Implementation of AHP TOPSIS and Sentiment Analysis for Identifying COVID-19 News That Have A Positive Impact On The Community

[1]Benny Richardson, [2]Kevin Hendy, [3]Vanessa Ardelia, [4]Ventryshia Andiyani, [5]Wilson Philips

***Abstract* — Currently, the number of COVID-19 cases is increasing with a worldwide number of 45 million by October 2020. Along with the increasing number of COVID-19 cases, much news on the subject has increasingly dominated reporting in Indonesia. Therefore, it cannot be denied that the coverage of COVID-19 in Indonesia affects people's behavior, which is evident from the mood and opinion in the comments column. Some messages also seem to have a negative impact on people's behavior. Information management is one of the efforts that can be made to filter news so that the news displayed to the public is news that has a positive impact on society. This study aims to assess the news related to COVID-19, which has the most positive impact on society with the AHP TOPSIS method and Lexicon-Based sentiment analysis, using a data set obtained from uploading COVID-19 news on the Facebook account of KementrianK KesehatanRI was derived. The results of this study show that the AHP-TOPSIS method is suitable as a method for determining the ranking of COVID-19 news, which is very effective and positive for news that has a negative impact. The accuracy obtained in this study reaches 80%.**

***Keywords* — AHP TOPSIS, COVID-19, Facebook, Lexicon-Based, Semantic Analysis.**

I. Introduction

Currently, the number of COVID-19 cases is increasing with the number worldwide reaching 45 million by October 2020. Meanwhile, the number of cases in Indonesia at that time reached more than 406 thousand. As the number of COVID-19 cases increases, a lot of news about this issue increasingly dominates news coverage in Indonesia throughout February - October 2020. Therefore, it cannot be denied that COVID-19 reporting in Indonesia affects people's behavior which can be seen from sentiment or opinion. given in the news comment column. One of the responses to high COVID-19 news is in the news uploaded by the Ministry of Health's account on Facebook.

Based on these problems, it requires proper information management related to COVID-19 news [1] by the mass media and the government, in order to spread news that has a positive impact on society. Proper information management can be done by reviewing the news that has been disseminated. This study aims to rank the news related to COVID-19 that has the most positive impact on society using the AHP TOPSIS method and lexicon-based sentiment analysis. In related research [3], a comparative analysis of the MCDM AHP, TOPSIS and AHP TOPSIS methods was carried out. The results show that the AHP TOPSIS method provides the best accuracy with a Hamming distance value of 96.02%. This shows that the AHP TOPSIS method is better than AHP and TOPSIS. Furthermore, to measure public sentiment, in related research [2] a lexicon-based analysis was carried out by taking the weight of the word sentiment provided by SentiWordNet which is integrated with the MOORA method to determine the best smartphone sequence. Therefore, in this research, the AHP TOPSIS method is applied to perform alternative weighting and ranking and lexicon-based sentiment analysis.

The stages of the research carried out included retrieving news data from Facebook, preprocessing data, analyzing comment data sentiment, making decision matrices, sorting alternatives with AHP TOPSIS, and calculating accuracy. The alternative that you want to rank is the last 10 news on the upload of COVID-19 news by the Ministry of Health's account on Facebook. The AHP TOPSIS criteria used are the number of likes, comments, shares, average sentiment value, and news duration. To get the average sentiment score, a lexicon-based sentiment analysis was carried out on the top 10 comments of each news post.

The program implementation is made using the Python programming language. News data was collected using web scraping techniques using the Facebook-Scraper library and the FacePager application. The sentiment value of the comment data is calculated using a predefined word sentiment weight dataset. The result data from the preprocessing are used to fill in the value in the decision matrix. The values in the decision matrix are then normalized by min-max. The decision matrix is then processed by AHP TOPSIS to sort alternatives. Then, the accuracy of the sequence results obtained is calculated by comparing the order given by the expert. The results of this study indicate that the AHP TOPSIS method and the Lexicon-based sentiment analysis are appropriate for information management with the accuracy value obtained in this study reaching 80%.

