Twitter Sentiment Analysis in Tourism with Polynomial Naïve Bayes Classifier

Ahmad Ashril Rizal[1,*], Gibran Satya Nugraha[2], Rian Asmara Putra[3] and Dara Puspita Anggraeni[4]

[1] Universitas Islam Negeri Mataram; ashril@uinmataram.ac.id
[2] Universitas Mataram; gibransn@unram.ac.id
[3] Universitas Bumigora; rian.asmara@gmail.com
[4] Universitas Nahdlatul Wathan; darapuspitaanggraeni40@gmail.com
[*] Correspondence: ashril@uinmataram.ac.id

Abstract: Lombok has become a favorited tourist destination in the world. Therefore, tourism is a mainstay sector in regional development in West Nusa Tenggara. The contribution of the tourism sector shows an increasing trend. Tourist expenditures are distributed to various sectors. The tourism sector has a positive impact on the regional economy. The local government has prepared to improve the quality and quantity of tourism in Lombok. The results of local government efforts need to be analyzed so that future policies are on target. Analysis can be done on the satisfaction of tourists who travel to Lombok. It would be very difficult to get satisfaction data from all tourists through questionnaires. But on the other hand, tourist satisfaction is usually posted on their social networks. One of the social media that is widely used by tourists is Twitter. Their tweets contain not only expressions of natural beauty but also criticism, suggestions, and complaints during their visit. In addition, the tweet data on twitter is open access. This study tries to analyze the sentiment on Twitter which contains tweets of tourists who have visited Lombok. Sentiment analysis is performed using the Polynomial Naive Bayes Classifier. Sentiment results are classified into positive and negative sentiments. The results of this sentiment are expected to help related agencies or other tourism actors to improve the quality and quantity of regional tourism. The results showed that the positive sentiment on the security factor were 50.65%, the cost 75.32%, accommodation 62.33% and the cleanness factor 77.92%.

Keywords: public opinion, sentiment analysis, polynomial naive bayes classifier, tourism factor

1. Introduction

Lombok is known as one of the best tourism destinations in Indonesia. In 2015, Lombok was awarded as World Best Halal Tourism Destination and World Best Halal Honeymoon Destination in World Halal Travel Summit 2015 which was held in Abu Dhabi, UEA. These categories have been voted by people through online voting. Nowadays, the number of tourists visiting Lombok increase. They are really interested to visit many gillies, beach, and natural waterfalls. Besides, there is a Rinjani mount which was interested to visit by some tourists. It is hard to make Lombok as a famous destination of tourism. But there is an internet technology which can help a lot of tourists to find the information of Lombok. By using internet technology, tourists can share about many places that have been visited by them in their social media. These give positive and negative perception in their feeds [1]. Based on other results of the research about system prediction tourist arrivals using Recurrent Neural Network with Extended Kalman Filter with the best MSE value when training process is 0.000949259 at the system with η=0.2. Prediction time series data of tourist arrival using the RNN-EKF has the best accuracy rate of 86.059%. The best

accuracy is obtained from RNN system with η=0.3 [2], [3]. The other research tried to track trending topics in social media Instagram since COVID-19 hit. The results of trending topics will be classified by sentiment analysis using a Lexicon-based and Naive Bayes Classifier. Based on Instagram data taken since January 2020, it shows the five highest topics in the tourism sector, namely health protocols, hotels, homes, streets, and beaches. Of the five topics, sentiment analysis was carried out with the Lexicon-based and Naive Bayes classifier, showing that beaches get an incredibly positive sentiment, namely 80.87%, and hotels provide the highest negative sentiment 57.89%. The accuracy of the Confusion matrix's sentiment results shows that the accuracy, precision, and recall are 82.53%, 86.99%, and 83.43%, respectively [4].

