

Sentiment Analysis on Jurassic Park Development in the Komodo Conservation Area

Annisa Septiana Sani¹, Indra Budi²
Faculty of Computer Science, Universitas Indonesia
Jakarta, Indonesia
¹annisa.septiana91@ui.ac.id
²indra@ui.ac.id

Abstract— Rinca Island is included in one of the islands in NTT Province, which is used as a National Park to protect Komodo dragons. According to the government's agenda, in June 2021, Rinca Island will build a world-class premium tourist destination project for tourists who want to see Komodo, which is called Jurassic Park Komodo. The existence of tourism is expected to be able to attract tourists and investment. However, the news about project development that is not in accordance with the conservation value has created public sentiment towards the project development process. The purpose of this study is to analyze public sentiment related to the classification of support & against the development of Jurassic Park in the Komodo Conservation Area and to measure the accuracy using two methods, namely Naïve Bayes and Decision Tree.

Keywords— sentiment analysis, komodo, twitter, Naïve Bayes, Decision Tree.

I. INTRODUCTION

In 1980, Komodo Island, Padar Island, and Rinca Island were designated as national parks to protect the komodo dragon or *Varanus komodoensis* [1]. Komodo is one of the ancient animals and legends; these animals are protected due to the fact that the animal is vulnerable to extinction, and currently, its natural habitat can only be found in the Province of Nusa Tenggara Timur (NTT). Tourists who have a vacation in NTT usually enter their agenda to see komodo dragons, but unfortunately, the komodo dragon is a dangerous animal where the saliva they spit can be very deadly. Therefore, it is not uncommon for tourists not to have the opportunity to be able to see and take long-distance selfies with the komodo dragons. The Coordinating Minister for Maritime Affairs sees this as an opportunity to bring in investors by creating a world-class premium tourist destination for tourists who want to see komodo dragons. The premium tourism, which is called Jurassic Park Komodo, is planned to be completed in June 2021 [1].

Of the total area of Rinca Island, which is 19800 hectares, 2.5% will be used for tourist recreation locations. The plan is that with the construction of Jurassic Park, it will have a research center and it is hoped that the local community will be more prosperous [2]. According to the government, development is in accordance with the regulations because it is carried out in a special area that does not disturb the ecosystem. This process has also been overseen by UNESCO and the Ministry of Environment and Forestry. However, this is in contrast to the tweet uploaded by KawanBaikKomodo,

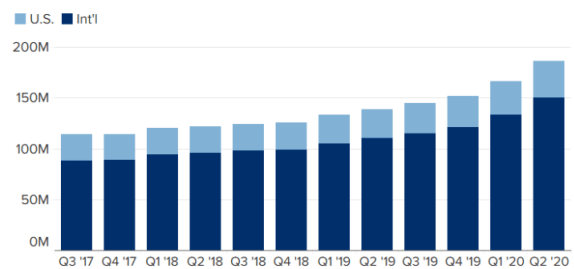
which displays a photo of a truck facing Komodo on October 26, 2020.

Since this tweet appeared and went viral on Twitter, various public sentiments have surfaced on the platform. Many discrepancies occurred in the field. The process and concept of tourism in the komodo dragon habitat are not environmentally friendly and contradict the concept of conservation, so the Jurassic Park development process can depress the animals there. In addition, UNESCO only gives titles for branding with an economic purpose but does not have an ethical responsibility for environmental and social issues that arise [2].

Many of the viral news originating from Twitter as the number of daily Twitter users globally are claimed to increase can be seen in Figure 1. In the 2nd quarter of 2020, Twitter's financial reports, daily active users on the Twitter platform were recorded to have increased to 180 million users [3]. Country Industry Head Twitter Indonesia claims that Indonesia is one of the countries with the largest growth in daily active Twitter users. Twitter users can send and read text-based messages known as tweets, and tweets usually contain opinions on something. Therefore, Twitter is a medium that can be used to collect sentiment for analysis. The sentiment itself is a term for presenting topics objectively and subjectively as well as facts and opinions by showing the difference between positive and negative topics [4].

Twitter user growth

Daily users defined as monetizable by Twitter ("mDAU")



SOURCE: Twitter. Company defines monetizable users as "Twitter users who log in and access Twitter on any given day through Twitter.com or our Twitter applications that are able to show ads."



Figure 1 Twitter User Development
(Source: [3])

From the problems that have been described, the purpose of writing this research is to analyze public sentiment related

to the classification of support & resistance to the development of Jurassic Park in the Komodo Conservation Area and measure the results of accuracy with two methods, namely Naïve Bayes and Decision Tree. From the problems that have been described, the purpose of writing this research is to analyze public sentiment related to the classification of support & rejection of the Jurassic Park development in the Komodo Conservation Area and measure the accuracy results with two methods, namely Naïve Bayes and Decision Tree. It is hoped that the results of this sentiment analysis can show the tendency of opinions from users on the Twitter platform to be taken into consideration and more serious attention by the government, seeing that Komodo is a world heritage and shows the accuracy of the comparison of the two methods that will be used.