II. RESEARCH METHOD

The research carried out can be broken down into several stages, namely:

1. Retrieval of news data from Facebook
2. Preprocessing news data
3. Conduct sentiment analysis on comment data
4. Prepare a decision matrix
5. Sorting alternatives with AHP TOPSIS
6. Calculating accuracy

The process flow in this research can be seen in Figure 2.1.

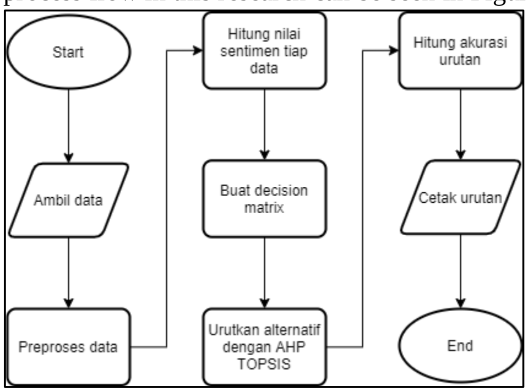

Figure 2.1 Flowchart of The Research

A. Data Retrieval

News data is collected using web scrapping and the FacePager application to retrieve data on Facebook pages. The Facebook page visited was the home page belonging to the Ministry of Health of the Republic of Indonesia (Kemenkes), namely on facebook.com/KementerianKesehatanRI. The Ministry of Health is a government agency in charge of controlling COVID-19 cases in Indonesia. The data taken includes the 10 most recent news uploads by the Ministry of Health, consisting of news headlines, number of likes, comments, shares, top 10 comments, and date of upload.

B. Preprocessing Data

Preprocessing Data is done to transform data into a decision matrix. The first step is data cleaning, which includes deleting irrelevant data, duplicate data, and missing data. Then, the data preprocessing is continued to process the comment data, each of which is in the form of a sentence into a collection of tokens. The stages for processing comment data are as follows:

1. Case folding

   Is a process for converting text into standard forms. Case folding operations include deleting characters other than letters, changing text to lowercase, deleting punctuation marks and whitespace.
2. Filtering

   Filtering is a process to get rid of unnecessary words or what is called a stopword. Stopwords are considered insignificant because they have no meaning that can affect sentiment values, and generally these words appear in large numbers.
3. Tokenizing

   Is the process of breaking a sentence into a collection of tokens that appear in that sentence. Words that appear more than once will not be included in the token list.

C. Sentiment Analysis

The sentiment analysis approach used is lexicon-based. With a lexicon-based approach, the calculation of the value or sentiment score is obtained by adding up all the word sentiment weights from the token collection of a sentence, based on the lexicon dataset that has been defined previously in related research [4].

$$Sentiment\ Score = \sum Weighted\ Sentiment\ Word \quad (1)$$

To get the overall sentiment score for comments on a news, the average value of all sentiment scores is used. The formulation in finding the average sentiment can be described as follows

$$Sentiment\ Average = \frac{\sum Sentiment\ Score}{Number\ of\ Comments} \quad (2)$$

Because the lexicon dataset used defines words in the form of inflections (words and their affixes) as well as in addition to the root word, as well as word combinations (consisting of several word combinations), a strategy is needed to take the weight of word sentiments. The strategy applied is to take the weight of the word sentiment starting from the inflected form. If not found in the lexicon dataset, the token is converted into the root word through a stemming process, which then searches for its weight again. If still not found, then the inflected token is combined with the previous one to two words.

D. Prepare a Decision Matrix

A decision matrix is needed before processing alternatives with AHP TOPSIS. The matrix contains the alternatives you want to sort along with the values for each criterion. From the results of the literature study, the criteria needed in ranking news are obtained, including:

1. Number of likes (C1, benefit)
2. Number of comments (C2, benefit)
3. Number of shares (C3, benefit)
4. Average comment sentiment (C4, benefits)
5. The duration of the news or the length of time the news is uploaded (C5, cost)

The value on each criterion will be normalized by the min-max method (between 1-5) so that the value has clear boundaries.