Twitter is a social media platform using text to express the user feelings such as angry, sad, and happy. Text mining clarify the sentiment which was had by common document called sentiment analysis. Sentiment analysis is a scientific discipline that concern with how to analyze opinions, sentiment, estimation, attitude, and emotion of an entity such as products, service, organization, individual, issues, and topics. The sentiment can be classified by some methods or classified. This study aims to find trending topics and conduct sentiment analysis on trending topics related to the province of West Nusa Tenggara. This analysis uses tweet data from January 1, 2019, to December 31, 2020. The method used is Lexicon Based with the python programming language. The research stages consist of crawling the dataset, preprocessing, finding trending topics, lexicon-based sentiment analysis, confusion matrix test and visualization. The conclusion from the analysis related to trending topics is that the top ten trending. The test accuracy value is above 80% and the average score for all topics is 100% on accuracy, 100% on precision and 100% on recall [5]. Next research by [6] on banking technology, a bank must collect the information about service perceptions if the service is suitable or not for customers. The easiest way to detect it is by asking the customers both to fill a form and to give some suggestions about the bank's services. These suggestions can be analyzed by sentiment analysis. The method used to classify is k-NN method. However, k-NN method is depend on the closest k value so other data is neglected. That is why, k-NN fuzzy is used to solve this problem. Based on the review of m-Banking application, the accuracy was 94% with F-Measure value was 0.9273. [7] have compared Naive Bayes method and k-Nearest Neighbor method to analyze work status in Demak regency in 2012. The result of this study showed that the calculation using APER, both of Naive Bayes method and k-NN method had small opportunity to find mistakes in classifying. But based on APER calculation, it was found that k-NN method is better than Naive Bayes method to classify work status in Demak regency in 2012. [8] also had classified the candidates of governor in West Java in 2018 using Naive Bayes method. The result of this research showed that the calculation using Naive Bayes method had good enough accuracy (76,65%). When the Naive Bayes algorithm was applied to test eligibility for aid recipients, the overall accuracy was 62.86%, recall 78.57% and precision 52.38% [9]

Sentiment mining is one of an aspect in data mining where the data can be classified based on positive class or negative class. Sentiment analysis is also known as an opinion mining [10]. Research on sentiment analysis has been carried out using kNN and naive Bayes algorithms. The data used comes from comments on the online travel agent's Facebook fanpage[11]. In this research, it will be concerned in classifying process of sentiment analysis using Naive Bayes Classifier based on data set which contain supporting factors such as accommodation, cleanness, safety, and cost [12]. Sentiment Analysis's is used to analyze the sentiment or emotion of the text. In this study, some tweets from twitter user publishing about tourism destination in Lombok Island will be used as an analyzed experiment. The result of this analysis is expected can give some ideas for entrepreneurs in tourism area and the government to increase the service quality in tourism sector [13].

2.   Materials and Methods

2.1. Data Analysis

There were two types of data which was used in this research including tweets data and the data which was divided into two types: training data and testing data [14]. Tweets data is tweets document used as model to find out the sentiment result. Tweets data is the document using Indonesian language obtaining from crawling through a created system. The data were divided into training data and testing data. Training data is labelled data while testing data is data that will be analyzed using a created model. The result of testing become the prediction of used method [15].

2.2. Pre-processing

Pre-processing is important process to confirm the sentiment result by Naive Bayes method. Pre-processing is the beginning process in determining the sentiment result from tweets data. This process is divided into some process including removing symbols stemming, tokenizing, and lowercase.

2.2.1. Lowercase. Lowercase process is processed to change all the characters in tweet data into small alphabet. This process will help us in determining the sentiment. For example: it can be seen in Table 1

Table 1. Lowercase Example

No	Tweet before pre-processing	Tweet after pre-processing
1	@rlthingy Ini yg ada di lombok jg ref hehe gimana gimana? :D @sesukaqu selama aku idup di lombok aku gaperna ikut b… ttps://t.co/WsyTwz8tjr	rlthingy ini yg ada di lombok jg ref hehe gimana gimana d sesukaqu lama aku idup di lombok aku gaperna ikut b https t co wsytwz8tjr
2	HUT Korem 162/WB Meriahkan Festival Bau Nyale 2019 https://t.co/GPIlDVHGhc lewat @KO-DIM 1620/LOTENG https://t.co/nlZ2KP5hpd	hut korem 162 wb riah festival bau nyale 2019 https t co gpildvhghc lewat kodim 1620 loteng https t co nlz2kp5hpd
3	Meriahkan Pesona Bau Nyale 2019, Mandalika Fashion Carnaval Digelar https://t.co/meM2AnpYEA https://t.co/8eejHKxc2j	riah pesona bau nyale 2019 mandalika fashion carnaval gelar https t co mem2anpyea https t co 8eejhkxc2j
4	Acara festival budaya sasak sekali setahun dilombok, yaitu acara "BAU NYALE" yang sangat dinanti-nantikan oleh masy… https://t.co/InELWPMXC3	acara festival budaya sasak sekali tahun lombok yaitu acara bau nyale yang sangat nanti oleh masy https t co inelwpmxc3

