

Web-Based System for Bali Tourism Sentiment Analysis during The Covid-19 Pandemic using Django Web Framework and Naive Bayes Method

Gede Aditra Pradnyana^{1,*} I Gede Mahendra Darmawiguna¹

¹ *Information Systems Study Program, Universtas Pendidikan Ganesha, Singaraja, Indonesia*

*Corresponding author. Email: gede.aditra@undiksha.ac.id

ABSTRACT

The province of Bali, which relies heavily on the tourism sector, is certainly affected by the Covid-19 pandemic. The various policies taken by the Bali Provincial Government in order to break the chain of the spread of Covid-19 always experience pros and cons in the community. The data from the analysis of public opinion on social media can be used as feedback to the government. The results of the analysis of social media data can be used as consideration for formulating more targeted policies. Analyzing public opinion on social media is not an easy thing. The challenges of processing opinion data from social media include the large amount of data that requires high costs such as time and effort if you have to analyze it manually one by one. The problem is complicated by dirty opinion data because it does not use standard language. In this study, a web-based system was developed to analyze Bali tourism sentiment using the Naïve Bayes method. The web-based system was developed with the Django web framework with a waterfall software development model. Based on performance testing, the Naive Bayes method can perform sentiment classification very well, which is indicated by an accuracy value of 81,70% and an f-measure value of 82,02%. From the results of sentiment analysis related to the topic of Bali tourism during the COVID-19 pandemic, the classification results obtained 65.67% of tweet data with positive sentiment and 34.33% of tweet data with negative sentiment from a total of 2377 tweet data collected in the period March 1, 2020 to March 1 2021.

Keywords: *Bali Tourism, Sentiment Analysis, Covid-19, Naïve Bayes, Django.*

1. INTRODUCTION

The World Health Organization (WHO) designated Covid-19 as a global pandemic in March 2020. The rapid spread of the virus through physical contact forced all countries to implement social distancing and physical distancing to reduce interactions. The Covid-19 pandemic certainly has an impact on various sectors and communities in Indonesia. The province of Bali, which relies heavily on the tourism sector, is certainly affected. Bali experienced a drastic decline in the level of tourist arrivals and various businesses in the tourism sector went bankrupt. Various policies taken by the Bali Provincial Government in order to break the chain of the spread of Covid-19 often experience pros and cons in the community.

Social media is one of the media used by the public to express their thoughts, opinions, opinions, and complaints about various things in this pandemic situation. The data from the analysis of public opinion

on social media can be used as feedback to the government. The results of this data analysis can be taken into consideration to formulate more targeted policies [1]. Analyzing public opinion on social media is not easy. Challenges in processing opinion data on social media include large amounts of data so that it requires high costs such as time and effort if analyzing it manually one by one [2], [3]. Moreover, coupled with the large amount of data and dirty opinion data because it does not use standard language [1].

Using a technique known as sentiment analysis, a data set can be mined for information about people's attitudes toward a certain issue (such as businesses or products) [4]–[6]. Several machine learning algorithm are often used for this sentiment analysis problem, including Support Vector Machine (SVM), Decision Tree, Maximum Entropy Classification, and Naïve Bayes (NB) [7], [8]. In various cases of sentiment analysis, the nave Bayes method is the method most often used. Naive Bayes is a simple method that was

developed based on Bayes' rules by looking at the existing conditions and the opportunities for each condition [9].

Research related to sentiment analysis was carried out by Sunardi, et al., (2018) who applied the Naïve Bayes Classifier method for sentiment analysis of lecturers' open questionnaires [4]. The apse Probability and the Bayesian theorem are used in the Bayes classifier method. Students' responses to a questionnaire satisfaction form are analyzed using this method to obtain the desired automatic classification. The findings of this study are utilized to classify student questionnaire data so that the data may be easily interpreted.. This method classification has a 75% precision, 75% recall and 80% accuracy. Research related to the analysis of public sentiment during the Covid-19 pandemic was also carried out by Samsir, et al., (2021). This study analyzes sentiment on Twitter social media related to public opinion about online learning during the pandemic using the Naïve Bayes algorithm. The findings show that online learning has 30% positive sentiment, 69% negative sentiment, and 1% neutral in that period. The high negative sentiment is generated due to public dissatisfaction with online learning. Several tweets expressed disappointment with the words 'stressed' and 'lazy' as words that have a high frequency in conversation [10].

