified while the other 50 comments used Rapidminer with the Naive Bayes method. Analysis of similar videos as comparative data also used the same steps. The results of the analysis showed a positive response of 65% with advantages in the use of animation and accurate scientific presentation. Based on the analysis, it can be concluded that YouTube is effectively used as an educational medium.

Keywords: *youtube, social media analysis, sentiment analysis, education, social media*

1. INTRODUCTION

Nowadays we can no longer be separated from social media, social media now plays a very important role in life, originally social media was only used to communicate and interact between users. Currently it includes the dissemination of information, knowledge, social activities, distribution of invitations, and places of work. In this modern era, it is impossible for someone not to have social media, and it has become something that is

embedded in people's daily lives. In the beginning communication could only be done in one direction, now with the development of social media you can see what is happening in this world. Social media is also used by many people to connect with each other using the applications that are provided, such as Instagram, Facebook, Twitter, and many more applications that are used [1].

This makes people who are far away feel close with the help of this technology. As explained by Nasrullah [2] social media is a medium on the internet that allows users to represent themselves and work together, share information, and communicate with other users to form virtual social bonds. From this understanding, we can see that social media is a way for people to socialize without having to meet face to face. Social media has given all people have the same opportunity to find information. Social media has also been able to realize human collaboration without time and place limitations, thus making social media a communication tool for the current generation.

Social media, which has developed rapidly, has both positive and negative impacts [3]. One of the positive impacts of social media is that it has become an educational medium. During this pandemic, the government's call for distance learning [4] has resulted in many educational videos being uploaded to the YouTube platform as a way to learn without meeting face to face. YouTube is used as an educational medium because the media is in the form of videos that can be uploaded in long duration so that the explanation will not be cut off, video editing can be done to make the video content more interesting, it can explain learning through audio and visuals, as well as features on YouTube itself that make it easier to use. Before the pandemic, there were already educational videos spread across various social media, but the effectiveness of social media as an educational medium was not yet known. Therefore, how effective social media, especially YouTube, is as an educational medium, will be explained further in this research [5] [6].

1. How did the public respond to the video “How Old is the Universe? (And How Do You Know?)”?
2. How is YouTube social media suitable as an educational medium?
3. Can social media be a good educational platform?

This research aims to determine the public's response to the use of YouTube videos and summarize the benefits of YouTube social media as a

means of education and also how effective the use of social media, especially YouTube, is as a means of education.

2. METHODS

First, researchers conducted a literature review to find a basic understanding of social media analytics, Naive Bayes methods, and social media. From literature, researchers found that Social media analysis is the science and art of extracting valuable insights from the abundance of unstructured and semi-structured social media data in order to be informed and make informed decisions [7]. There are three levels of social media analysis, namely media analysis (reach, relationships, and virality), conversation analysis (share of voice, sentiment analysis, and ethnographic analysis), and network analysis (influencer identification and network dynamics). Social media analysis is carried out with the following steps: collecting data, looking for meaning, and measuring. In this research, the analysis conducted will be conversation analysis which is sentiment analysis because researcher will conduct the research based on the comment on some social media. Sentiment analysis is a computational study of mass opinion, which can represent opinions on a particular topic or event expressed in a text [8]. Sentiment Analysis is categorized into two, namely: Coarse (positive or negative sentiment) and fine (2 or more sentiment) grained sentiment analysis. Using a simple probabilistic classifier or Naive Bayes method that calculates a set of probabilities by adding the frequencies and combinations of values from a given data set to predict future opportunities based on past experiences. The advantage of using Naive Bayes is that this method only requires a small set of training data to determine parameter estimates needed in the classification process. Naive Bayes often performs much better than expected in the most complex real-world situations [9].

According to Andreas and Haenlein [10] social media is a platform that enables people to build social networks and share information and opinions. The functions of social media include the following: a place to get knowledge and information, a medium of communication and popularity, and provides an opportunity for entrepreneurs to get closer to their consumers. One of the social media in question is youtube. According to Sugama and Dewi [11]; Emiliana and Sari [6] YouTube is a social media platform which contains videos uploaded by account owners with the aim of sharing. As well as account owners can view videos shared by other

people. Based on DataIndonesia.id [12] YouTube users in Indonesia reached 127 million users. There are several structures of YouTube itself, namely: videos uploaded by account owners, a description that contains the text that the account owner wants to convey, and the comments page is part of the discussion for viewers or account owners.