2.2.2. Removing Symbol. This process will remove all of symbols in tweets. For example: dot symbol (.), commas symbol (,), exclamation mark (!), and question symbol (?), etc. These symbols were removed because they were not giving some impacts in determining the sentiment. The example of removing symbol is shown in Table 2.

Table 2. Removing Symbol Example

No	Tweet before pre-processing	Tweet after pre-processing
1	@rlthingy Ini yg ada di lombok jg ref hehe gimana gimana? :D @sesukaqu selama aku idup di lombok aku gaperna ikut b… https://t.co/WsyTwz8tjr	rlthingy ini yg ada di lombok jg ref hehe gimana gimana d sesukaqu lama aku idup di lombok aku gaperna ikut b https t co wsytwz8tjr

No	Tweet before pre-processing	Tweet after pre-processing
2	HUT Korem 162/WB Meriahkan Festival Bau Nyale 2019 https://t.co/GPIlDVHGhc lewat @KODIM1620/LOTENG https://t.co/nlZ2KP5hpd	hut korem 162 wb riah festival bau nyale 2019 https t co gpildvhghc lewat kodim 1620 loteng https t co nlz2kp5hpd
3	Meriahkan Pesona Bau Nyale 2019, Mandalika Fashion Carnaval Digelar https://t.co/meM2AnpYEA https://t.co/8eejHKxc2j	riah pesona bau nyale 2019 mandalika fashion carnaval gelar https t co mem2anpyea https t co 8eejhkxc2j
4	Acara festival budaya sasak sekali setahun dilombok, yaitu acara "BAU NYALE" yang sangat dinanti-nantikan oleh masy… https://t.co/InELWPMXC3	acara festival budaya sasak sekali tahun lombok yaitu acara bau nyale yang sangat nanti oleh masy https t co inelwpmxc3

2.2.3. Stemming. This process is used to replace words into their basic words by removing all the affixes including prefix, confix, and suffix. In this process, it will be used library sastrawi to replace words into their basic words. The example of this process is shown in Table 3.

Table 3. Stemming process

No.	Affixes	Basic
1	Meriahkan	Riah
2	Digelar	Gelar
3	Setahun	Tahun
4	…	…

2.3. *Naïve Bayes Classifier*

The next process is to determine the sentiment using Naïve Bayes. This process is a main process in determining both of positive and negative sentiments. Before doing this step, we begin it by crawling process of data set, then determine the sentiment of tweets data which was obtained. Naïve Bayes is one of classifying method based on simple probability which is assumed that inter class were independent to each other [16]. In Naive Bayes, it is concerned to estimate the probabilities. One of benefits in using Naive Bayes approach is smaller error values although we used big number of data set. Besides, classifying by Naïve Bayes has good accuracy and high velocity when it was applied in big number of data set. In this research, we used Naïve Bayes method because of there are big number data set so it will need a fast method to classify and give high accuracy in analysis [17].

There are two processes for Naïve Bayes method to classify the texts. They were training process and classifying process. In training process, the trained data which were become representation will be processed. Then, we will determine prior probabilities for every category based on the data sample. In classifying process, a category value will be determined based on the term of classified data. Formulation of Naive Bayes [4] is shown below [18].

$$P(c|X) = P(x_1|c)xP(x_2|c)xP(x_3|c)x\dots xP(x_n|c)xP(c) \tag{1}$$

$P(c|X)$ is posterior (probability of c in X. It can be calculated from prior probability of X in c divided with the number of all probability X in all of c).