Based on several related studies that have been done before, the method used in this study is the Naïve Bayes method. The sentiment analysis system was developed in the form of a web-based system using the Django framework. The web-based system is expected to facilitate system use, flexibility, maintenance, and system updates in the future. Full-stack framework for Python web applications, Django is a full-stack framework. Additionally, Django is an advanced python framework that can be used to swiftly build applications with a clean and pragmatic architecture. In terms of application development, Django can make it easier, faster, and less code-intensive [11].

2. METHOD

The processes that occur in the system when carrying out the sentiment analysis process for Bali tourism during the Covid-19 pandemic using the Naïve Bayes method and Query Expansion Ranking can be described in a flow chart as shown in Figure 1.

Figure 1 illustrates the processes that must be taken in order to accurately determine sentiment on social media platforms. The performance of the Naive Bayes algorithm in classifying sentiments is evaluated using a confusion matrix. There are two indicators that are used as a reference for evaluating the quality of method performance in the confusion matrix, namely accuracy and F-measure. Furthermore, the Naive Bayes method

will be validated using K-Fold Cross Validation to provide an overview of the variation in the results of the Nave Bayes method to avoid overfitting.

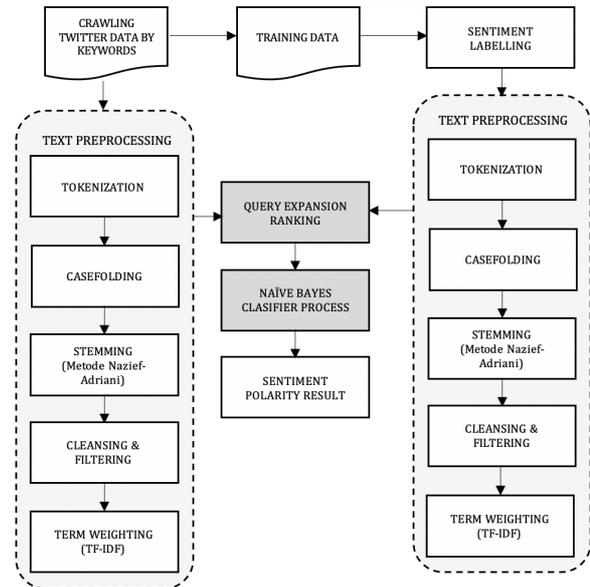


Figure 1 Flow Chart of Bali Tourism Sentiment Analysis Using the Naïve Bayes Method

3. RESULTS AND DISCUSSION

The source of sentiment data in this study is Twitter. The developed system retrieves tweet data on Twitter social media based on the specified keyword input. Sentiment data extraction from Twitter in this study used a web scraping technique. The Twint library in the Python programming language is used for this scraping. Twint uses the Twitter search facility to scrape data based on keywords, hashtags, or specific user names. The web-based system was developed using the django web framework.

After the data has been collected, the data labeling process is carried out. The labeling process is carried out to determine whether the tweet is included in the positive class which contains praise, suggestions, input, and a reflection of positive emotions such as satisfied, happy, and happy. While the negative class contains complaints, satire sentences, criticisms, and reflections of negative emotions such as anger, upset, and disappointment. In this study, the keywords 'tourism in Bali', 'covid19 Bali', 'tourism covid19', 'government of Bali' and #bali return are used with the tweet data taken from August 1, 2021 to September 1, 2021. From the crawling process the data is obtained. 1133 Data, which after labeling is divided into 692 positive tweets and 441negative tweets.

Before the classification process is carried out using the Naïve Bayes method, sentiment data must go

average accuracy value of 81.70% and the F-Measure of 82.02%. This value is strongly influenced by the quality of the data used. In this testing process, there were a lot of comments that used abbreviations, non-standard words, and also the use of emojis.