Next, researchers choose social media, topics, and two comparative accounts that will be researched. The procedure of the research will come as follows: collecting the comments with Coberry, then using the sentiment analysis steps such as pre-processing comments which will filtering the words, the filtered comments will be classified to three cluster namely: positive, negative, and neutral. The clustering will be done to 50 comments only that later will be named training data. The training data is imported to the RapidMiner Studio. Then, the other 50 will also be imported after setting up the training data on RapidMiner Studio. This step will make the RapidMiner Studio automatically predict the rest of comments classification with Naive Bayes Method. Lastly, an analysis is carried out based on the data results from both accounts.

3. RESULTS AND DISCUSSION

In this research, the topic chosen was a video on YouTube entitled "How Old is the Universe? (And how do you know?)" This educational content was uploaded by an account called *Kok Bisa?*. See Figure 1.

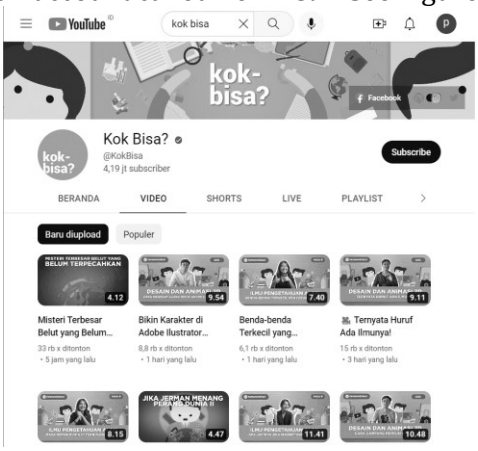

Figure 1. Account *Kok Bisa?*

After determining the social media and topic to be analyzed, then collect comments on the video using the Coberry application by entering the

YouTube link in the column provided as in Figure 2. Then the Coberry application will collect comments on the video "How Old is the Universe? (And how do you know?)". Of the 611 comments, a selection was carried out to get comments that were suitable for analysis, the criteria were that they had words and were not spam, then from the selection we got 100 comments.

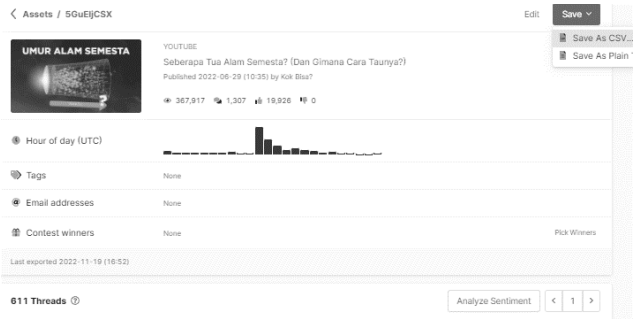

Figure 2. Coberry operation

The next step, namely pre-processing comments. In the excel table there are four columns, in the parsing column each comment is entered in one row, in the next column tokenization is carried out, namely the punctuation marks in comments are removed, abbreviated words are lengthened, letters are changed to lowercase and words are standardized. Then in the next column a filtering process is carried out, namely removing unimportant and meaningless words (stopword removal). In the last column, stemming is carried out or returning the words to their base words. An example of work on one of the comments as shown in Figure 3.

Isi Dokumen	Tokenisasi	Filtering	Stemming
Indahnya sains. Lebih indah lgi cara kerja di baliknya. Memang otak manusia adalah keajaiban dari Tuhan yang keajaiban nyata	indahnya sains lebih indah lagi cara kerja di baliknya memang otak manusia adalah keajaiban dari tuhan yang keajaiban nyata	indahnya sains lebih indah lagi cara kerja baliknya memang otak manusia adalah keajaiban tuhan keajaiban nyata	indah sains lebih indah lagi cara kerja balik memang otak manusia adalah ajain tuhan ajaib nyata

Figure 3. Example of pre-processing

After pre-processing, comments are obtained that are clean and ready to be processed. The next process is clustering, namely making comments

into several groups based on the appearance of similar or synonymous words. In this research, four clusters were obtained with each cluster having different similarities, such as in cluster 1 consisting of the appearance of the word 'cool' and its synonyms such as 'steady', 'amazed', and the like. In cluster 2 based on the appearance of the word 'understand' and its synonyms 'grasp', 'understood', and so on. cluster 3 based on the occurrence of the words 'god' and 'allah'. Finally, cluster four is based on the appearance of the word 'terrible', and its synonym 'scary'. Then the comments are classified into three columns, namely negative, neutral, and positive. In the first 50 comments, comments were classified manually with the aim of creating training data. The other 50 comments were classified using the RapidMiner Studio application with the Naive Bayes method. In manual classification, comments are classified based on sentences that show meanings according to the predetermined classification so that the negative category is given to cluster four because it has words with negative meanings, the neutral category is given to comments in cluster three because it has negative words. have a negative or positive meaning, and the positive category is given to clusters one and two because they contain words with a positive meaning.