2.3.1. Crawling Tweets

Pre-processing is an important process in determining sentiment results using Naïve Bayes method. This pre-processing step will decide the sentiment result. Pre-processing

has so many processes that will be explain one by one, like symbols, stemming, tokenizing, and lowercase.

2.3.2. Sentiment Decision

The last process is determining the negative and positive sentiments using Naïve Bayes classifier. The final result of sentiment depends on both pre-processing and the dictionary that was used in this step.

3.　　Results

3.1. Dataset

The table of data set is data groups which was used in training process. The data set was divided based on their own class where 0 was recognized as negative sentiment and 1 was recognized as positive sentiment. For example: the writer used 4 data (shown in Table 4).

Table 4. Dataset

No	Tweets	Class
1	Lombok itu indah, tapi mahal	0
2	Lombok sudah kondusif	1
3	Liburan di Lombok menyenangkan	1
4	Di Lombok banyak begal	0

3.2. Tokenize

Tokenizing is used to collect the number of words in data set. The data can be a single word. It means that if there are two words or more than two words in data set, we only can use one word. The result of tokenize can be seen in Table 5.

Table 5. Tokenize results

| No	Word
1	Lombok
2	Itu
3	Indah
4	Tapi
5	Mahal
6	Sudah
7	Kondusif
8	Liburan
9	Di
10	Meneyenangkan
11	Banyak
12	Begal

3.3. Bag of Words

Bag of words is a concept from text analysis, this concept represents document as important information pocket without sorting its words. This method works by counting the total of words' frequency who appeared in a document dataset. So, the output of Bag of words model is a frequency vector. The process of bag of words can be seen in Table 6.

Table 6. Bag of words process

Word	No.			
	1	**2**	**3**	**4**
lombok	1	1	1	1
itu	1			
indah	1			
tapi	1			
mahal	1			
sudah		1		
kondusif		1		
liburan			1	
di			1	1
menyenangkan			1	
banyak				1
begal				1
Kelas	0	1	1	0

3.4. *The Result of Sentiment Analysis in Security Factor*

After finishing the pre-processing step and data training, the model that has been trained will be tested in testing data of safety factor. The number of data set used in this test was 77 data. Those data were some tweets which related to safety factor in tourism area. Those data were raw data which were not labelling yet. The result of prediction for safety factor is shown in Figure 1.
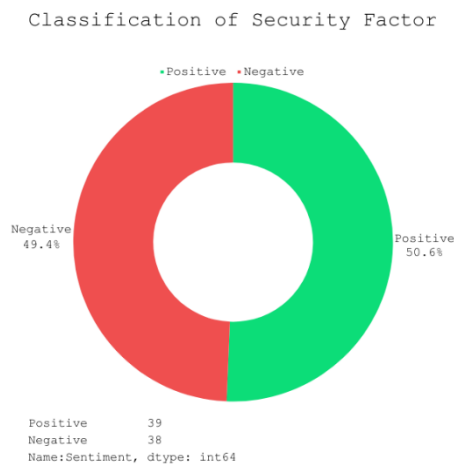
Figure 1. Classification graphic of security factor

Figure 1 showed that positive sentiment was recognized by 1 and negative sentiment was recognized by 0. The number of positive data was 39 meanwhile the number of negative data was 38. The testing using confusion matrix is to make sure the classification results. The result of this test is shown in Table 7 below.

Table 7. Confusion Matrix Of Security Factor

Class	Classified as Positive	Classified as Negative
Positive	31	0
Negative	3	43

Accuracy:

$$\frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{31 + 43}{31 + 43 + 3 + 0} \times 100\% = 96{,}1\%$$