E. AHP TOPSIS

The MCDM method used is AHP TOPSIS to sort alternatives based on the criteria previously mentioned. AHP TOPSIS was chosen because it has been proven to provide better accuracy, when compared to the AHP and TOPSIS methods. The criteria importance matrix used is as follows

	C1	C2	C3	C4	C5
C1	1	1/5	3	1/7	3
C2	5	1	5	1/3	3
C3	1/3	1/5	1	1/7	1/3
C4	7	3	7	1	7
C5	1/3	1/3	3	1/7	1

Figure 2.2 Criteria Importance Matrix

F. Calculate Accuracy

The result of the alternative ordering obtained will be measured for its level of accuracy with the following equation, referring to [5]

$$Accuracy = \frac{Number\ of\ Exact\ Sequence}{Amount\ of\ Data} * 100\% \qquad (3)$$

III. Results and Analysis

The representation of the implementation results of each designed work step can be described as follows

A. Scraping Data from Facebook

The data set to be used in this study comes from the Ministry of Health's facebook. The data set will consist of the last 10 posts dated 29 November 2020. The data collection process is carried out using 2 web scrapping techniques. namely coding through and through the FacePager application.

The coding technique is carried out using the python programming language and utilizing the get_posts class from the facebook_scraper library. This library functions to retrieve data from the Facebook public page without requiring an API key, so that in use it only requires sending a unique page name as the first parameter. As for the use of other parameters, namely the page to determine the number of frame pages, where usually the first page has 2 posts and the remaining 4 posts. In this case, the unique page name used is 'KementerianK KesehatanRI' and the page parameter is 3 to get a total of 10 posts. Scrape through this coding technique resulted in 10 posts consisting of the post_id attribute as the id of a post, text as the content of the post, time as posting time, likes as the number of likes of a post, comments as the number of comments of a post, share as the number of shares of a post, etc. which will be saved in the csv file.

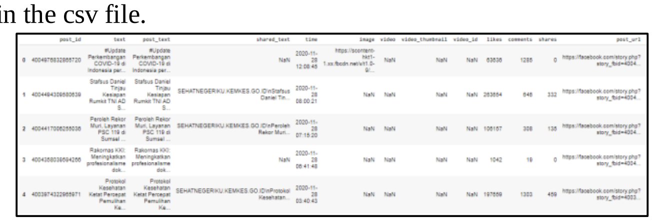

Figure 3.1 Hasil pengambilan data dengan library facebook_scraper

The second technique is through an application called a facepager. This application is used to get the content of comments from each post. The appearance of the facepager application can be seen in Figure 2. The output of this application will be stored in a csv file.

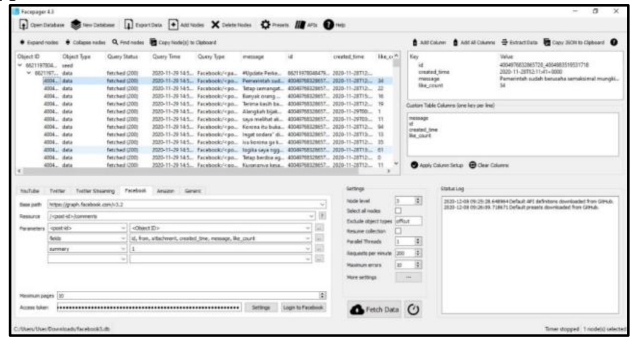

Figure 3.1 Tampilan dan hasil pengambilan data dengan Facepager

B. Preprocessing Data

After the data is obtained by scraping data from Facebook, the next process is to preprocess the data. The preprocessing process is carried out using the NLTK and Literature python libraries. First, a new dataframe is created that is used to store preprocessed data. The process begins with case folding or one form of text preprocessing to convert all letters in the document to lowercase. Only letters 'a' through 'z' will be accepted. There are several ways to do case folding, including changing comments to lowercase, deleting numbers, deleting punctuation marks, and deleting whitespace or empty characters. In converting comments to lowercase, I make use of the built-in method for handling strings to lowercase. Next, the authors erase the numbers in the comments using regular expressions (regex). In removing punctuation marks, the author makes use of maketrans to remove punctuation in comments. To remove whitespace, I use python's strip () function.