Table 1. Example Sentences in Each Class

Sentiment	Class
Very amazed to know the age of the earth	Positive
there are those who believe it	Negative
The more you know how great nature is, the more you know how big God is	Neutral

In Table 1, each comment, which is called a sentiment, has a class. These classes consist of positive, negative, and neutral. These classes are the result of classifying existing sentiments through a clustering process.

The manual classification that has been carried out will become training data that will be used when operating RapidMiner Studio. The first thing to do in operating RapidMiner Studio is to use manual classification data as training data for the application. The data is imported into the application, then using the example filter operator enters the data that already has a label, after that uses nominal to text so that the data can be read as text. Document processes are carried out such as pre-processing, namely tokenization, changing words to lowercase, filtering stop words

using previously created stopword data, and filtering tokens based on word length. After processing the document, Naive Bayes automatically calculates the probability based on the Naive Bayes algorithm. The data that has been used as training data is then saved and then used to predict the classification of 50 other comments.

Test data operations are carried out in the following stages: importing excel data into application, use filter to data that does not yet have a label, use nominal to text, process documents like training data steps, then combine the training and test data with union operation, use the replace missing values operation so that the value does not have a question mark (?) written on it, finally, apply the model with input from operations and test data. An example of the results of RapidMiner Studio operations can be seen in Figure 4.

Row No.	kategori	prediction(...
1	?	Netral
2	?	netral
3	?	netral
4	?	Positif
5	?	Positif
6	?	netral
7	?	positif
8	?	positif
9	?	netral
10	?	positif
11	?	netral
12	?	positif
13	?	positif
14	?	positif
15	?	Positif
16	?	positif
17	?	positif
18	?	positif
19	?	positif
20	?	positif
21	?	positif
22	?	Netral
23	?	positif
24	?	netral
25	?	netral
26	positif	positif
27	positif	positif
28	positif	positif
29	positif	positif
30	positif	positif
31	positif	positif
32	positif	positif
33	positif	positif
34	positif	positif
35	positif	positif
36	positif	positif
37	positif	positif
38	positif	positif
39	positif	positif
40	positif	positif
41	positif	positif
42	netral	netral
43	Negatif	Negatif
44	positif	positif
45	netral	netral
46	Netral	Netral
47	Netral	Netral
48	Netral	Netral
49	Positif	Positif
50	Positif	Positif

Figure 4. Example of RapidMiner Studio operation results

Vol. 5, No. 1, January 2023 e-ISSN: 2775-2496
https://journal-computing.org/index.php/journal-cisa/index

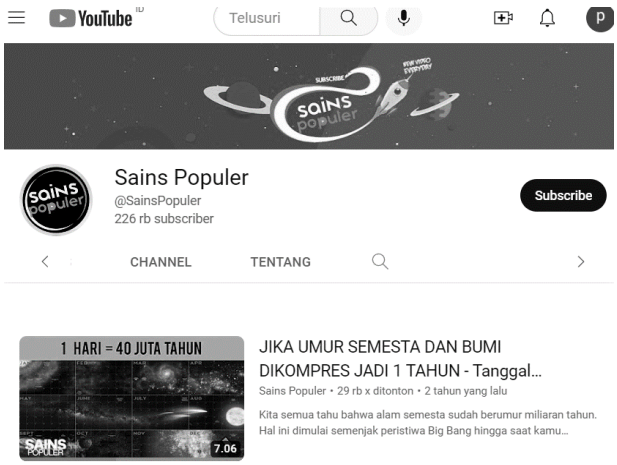

Figure 5. Comparative accounts of Sains Populer

After completing a series of steps on one video, then carry out the same steps on another video with the same topic but published by a different account which will later be used as comparative data in carrying out the analysis. After both videos have data that has been extracted, analysis is carried out by comparing the number of positive, neutral, and negative category classifications from the two videos which then produces decisions, policies and strategies.

Based on sentiment analysis through classification carried out as in table 1. of 50 sentiment responses it was found that 70% or a total of 35 comments were positive, which means that viewers liked the content "How Old is the Universe?" published by the account *Kok Bisa?* And 26% or 13 comments were neutral.