Precise:

$$\frac{TP}{FP + TP} \times 100\% = \frac{31}{3 + 31} \times 100\% = 91{,}1\%$$

Recall:

$$\frac{TP}{FN + TP} \times 100\% = \frac{31}{0 + 31} \times 100\% = 100\%$$

3.5. *The Result of Sentiment Analysis in Cost Factor*

Next step is testing model that has been trained on testing data for cost factor. The number of data set used in this test was 77 data. Those data were some tweets which related to cost factor in tourism sector. Those data were raw data which were not labeling yet. The result of prediction for safety factor is shown in Figure 2.
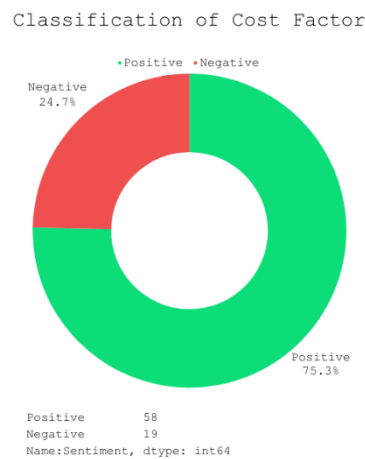

Figure 2. Classification graphic of cost factor

Based on graph in Figure 2, it showed that the number of positive data was 58 and the number of negative data was 19. The aim of analyzing by confusion matrix is to make sure the classification results. The result of this test is shown in Table 8 below.

Table 8. Confusion matrix of cost factor

Class	Classified as Positive	Classified as Negative
Positive	53	0
Negative	5	19

Accuracy:

$$\frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{53 + 19}{53 + 19 + 5 + 0} \times 100\% = 93{,}5\%$$

Precise:

$$\frac{TP}{FP + TP} \times 100\% = \frac{53}{5 + 53} \times 100\% = 91,3\%$$

Recall:

$$\frac{TP}{FN + TP} \times 100\% = \frac{53}{0 + 53} \times 100\% = 100\%$$

3.6. *The Result of Sentiment Analysis in Accommodation Factor*

Next step is testing model that has been trained on testing data for accommodation factor. The number of data set used in this test was 77 data. Those data were some tweets which related to accommodation factor in tourism sector. Those data were raw data which were not labeling yet. The result of prediction for safety factor is shown in Figure 3.
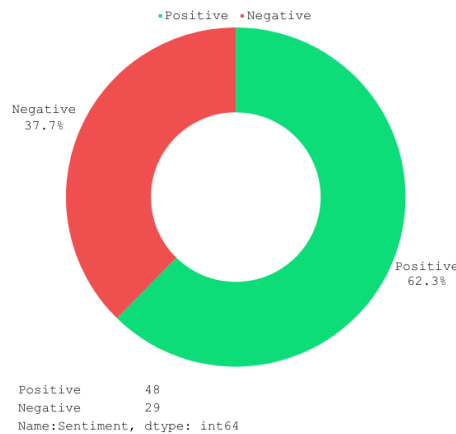

Figure 3. Classification graphic of accommodation factor

Based on graph in Figure 3, it showed that the number of positive data (recognized by 1) was 48 and the number of negative data (recognized by 0) was 29. The aim of analyzing by confusion matrix is to make sure the classification results. The result of this test is shown in Table 9 below.

Table 9. Confusion matrix of accommodation factor

Class	Classified as Positive	Classified as Negative
Positive	43	1
Negative	5	28

Accuracy:

$$\frac{TP + TN}{TP + TN + FP + FN} x100\% = \frac{43 + 28}{43 + 28 + 5 + 1} x100\% = 92,2\%$$

Precise:

$$\frac{TP}{FP + TP} x100\% = \frac{43}{5 + 43} x100\% = 89,5\%$$

Recall :

$$\frac{TP}{FN + TP} x100\% = \frac{43}{1 + 43} x100\% = 97,7\%$$

3.7. The Result of Sentiment Analysis in Cleanness Factor

Next step is testing model that has been trained on testing data for cleanness factor. The number of data set used in this test was 77 data. Those data were some tweets which related to cleanness factor in tourism sector. Those data were raw data which were not labeling yet. The result of prediction for safety factor is shown in Figure 4.
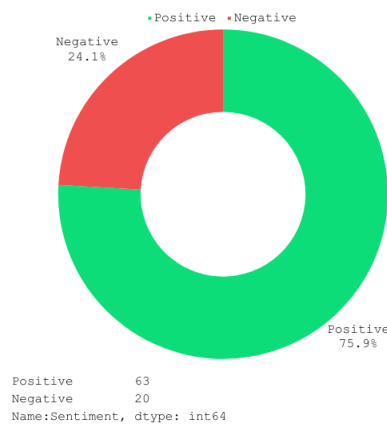

Figure 4. Classification graphic of cleanliness factor

Based on graph in Figure 4, it showed that the number of positive data (recognized by 1) was 63 and the number of negative data (recognized by 0) was 20. The aim of analyzing by confusion matrix is to make sure the classification results. The result of this test is shown in Table 10 below.