II. LITERATURE REVIEW

A. Jurassic Park development on Rinca Island

Rinca Island will build a geopark planned to be named Jurassic Park Komodo. The goal is that this geopark will become world-class premium tourism, and it is hoped that it will attract a lot of tourist attention. The geopark development that is currently underway has pocketed the permit—starting from the issuance of Ministerial Regulation of Environment and Forestry concerning natural tourism exploitation permits in wildlife reserves, grand forest parks, national parks, and nature tourism parks. Until finally, in 2019, KLHK revised the previous Ministerial Regulation, which was published in 2020.

Through this regulation, the Ministry of Environment and Forestry shares authority with the Ministry of Maritime Affairs and Investment, the Ministry of Tourism and the NTT Government were arranging Komodo Island into an exclusive tourist destination. Supported by the Minister of Environment and Forestry's statement stating that Rinca Island will be jointly managed by the central government and the NTT Provincial Government as world-class tourism and investment.

From Sentiment Analysis of Labuan Bajo Tourist Destination on Instagram by Vegeterrikin Gousander, in his research, the data obtained from Instagram are managed in the form of feeds that are generally published by users on Instagram. The data is in the form of text, sentiment analysis is carried out using a Lexicon-based approach using the VADER sentiment model and analysis is carried out on the use of hashtags to find out which destinations are often posted. The results of the study stated that the lexicon-based sentiment method using VADER sentiment gave an accuracy value of 72% [5]. The similarity between the research conducted by Gousander and this research is that they both conduct sentiment analysis with the object of the same place, namely the Komodo conservation area and differences can be found in the data collection and the method of approach accuracy.

B. Sentiment Analysis

Sentiment is a term used to describe subjective and objective topics and factual or non-factual topics that transcend the difference between positive or negative topics

[6]. Sentiment analysis is an analytical approach used to analyze text. Sentiment analysis aims to determine the subjectivity of opinions, reviews, or tweets. Based on sentiment analysis, a person's opinion can be classified into various categories based on the size of the data and the type of document [7]. Sentiment analysis is carried out by grouping the text in a sentence or document then determining the opinion expressed in the sentence or document, whether it is positive, negative, or neutral. Sentiment analysis can also be carried out to analyze people's opinions on an emerging issue. Sentiment analysis or opinion mining can also be used to identify the polarity of consumer reviews on the various dimensions that will be the trust score [8].

In a study conducted by Wongkar and Angdresey, they conducted a sentiment analysis on the Indonesia presidential candidates using Twitter data. The method they use is to do text processing from the data collected and then use Naïve Bayes to make predictions. Where the accuracy count for Naïve Bayes is better, which is 80.90%, when compared to the KNN method, which has 75.58% accuracy and SVM reaches 63.99% [9]. Their study has similar topic and method compared to this research, scilicet sentiment analysis using the Naïve Bayes method.

C. Naïve Bayes

One method for measuring accuracy is Naïve Bayes, which is a method that can be trained on small-scale data and can provide real-time predictive results. This method can also help in classifying classes whose results can be used in parallel in increasing the scale of the dataset, especially in case studies of large-scale data. [10].

D. Decision Tree

Decision Tree is a decision support method that is popular in machine learning. The Decision Tree uses a tree-like model and a way of displaying algorithms that contain only conditional control statements. Decision Tree can be used in research, especially in decision analysis, to help identify the strategies that are most likely to achieve goals [11].

III. METHODOLOGY

This section will explain some of the stages carried out in this research. The research stages can be seen in Figure 2.

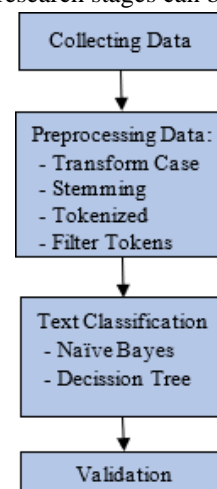


Figure 2. Research methodology

IV. RESULT AND DISCUSSION

A. Collecting Data

The data collected were in the form of tweets with the keyword Rinca Island which were taken from October 25, 2020 to January 6, 2021, because on October 26, this issue became a trending topic in Indonesia. Retrieval of data using a crawling technique using the Twitter API. The resulting dataset is 695, with 489 as training data and 200 as test data.