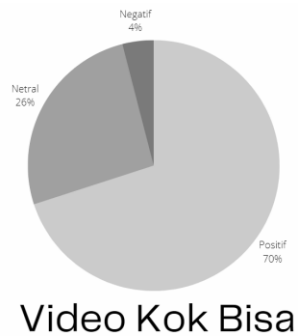

Figure 6. Visualization of account percentages of *Kok Bisa?* accounts

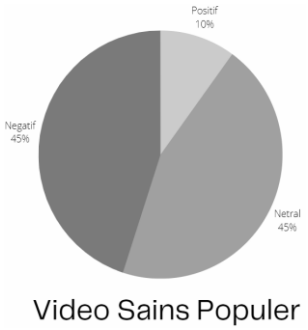

Figure 7. Visualization of account percentages of *Sains Populer* Accounts

To get a comparative value, the same content was taken from a different account, namely *Sains Populer* which received 10% positive comments, 45% negative comments, and the other 45% neutral, which means there were only 2 positive comments while there were 9 negative comments and 9 neutral comments. Based on the data obtained, it can be found that the presentation was carried out by the account *Kok Bisa?* much better than its competitors. There are several factors that this can happen, namely:

1. The use of animation, the use of animation really supports the ease of explaining knowledge. This can be proven by the many comments praising the animation created by the account *Kok Bisa?*
2. Use of research results, with this account *Kok Bisa?* Can prove that the knowledge provided is an accurate fact and has been explained by experts.

From the factors found, the policies and strategies carried out by the account. *Kok Bisa?* It is appropriate to see from the account's consistency in using animation to explain information. This account also provides facts that support the explanation of the information, so that people can believe it. In its function as an educational video account *Kok Bisa?* He has been good at delivering informative videos so that viewers are interested in learning more. However, everything has a negative side where there are some viewers who do not understand or despise the science being explained. From seeing the sentiment and enthusiasm of viewers who ask for other educational videos, it can be concluded that YouTube social media is very effective as an educational medium because it can provide education to the general public wherever and whenever and with easy-to-understand explanations so that viewers can easily understand.

4. CONCLUSION

Social media such as YouTube has advantages and disadvantages in its function as an educational medium. Interesting and informative videos can easily provide new information for viewers and develop viewers interest in learning, but in educational videos there are still viewers who respond negatively and don't like the information conveyed in the video. Social media as an educational medium is quite effective if the content presented is interesting and easy for the audience to understand. This analysis can clarify people's interest in the content presented and can find out what people want. So, comment analysis needs to be carried out by content creators.

Before analyzing, selection needs to be done so that the data obtained can be more accurate. Although sometimes the data obtained does not match expectations because the public response cannot be controlled. Suggestions for further research are to look for videos that are newer and have more comments and increase the sample, so that it is more representative of society.

REFERENCES

[1] S. Ahmad, "Pemanfaatan media sosial untuk efektifitas komunikasi." *Cakrawala: Jurnal Humaniora Bina Sarana Informatika,* vol. 16, no.2, 2016.

[2] N. Rulli, "Communication, Culture and Social Media Perspectives." *Sociotechnology: Rekatama Media Symbiosis,* 2015.

[3] A. Waseem, and R. Kumar, "A study on positive and negative effects of social media on society." *International journal of computer sciences and engineering,* vol. 5, no.10, pp. 351-354, 2017.

[4] Kementerian Pendidikan dan Kebudayaan. "Kemendikbud terbitkan pedoman penyelenggaraan belajar dari rumah*".* kemdikbud.go.id. Accessed: Nov 23, 2022. [Online]. Available: https://www.kemdikbud.go.id

[5] M. Sugama, and P. S. Dewi, "Practicality and effectiveness of geogebra-assisted online calculus teaching materials." *Scholar's Journal: Journal of Mathematics Education,* vol. 4, no. 2, pp. 888-899, 2020.

Vol. 5, No. 1, January 2023 e-ISSN: **2775-2496**

https://journal-computing.org/index.php/journal-cisa/index

[6] P. Emiliana, and F. M. Sari, "Indonesian EFL student's perspectives towards learning management system software." *Journal of English Language Teaching and Learning* vol. 1, no. 1, pp. 20-24, 2020.

[7] S. Marshall, and G. F. Khan*. Digital analytics for marketing*. Routledge, 2018.

[8] T. Mikalai, and T. Palpanas, "Survey on mining subjective data on the web." *Data Mining and Knowledge Discovery,* vol. 24, no. 3, pp. 478-514, 2012.

[9] A. Saleh, "Implementation of the Naive Bayes Classification Method in Predicting the Amount of Household Electricity Use". *Creative Information Technology Journal,* vol. 2, no. 3, pp. 207-217, 2015.

[10] K. Andreas M., and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media." *Business horizons,* vol. 53, no. 1, pp. 59-68, 2010.

[11] M. Sugama, and P. S. Dewi, "Practicality and effectiveness of geogebra-assisted online calculus teaching materials." *Scholar's Journal: Journal of Mathematics Education,* vol. 4, no. 2, pp. 888-899, 2020.

[12] M. Ali, "Third Largest Indonesian Youtube Users in the World in 2022." DataIndonesia. id. Accessed: Nov 23, 2022. [Online]. Available: https://dataindonesia.id