Table 10. Confusion matrix of cleanness factor

Class	Classified as Positive	Classified as Negative
Positive	35	11
Negative	24	7

Accuracy:

$$\frac{TP + TN}{TP + TN + FP + FN} x100\% = \frac{35 + 7}{35 + 7 + 24 + 11} x100\% = 54,5\%$$

Precise :

$$\frac{TP}{FP + TP} x100\% = \frac{35}{24 + 35} x100\% = 59,3\%$$

Recall :

$$\frac{TP}{FN + TP} x100\% = \frac{35}{11 + 35} x100\% = 76\%$$

4. Discussion

The classification result of the security factor has a tendency toward positive sentiment (39 tweets). One of the local media outlets in Lombok, called Radar Lombok, published on March 18, 2019, said that the destination on Lombok Island is now safe to visit, although it is still not stable yet because of the earthquake that happened on Lombok Island a few months ago. However, based on an interview with Radar Lombok's journalist, Lombok is now safe and can be visited by tourists. In safety factors, researchers not only take tweets about criminality but also take tweets about destination safety on Lombok Island.

Classification result of cost factor has tendency to positive sentiment (58 tweets). It caused by the price of tour packet likes accommodation, tour, and food quietly cheap said online media in Lombok. And tour packet that offered by travel agent is quietly cheap, so it is not bothered enough the tourist who will come to Lombok. Classification result of accommodation factor get 48 values for positive sentiment. As the travel agent said that tour packet in Lombok has various type with cheap price, and easy to find them. The hotel provides rent car or motorcycle to facilitate the tourist transportation.

Classification result of cleanness factor has tendency to positive sentiment with 60 values. Because cleanness becomes main problem for tourism place. Some tweets said some tourism place in Lombok is quietly clean, it cannot be separated from the awareness level of cleanness by the citizen who raised up lately.

5. Conclusions

Sentiment analysis for tourism sector is successfully conducted using Naïve Bayes. It comes with some phase likes crawling dataset from Twitter, pre-processing, grouping positive and negative sentiment, testing process, and the last is data visualization. Sentiment analysis for tourism sector grouped into four factors: security, cost, accommodation, and cleanness. Those 4 factors have more positive sentiment than negative. Cleanness factor has the most percentage with 77.92%, followed by cost factor (75.32%), accommodation factor (62.33%), and security factor (50.64%)

Cleanness factor has the most percentage because it becomes the main concern for a tourism place in Lombok to improve comfortable feeling of the tourist. Besides security factor has the less percentage because of the earthquake that happened last year who still shadowed citizen in Lombok especially for tourism place like beach.

References

[1]   I. N. Wijaya, "Pengaruh Jumlah Wisatawan Mancanegara, Lama Tinggal, Dan Kurs Dollar Amerika terhadap Penerimaan Produk Domestik Regional Bruto Industri Pariwisata di Badung," Universitas Udayana, 2011.

[2]   A. A. Rizal and S. Hartati, "Recurrent neural network with Extended Kalman Filter for prediction of the number of tourist arrival in Lombok," in *2016 International Conference on Informatics and Computing (ICIC)*, IEEE, 2016, pp. 180–185. doi: 10.1109/IAC.2016.7905712.

[3]   A. A. R. Rizal and S. Hartati, "Prediksi kunjungan wisatawan dengan recurrent neural network extended kalman filter," *Ilmu Komputer, Udayana*, vol. I, no. 1, p. 1, 2017.