Table 1. Naive Bayes Performance Evaluation Results

Fold	Accuracy (%)	F-Measure (%)
1	83,22	81,67
2	80,96	81,21
3	77,34	80,06
4	84,82	82,63
5	82,17	84,51
Average	81,70	82,02

The next stage is the implementation process on the original data using twitter data from March 1, 2020 to March 1, 2021, as 1 year of the Covid-19 pandemic hitting Indonesia and Bali in particular. From the results of sentiment analysis, positive sentiment results were obtained as much as 65.67% and negative sentiment as much as 34.33% of the total 2377 tweet data collected. From the classification data above, it can be seen that most of the comments are positive comments such as related to the vaccination program in Bali and support for health workers. The negative comments that often appear are related to the scarcity of masks, the closure of various tourist attractions and hotels in Bali, and the loss of jobs. The five words that most often appear in twitter users' sentiments related to Bali tourism are vaccines, restrictions, masks, government, close.

4. CONCLUSION

In this study, a web-based system has been successfully developed to analyze sentiment related to Bali tourism during the Covid-19 pandemic. The system was developed using the Django web framework and using Naïve Bayes for the classification process. The accuracy of the Naïve Bayes method in conducting the classification process is 81.70% with an F-Measure value of 82.02%. From the results of sentiment analysis related to the topic of Bali tourism during the COVID-19 pandemic, the classification results obtained 65.67% of tweet data with positive sentiment and 34.33% of tweet data with negative sentiment from a total of 2377 tweet data collected in the period March 1, 2020 to March 1 2021. Future research can focus on the problem of handling abbreviations or typos in comment data and also analysis of sentiment in English considering that Bali is the center of world attention, not only Indonesia.

REFERENCES

- [1] P. H. Prastyo, A. S. Sumi, A. W. Dian, dan A. E. Permanasari, "Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 2, hal. 112, 2020.
- [2] J. Riany, M. Fajar, dan M. P. Lukman, "Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor," *J. Sisfo*, vol. 06, no. 01, hal. 147–156, 2016.
- [3] G. A. Pradnyana, I. G. M. Darmawiguna, D. K. S. Suditresna Jaya, dan A. Sasmita, "Performance analysis of support vector machines with polynomial kernel for sentiment polarity identification: A case study in lecturer's performance questionnaire," *J. Phys. Conf. Ser.*, vol. 1810, no. 1, 2021.
- [4] Sunardi, A. Fadlil, dan Suprianto, "Analisis sentimen menggunakan metode naïve bayes classifier pada angket mahasiswa," *SAINTEKBU J. Sains dan Teknol.*, vol. 10, no. 02, hal. 1–9, 2018.
- [5] W. Wilana, S. H. Hanasbey, M. R. Awinero, J. V. A. Modouw, dan M. N. N. Sitokdana, "Analisis Sentimen Terhadap Opini Masyarakat Indonesia Mengenai Bukalapak," *Semin. Nas. Teknol. Inf. dan Multimed. 2018 Univ. AMIKOM Yogyakarta*, hal. 1–6, 2018.
- [6] B. Haryanto, Y. Ruldeviyani, F. Rohman, J. D. T.N., R. Magdalena, dan Y. F. Muhamad, "Facebook Analysis of Community Sentiment on 2019 Indonesian Presidential Candidates from Facebook Opinion Data," *Procedia Comput. Sci.*, vol. 161, hal. 715–722, 2019.
- [7] S. Adi, M. Wulandari, A. K. Mardiana, dan A. Muzakki, "Survei: Topik Dan Tren Analisis Sentimen Pada Media Online," *Semin. Nas. Teknol. dan Multimed.*, hal. 55–60, 2018.
- [8] C. Song, X. K. Wang, P. fei Cheng, J. qiang Wang, dan L. Li, "SACPC: A framework based on probabilistic linguistic terms for short text sentiment analysis," *Knowledge-Based Syst.*, no. xxxx, hal. 105572, 2020.
- [9] M. A. Fauzi dan S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking Image Processing View project Smart Wheelchair View project," *Researchgate.Net*, vol. 2, no. 8, hal. 2766–2770, 2018.
- [10] Samsir, U. Verawardina, F. Edi, dan R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J.*

Media Inform. Budidarma, vol. 5, no. 1, hal. 157–163, 2021.

- [11] J. Plekhanova, “Evaluating web development frameworks: Django, Ruby on Rails and CakePHP,” *Inst. Bus. Inf. Technol.*, no. September, hal. 20, 2009.
- [12] M. S. Hadna, P. I. Santosa, dan W. W. Winarno, “Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter,” *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. Sentika, hal. 57–64, 2016.