After doing case folding, the authors filter using literature. However, after filtering it turns out that there are still many slang words that are not filtered out. Therefore, the author adds new words combined with the literary dictionary to filter sentences. The writer got these words from the open dataset of other researchers on the github account. Then it is filtered again using the combined dataset.

After filtering, the author tokenizes the comments. Tokenizing is the process of separating text into pieces known as tokens for later analysis. Tokenizing is done using the nltk library.

C. Sentiment Analysis

The number of tokens obtained is 1.131, with the highest frequency of tokens being 'semoga', 'banyak', 'semua', 'covid', 'pandemi', 'allah', 'cepat', and 'berakhir'. These words describe the response of the people who want the COVID-19 pandemic to end quickly. The token collection is then processed by stemming, and the availability is checked on the lexicon dataset. The lexicon dataset has 10.248 words, with a range of sentiment values ranging from +5 (most positive sentiment) to -5 (most negative sentiment). Of the 1.131 tokens, there are 604 tokens that are not available. The non-available tokens with the most occurrences include the words 'covid', 'berakhir', 'corona', 'orang', 'selalu', and 'pemerintahan'.

```
{'semoga': 46, 'banyak': 32, 'semua': 31, 'covid': 31, 'pandemi': 26, 'allah': 24, 'berakhir': 23, 'cepat': 23,

len(word_list)

1131
```

Figure 3.2 The most number of tokens found

```
covid 31
berakhir 23
corona 20
orang 19
selalu 17
medis 16
pemerintah 16
```

Figure 3.3 Results of the most tokens that were not found

The sentiment score calculation is done by adding up all the weight of the word sentiment. If there is a negation word in the previous token, the weight value is a negative sentiment because the word negation reverses the meaning of a word. The negation words that are covered are 'bukan', 'tidak', 'ga', and 'gk'. The strategy of taking word weight is applied starting from the word in the form of inflection, the root word, a combination of one word before, and a combination of two words before. All sentiment scores are obtained, then averaged.

```
[[ 45  11  10   2  -5  23  19  12  10  -2]
 [ 37  -1  25  10  -3   3  11  14 -21  21]
 [ -1  39 -29  11 -10   1   3  40  27  17]
 [ 17  29   1   4  11  16   5  35   9   2]
 [ 18  11 -10   5 -11   9  45  12  14   8]
 [ 11  35   1  -1  11  45  -6  40  14  28]
 [ 11  -1  37  14  32  12   7  19  -6   0]
 [ 25   6  -3  22  -5  -1  12  11   1 -27]
 [ 37  -5  16  11  -3  11  29  14  23  22]
 [ 45  37  14  12  18   6   8  -2   7  10]]
```

Figure 3.4 The results of the comment sentiment score on the news

```
[12.5  9.6  9.8 12.9 10.1 17.8 12.5  4.1 15.5 15.5]
```

Figure 3.5 The average sentiment score calculation result

D. *Decision Matrix*

The decision matrix is obtained by taking 5 criteria values from each news, namely the number of likes, comments, shares, average comment sentiment value, and news duration. The duration of the news is taken from the time difference between the upload of the news and the current time, which is converted into days. The value of each criterion is then normalized by the min-max method, so that it has a clear value limit. The minimum value limit used is 1, while the maximum value used is 5.

	likes	comments	shares	sentimen_val	durasi_berita
0	63636	1285	0	12.5	10.725087
1	263654	646	332	9.6	10.897587
2	106157	308	135	9.8	10.928849
3	1042	19	0	12.9	10.952136
4	197659	1303	459	10.1	11.077888
5	2833	49	0	17.8	11.184057
6	222208	733	478	12.5	11.622634
7	170032	1874	588	4.1	11.827796
8	73346	1489	0	15.5	11.853664
9	63068	714	0	15.5	12.210203

Figure 3.6 Result of Decision Matrix

E. *AHP TOPSIS*

Calculations using the AHP-TOPSIS method are carried out by finding the weights in the AHP method, and getting the order in the TOPSIS method. The calculation begins with initializing the value on a pairwise matrix based on data from experts. The next step is to normalize the pairwise matrix to get weight in the AHP process.