[4]   A. Priadana and A. A. Rizal, "Sentiment Analysis on Government Performance in Tourism During The COVID-19 Pandemic Period With Lexicon Based," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 7, no. 1, pp. 28–39, Nov. 2021, doi: 10.18860/ca.v7i1.12488.

[5]   M. N. Asti, I. Ismarmiaty, and A. A. Rizal, "Lexicon Based Sentiment Analysis pada Trending Topics di Nusa Tenggara Barat," *Jurnal Informatika dan Teknologi Komputer*, vol. 3, no. 2, pp. 93–98, 2022, doi: https://doi.org/10.33059/j-icom.v3i2.6136.

[6]   D. G. H. Divayana, P. W. Arta Suyasa, I. M. Agus Wirawan, and I. M. Putrama, "Pemberdayaan Materi Ajar Berbentuk Digital Menggunakan Aplikasi Open Office Sun Microsystem Bagi Guru-Guru Sma Se-Kecamatan Ubud," *Jurnal Widya Laksana*, vol. 5, no. 2, p. 69, 2017, doi: 10.23887/jwl.v5i2.8870.

[7]   R. Eko Putri and R. Rahmawati, "Perbandingan Metode Klasifikasi Naive Bayes dan K-Nearest Neighbor Pada Analisis Data Status Kerja di Kabupaten Demak Tahun 2012," vol. 3, no. 4, pp. 831–838, 2014, doi: https://doi.org/10.14710/j.gauss.3.4.831-838.

[8]   H. Simada Ginting, K. Muslim Lhaksmana, and D. Triantoro Murdiansyah, "Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter Menggunakan Naive Bayes," in *e-Proceeding of Engineering*, 2018.

[9]   N. Riyanah, S. Informasi, S. Tinggi, M. Informatika, D. Komputer, and N. Mandiri, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu (Implementation of Algorithms Naïve Bayes for Classification Recipients Help Letter Description Not Able)," *JTIM*, vol. 2, no. 4, pp. 206–213, 2021.

[10] C. R. Fink, D. S. Chou, J. J. Kopecky, and A. J. Llorens, "Coarse- and fine-grained sentiment analysis of social media text," *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)*, vol. 30, no. 1, pp. 22–30, 2011.

[11] E. Wahyu Sholeha, S. Yunita, R. Hammad, V. Cahya Hardita, T. Rekayasa Komputer Jaringan, and P. Tanah Laut, "Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor (Sentiment Analysis of Online Travel Agent Using Naïve Bayes and K-Nearest Neighbor)," *JTIM*, vol. 3, no. 4, pp. 203–208, 2022.

[12] A. P, Pi. R, and Pp. P, "Sentiment Analysis in Tourism," *International Journal of Innovative Science, Engineering & Technology*, vol. 1, no. 9, pp. 443–450, 2014.

[13] P. Kinerja, P. Di, J. Akuntansi, F. Ekonomi, D. A. N. Bisnis, and U. Hasanuddin, "Faktor-faktor yang memengaruhi pengungkapan kinerja perusahaan di website," 2013.

[14] A. A. Rizal and S. Soraya, "Multi Time Steps Prediction Dengan Recurrent Neural," *Matrik*, vol. 18, no. 1, pp. 115–124, 2018.

[15] M. Bouazizi and T. Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter," *IEEE Access*, vol. 5, pp. 20617–20639, 2017, doi: 10.1109/ACCESS.2017.2740982.

[16] S. Rana and A. Singh, "Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques," *Proceedings on 2016 2nd International Conference on Next Generation Computing Technologies, NGCT 2016*, no. October, pp. 106–111, 2017, doi: 10.1109/NGCT.2016.7877399.

[17] D. Hardt and F. K. Glückstad, "A social media analysis of travel preferences and attitudes, before and during Covid-19," *Tour Manag*, vol. 100, Feb. 2024, doi: 10.1016/j.tourman.2023.104821.

[18] A. C. Ojha, P. K. Shah, S. Gupta, and S. Sharma, "Classifying Twitter Sentiment on Multi-Levels using A Hybrid Machine Learning Model," *OriginInternational Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 12, no. 3s, pp. 328–333, 2024, [Online]. Available: www.ijisae.org