B. Data Preprocessing

Data on the internet is generally incomplete, noisy, and inconsistent. Data preprocessing helps in data cleaning, data reduction, and data discretization. The following preprocessing phases will make the dataset more precise [12]. At this stage, there are several methods used, namely transform case, stemming, tokenized fan filter tokens. The method increases the accuracy of the Naïve Bayes classifier and reduces processing time [12]. The following is an explanation of each method.

1) Case Folding

This process is carried out using the Python programming language. The goal is to change all words to lowercase so that all text has the same writing format. In addition, it removes unnecessary text such as links, numbers, usernames and punctuation marks.

2) Tokenized

The goal of doing this phase is to separate sentences into words, the constituent tokens. This aims to be able to give weight to each word that is formed so that the prediction results can be more accurate.

3) Stemming

Stemming is the process of changing a word into a root word. The purpose of stemming is so that one word has the same meaning, the writing is also the same. At this stage, it will remove the prefix, suffix, and confix prefixes. The stemming process is carried out programmatically using the Python programming language.

4) Stopwords Removal

Stopwords are a collection of common words that usually appear in sentences that don't really matter. Examples of stopwords are "and", "from", "at", "the", and others. The purpose of doing this step is to remove words that are considered unimportant which have low information from a text in order to focus on important words.

C. Validation

Validation is done by calculating the accuracy of each proposed method. Accuracy is the accumulation of the number of documents classified as true positive and true negative. Meanwhile, recall is the number of positive classifications of documents that have positive classifications. The selection of the Naïve Bayes method is because this method is quite simple but works well as well as the much more complicated method, while in the Decision Tree there is no need to standardize or normalize the data that has been collected. Decision Tree can handle continuous and categorical variables.

This section will contain the results and discussion that has been carried out in this study. Approximately 1500 tweets were obtained from crawling using the Twitter API with the keywords Rinca Island and Komodo, but after being independently sorted and analyzed, the results of the tweets were valid and correct, which were about 690 tweets. From these results, data preprocessing was carried out. In

Table 1 is the result of the preprocessing process. This process is carried out using the Python programmer and the libraries provided.

Table I. Results of the Preprocessing Process

<i>Transform, steam</i>	<i>Tokenizing</i>	<i>Token Filter</i>
setahuku yg namanya kawasan lindung seperti taman nasional, gak boleh ada perubahan, khususnya di zona inti. karena memang fungsinya demi kelestarian plasma nufatn yg ada di dalamnya. dan rinca setahuku masuk zona itu, selain pulau komodo yang panjang itu. savekomodo	"dalamnya", "yg", "zona", "savekomodo", "nufatn", "rinca", "nasional", "lindung", "pulau", "fungsinya", "kelestarian", "plasma", "gak", "masuk", "kawasan", "taman", "namanya", "perubahan", "inti", "komodo", "setahuku"	"zona", "savekomodo", "nufatn", "rinca", "nasional", "lindung", "pulau", "kelestarian", "plasma", "kawasan", "taman", "namanya", "perubahan", "inti", "komodo",

If the data is clean, then proceed with modeling the process to find out its accuracy. The accuracy process is carried out by means of text classification, which includes two stages, namely training data and testing data. Training data is a classification process on data that has a known label. Meanwhile, data testing is a classification process whose label is unknown.

Figure 3 and Figure 4 show the main process of finding accuracy using the Naïve Bayes method and the Decision Tree using the Rapid Miner tool.

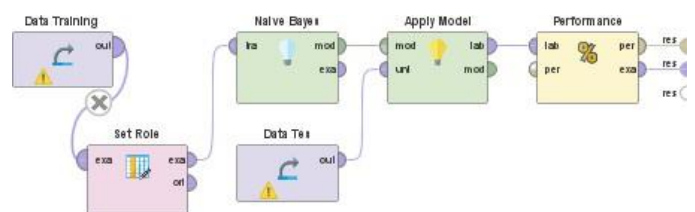


Figure 3 Naive Bayes Process on Rapid Miner

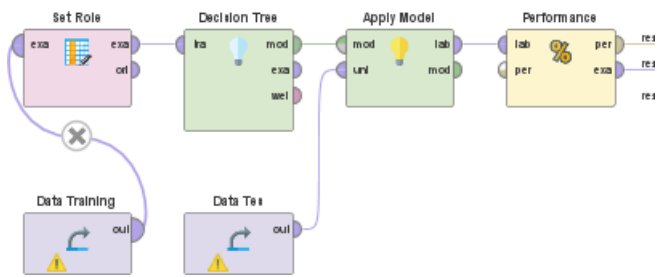


Figure 4 Decision Tree Process on Rapid Miner

Training Data and Test Data using the CSV input operator, which contains data that has been prepared for data processing. The set role is an attribute that functions to select label, here the role used is 'status', which contains positive and negative classification values. Naïve Bayes and Decision Tree are modeling operators whose function is to calculate the weight of the text in the training data. Apply model is a scoring operator that aims to make comparisons and assign values to the test data. The last stage in the Rapid Miner process is performance, namely the validation data operator, whose function is to calculate and summarize the results of the calculations carried out in the previous stage, namely apply the model.