ALT./CRIT.	C1 (max)	C2 (max)	C3 (max)	C4 (max)	C5 (min)
C1	1	0.2	3	0.142857	3
C2	5	1	5	0.333333	3
C3	0.333333	0.2	1	0.142857	0.333333
C4	7	3	7	1	7
C5	0.333333	0.333333	3	0.142857	1
Total	13.6667	4.73333	19	1.7619	14.3333

Figure 3.7 Matrix Pairwise

Next, perform consistency ratio (CR) calculations to ensure that the obtained consistency ratio value is less than or equal to 0.1. The results of the calculation of the consistency ratio can be seen in the following figure.

```
c_vector = np.true_divide(AX, np.transpose(w))
c_vector

array([[5.42867208],
       [5.84676271],
       [5.19193382],
       [5.68899806],
       [5.06103492]])

lambdaMax = c_vector.mean()
CI = (lambdaMax-column)/(column-1)
RI = 1.12
CR = CI/RI

print("Lambda Max = ",lambdaMax)
print("CI = ",CI)
print("CR = ",CR)

Lambda Max =  5.443480317747239
CI =  0.11087007943680982
CR =  0.09899114235429447
```

Figure 3.8 Consistency Ratio

Based on Figure 3.9, it can be seen that the value of the consistency ratio obtained is less than 0.1, which is 0.09899. This proves that the value of the consistency ratio is consistent so that the weight can be accepted and can be continued to the next stage. After the weight value is obtained, the next step is to start the TOPSIS process by initializing the decision matrix based on data obtained from Facebook. Then, the vector normalization is performed on the decision matrix. The results of normalization calculations can be seen in Figure 3.10.

Normalize Matrix

	likes	comments	shares	sentimen_val	durasi_berita
0	0.201853	0.385747	0.112247	0.313206	0.108019
1	0.516669	0.243245	0.365756	0.090717	0.158205
2	0.268778	0.167869	0.215330	0.313206	0.167301
3	0.103334	0.103420	0.112247	0.453585	0.174076
4	0.412797	0.389761	0.462731	0.323800	0.210662
5	0.106153	0.110110	0.112247	0.236394	0.241550
6	0.451436	0.262647	0.477239	0.249637	0.369148
7	0.369314	0.517098	0.561233	0.392666	0.428837
8	0.217136	0.431240	0.112247	0.392666	0.436363
9	0.200959	0.258410	0.112247	0.241691	0.540094

Figure 3.10 Normalized Decision Matrix

Next, calculate the weighting on the normalized decision matrix. The results of this calculation can be seen in Figure 3.11.

	likes	comments	shares	sentimen_val	durasi_berita
0	0.022757	0.095570	0.005020	0.161010	0.008718
1	0.058250	0.060265	0.016357	0.046635	0.012769
2	0.030302	0.041590	0.009630	0.161010	0.013503
3	0.011650	0.025623	0.005020	0.233175	0.014050
4	0.046539	0.096565	0.020694	0.166457	0.017003
5	0.011968	0.027280	0.005020	0.121524	0.019496
6	0.050895	0.065072	0.021343	0.128332	0.029794
7	0.041637	0.128113	0.025100	0.201858	0.034612
8	0.024480	0.106842	0.005020	0.201858	0.035219
9	0.022656	0.064022	0.005020	0.124247	0.043592

Figure 3.11 Weighted Normalized Decision Matrix

After weighting, then determining the ideal solution is positive (A*) and the ideal is negative (A'). The next step, calculating the value of positive separation (S*) and negative separation (S'). The results of the calculation of the separation value can be seen in Figure 3.12.

```
Positive Ideal (S*)
[0.08904926 0.19872971 0.11720627 0.11448783 0.07531181 0.15904338
 0.12441399 0.04390016 0.06065575 0.13732935]


Negative Ideal (S')
[0.13897431 0.06670978 0.12087641 0.18886503 0.1468329  0.07868759
 0.10113136 0.18968824 0.17585648 0.08728817]
```