The accuracy results obtained from Naïve Bayes and the Decision Tree can be seen in Table II and Table III. It is known that the Naïve Bayes method has higher total accuracy than the Decision Tree method.

Table IIIII. Accuracy Results

Method	Accuracy
Naïve Bayes	75%
Decision Tree	69%

Table III. Final Value of Classification Opinion Test Data using Naïve Bayes Method and Decision Tree

No	Category	Analysis	NB	DT
1	Community support of the construction of Jurassic Park	31%	6%	0%
2	Community rejection of the construction of Jurassic Park	69%	94%	100%

V. CONCLUSION

Based on the results of the analysis that has been carried out, it can be concluded that the sentiment analysis on public opinion whose data is taken from Twitter in October 2020-January 2021 regarding the support and rejection of the development of Jurassic Park Komodo using the Naïve Bayes and Decision Tree methods, obtained predictions of up to 94% and 100% in Negative opinion or show rejection of making Jurassic Park premium tourism and the remaining 6% support the government's decision to develop Jurassic Park Komodo premium tourism. The probability of positive and negative sentiments using the Naïve Bayes algorithm is

supported with an accuracy rate of up to 75%, and the Decision Tree method the accuracy rate reaches 69%. For further research, more and more quality data is needed so that the level of accuracy can be higher.

REFERENCES

- [1] "‘Jurassic Park Komodo’ di Rinca: ‘Harap tenang’, namun truk keluar masuk, pembangunan resor dikritik tak perhatikan keselamatan reptil purba setelah muncul foto viral," *BBC News Indonesia*, 2020. [https://www.bbc.com/indonesia/indonesia-54693351#:~:text=Komodo berhadap-hadapan dengan truk,Wisata Jurassic di Pulau Rinca.%22&text=Wiratno mencatat jumlah hewan karnivora,proyek pembangunan "Jurassic Park".](https://www.bbc.com/indonesia/indonesia-54693351#:~:text=Komodo%20berhadap-hadapan%20dengan%20truk,Wisata%20Jurassic%20di%20Pulau%20Rinca.&text=Wiratno%20mencatat%20jumlah%20hewan%20karnivora,proyek%20pembangunan%20‘Jurassic%20Park’) (accessed Jan. 07, 2021).
- [2] "Mengenal Proyek ‘Jurassic Park’ di NTT," *CNN Indonesia*, 2020. <https://www.cnnindonesia.com/ekonomi/20201026142334-92-562801/mengenal-proyek-jurassic-park-di-ntt>.
- [3] L. Feiner, "Twitter reports strong user growth, but ad revenue is down 23%," *CNBC*, 2020. <https://www.cnbc.com/2020/07/23/twitter-twtw-q2-2020-earnings.html>.
- [4] F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, *Sentiment Analysis in Social Networks*. United States: Morgan Kaufmann, 2017.
- [5] V. Gousander, "Analysis of the Sentiment of Labuan Bajo Tourism Destinations on Instagram," Universitas Multimedia Nusantara, 2020.
- [6] F. A. Pozzi and E. Fersini, *Sentiment Analysis in Social Networks*. United States: Todd Green, 2017.
- [7] D. S. Rajput, R. S. Thakur, and S. M. Basha, *Sentiment Analysis and Knowledge Discovery in Contemporary Business*. United States of America: IGI Global, 2018.
- [8] P. Porntrakoon and C. Moemeng, "Thai Sentiment Analysis for Consumer’s Review in Multiple Dimensions Using Sentiment Compensation Technique (SenseComp)," in *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Jul. 2018, pp. 25–28, doi: 10.1109/ECTICon.2018.8619892.
- [9] M. Wongkar and A. Angdresy, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, Oct. 2019, pp. 1–5, doi: 10.1109/ICIC47613.2019.8985884.
- [10] G. Chakraborty, M. Pagolu, and S. Garla, *Text Mining And Analysis Practical Methods, Examples, And Case Studies Using SAS*. USA: SAS Institute Inc, 2013.
- [11] B. Kamiński, M. Jakubczyk, and P. Szufel, "A framework for sensitivity analysis of decision trees," *Cent. Eur. J. Oper. Res.*, vol. 26, no. 1, pp. 135–159, 2018, doi: 10.1007/s10100-017-0479-6.
- [12] P. Chandrasekar and K. Qian, "The Impact of Data Preprocessing on the Performance of a Naive Bayes Classifier," in *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, Jun. 2016, vol. 2, pp. 618–619, doi: 10.1109/COMPSAC.2016.205.