Figure 3.12 Positive and Negative Separation Values

The final step taken is to calculate the relative proximity of the alternatives to the positive and negative ideal solutions. The results of the calculation of relative proximity can be seen in Figure 3.13.

	text	c*	rank
0	#Update Perkembangan COVID-19 di Indonesia per...	0.609473	5.0
1	Stafsus Daniel Tinjau Kesiapan Rumkit TNI AD S...	0.251318	10.0
2	Peroleh Rekor Muri, Layanan PSC 119 di Sumsel ...	0.507708	6.0
3	Rakornas KKI: Meningkatkan profesionalisme dok...	0.622592	4.0
4	Protokol Kesehatan Ketat Percepat Pemulihan Ke...	0.660979	3.0
5	Selamat Pagi #Healthies\n\nAkhir-akhir ini beb...	0.330994	9.0
6	RSUD Buleleng Mampu Tes PCR Mandiri\n\nBali, 2...	0.448386	7.0
7	Perbaiki Sektor Pariwisata di Masa Pandemi COV...	0.812062	1.0
8	#Update Perkembangan COVID-19 di Indonesia per...	0.743541	2.0
9	Hai #Healthies\n\nSaat ini di Indonesia ada se...	0.388608	8.0

Figure 3.13 Relative Proximity

In Figure 4.13, it can be seen that the 7th alternative has the greatest relative durability value among all alternatives by obtaining the first rank, so that the 7th alternative with a news headline, namely "Improve the Tourism Sector during the COVID-19 Pandemic" is an alternative news with a positive impact best.

F. *Calculate Accuracy*

Accuracy calculations are performed using the accuracy_score function from the Sklearn library. The accuracy function performs calculations by dividing the number of appropriate alternative sequences by the number of all alternatives. In this study, the correct alternative order of experts is as shown in Figure 3.14 with an accuracy value of 80%.

```
[7, 10, 4, 9, 3, 6, 1, 2, 5, 8]
[7, 10, 4, 9, 3, 6, 1, 2, 8, 5]

accuracy_score(y_true, y_pred)

0.8
```

Figure 3.13 News Ranking Accuracy

IV. Conclusion

The implementation of AHP TOPSIS and lexicon-based sentiment analysis can be used to carry out information management by ranking COVID-19 news to get news that has a positive impact on society. This research is expected to help the government and the mass media in managing information related to COVID-19 news. The results of this study indicate the accuracy value obtained in this study reaches 80%.

References

[1] M. Q. Khairuzzaman, “Informasi Negatif Pandemi Covid-19 Bisa Pengaruhi Perilaku,” 2016. https://republika.co.id/berita/qa3rfm382/informasi-negatif-pandemi-covid19-bisa-pengaruhi-perilaku.

[2] I. Hidayatulloh and M. Z. Naf’an, “Integrasi Sentiment Analysis SentiWordNet pada Metode MOORA untuk Rekomendasi Pemilihan Smartphone,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 7, no. 1, 2018, doi: 10.22146/jnteti.v7i1.396.

[3] E. N. S. Purnomo, S. W. Sihwi, and R.

Anggrainingsih, “Analisis Perbandingan Menggunakan Metode AHP, TOPSIS, dan AHP-TOPSIS dalam Studi Kasus Sistem Pendukung Keputusan Penerimaan Siswa Program Akselerasi,” *J. Itsmart*, vol. 2, no. 1, 2013.

[4] E. Martua, “Twitter-COVID19-Indonesia-Sentiment-Analysis---Lexicon-Based,” 2020. https://github.com/evanmartua34/Twitter-COVID19-Indonesia-Sentiment-Analysis---Lexicon-Based.

[5] A. N. Pramudhita, “Komparasi Algoritma Multi Criteria Decision Making Dengan Metode AHP dan SAW dalam perangkingan calon karyawan,” no. 01, pp. 85–89.