Sentiment analysis of public opinions on the welfare of honorary educators using Naive Bayes

D R Lazuardi¹, T A Munandar¹, H Harsiti², Z Mutaqin² and R N Hays¹

¹ Informatics Dept., Faculty of Information Technology, Universitas Serang Raya, Serang, Indonesia
² Information Systems Dept., Faculty of Information Technology, Universitas Serang Raya, Serang, Indonesia

*aafatwa25@gmail.com

Abstract. The development of the internet in Indonesia is very rapid, this is marked by the number of emerging social media. One of the most popular social media is Twitter. Many Twitter users who tweet their expression and opinion on a product, service, political issues or things that are trending. Even the current government is not spared from public comment on social media, one of which is an honorary teacher at an elementary school in Pandeglang Regency, who lives next to his school toilet. This case was widely discussed by almost all Indonesians through offline and online media, especially on social media. The impact of this case is the difference in response in the form of public sentiment towards public services, especially in improving the welfare of educators. Various sentiments of Netizens (internet users) appear on social media ranging from praise, criticism, suggestions, satire and even expressions of hatred towards the Pandeglang Government. This research was conducted to determine netizens’ sentiments towards the Pandeglang Regency government based on the case of honorary teacher news in Pandeglang Regency who live in school toilets. The approach used is Naive Bayes. The results of sentiment analysis showed that there are 16 Twitter posts grouped in neutral class, 19 posts in positive class and the remaining 16 posts are in the negative class. The level of accuracy generated from the sentiment analysis model formed is 88.24%.

1. Introduction

The development of information technology which has increased has an impact on the exchange of existing information and communication. Social media is a new set of communication and collaboration tools that enable many types of interactions that were not previously available to ordinary people [1]. Many social media users post their opinions and opinions on a product, service, political issue or things that are being viral. The current government is also not spared from public comment on social media, one of which is Pandeglang District Government. The current hot topic of social media users is the news of Honorary Teachers at the Primary School level in the Pandeglang Regency who live next to the toilet of the school where they teach. This case is widely discussed by almost all Indonesian people in offline and online media including social media, of course, this invites sentiments of Netizens (internet users) on social media ranging from praise, criticism, suggestions, satire and even expressions of hatred against the Pandeglang Government. The impact arising from this sentiment is the level of public confidence that changes in public services provided by the government. This research was conducted to determine
the netizens' sentiments due to the reporting of the honorary teacher's case of public trust in the local government, using a sentiment analysis approach.

Sentiment analysis is indeed not new in the world of computing. Sentiment analysis is also commonly known as opinion mining which refers more to natural language processing (NLP) counselling, analytical text and computational language with the main objective to identify and extract subjective information from various sources [2]. In this research, the level of public trust, especially in Pandeglang Regency, was analyzed based on opinions circulating in various social media, one of which was Twitter and Facebook. Opinions were previously collected and then analyzed to obtain preliminary data on the level of public trust, in the form of positive and negative sentiment classes. One example of opinion which contains these two classes of sentiments is as follows:

- Positive sentiments in subjective sentences: "The Sub-district will immediately repair the old house which collapsed, Allhmdullilah " . In the word "improve", there is a positive value.
- Negative sentiment in subjective sentences: "The local government officers are all sleeping yaa ... Poor citizens (teachers) to live in the toilet" sentence that is defined in "local government officers are all sleeping" expresses negative sentiment towards the government, so the assessment of the word "sleep" forms a negative.

Many methods can be used to determine the sentiment of analysis, one of which is using Naive Bayes. The use of Naive Bayes is more on its ability to be able to classify even with the availability of data that is not too much and works based on a simple probabilistic concept with the assumption of strong independence (dependence) [3,4]. Thus the use of the Bayes approach is expected to be able to do a better opinion classification based on the tendency of the highest probability value of a class from the given text input. The research conducted is expected to be able to map public sentiment information on public service efforts provided by the government, so that it can be used as an input for improving public services to the public.

2. Theoretical basis

2.1. Sentiment analysis
Sentiment analysis is a field of text classification that allows to determine people's opinions and attitudes on different products, services, and topics [5]. Meanwhile, according to research from Abo et al. Sentiment analysis (SA) techniques have become necessary for supporting knowledge discovery [6]. The aim of sentiment analysis is to define automatic tools able to extract subjective information from texts in natural language, such as opinions and sentiments, so as to create structured and actionable knowledge to be used by either a decision support system or decision-maker [7]. As the network comments are unstructured text information and have domain knowledge, therefore, how to carry on the thorough excavation to find sentiment tendency and its strength, text mining puts forward new challenges [8]. It can be identity opinions tuples, which consist of a target term, target attribute (aspect), and target sentiment [9].

2.2. Public service
According to the Government Website, at the Ministry of Home Affairs that public service is a public service, and the definition of public service is Public services are all forms of services, both in the form of public goods and public services which in principle are the responsibility and are carried out by government agencies at the central and regional levels. From the description above, public services can be interpreted as activities of providing services both by the government and private parties in certain ways that require sensitivity and interpersonal, thus creating a satisfaction of goods and services.

2.3. Naïve Bayes
Naïve Bayes is a classification method that can predict a class based on its probability value so that it can produce decisions based on learning data [10]. Naïve Bayes is a probabilistic classifier based on the
Bayes theorem, considering Naive (Strong) independence assumption [2]. Theoretically, Naïve Bayes classifier often works much better and has a minimum error rate compared to other classification methods even though it uses a "naive" design and with simplified assumptions [11]. Naïve Bayes can reduce errors in the classification of documents with an average value of 27%, even reaching 50% of experiments using multivariate Bernoulli.

3. Research method
The determination method used consists of data collection, data sharing, data labelling, data pre-processing, and sentiment calculation with naïve Bayes as shown in figure 1.

![Figure 1. Research stages.](image1)

Data is taken from social media Twitter.com in the form of tweets and stored in the form of CSV files. Labelling is done to distinguish between positive and negative tweets. The data is then divided into two new datasets namely training data and testing data. The next step is text pre-processing, such as tokening, stop-word removal, filtering and stemming. After the data goes through the pre-processing stage, then the sentiment classification stage uses the Naïve Bayes method and then tests the accuracy. After the sentiment classification process, the next step is the interpretation of the results of sentiment analysis.

4. Results and discussion
In this research the data obtained were divided into two types of datasets, the first training data was 90% training data and 10% testing data. Training data is used to build the sentiment analysis model while testing data is used to test the developed model. The analytical aid used is Rapid Miner. The analysis shows that there are 38 opinions from Twitter posts that indicate opinions with neutral target classes, 28 posts categorized into positive target classes and the remaining 68 posts indicate negative target classes. Figure 2 shows statistical information for the testing data used in this research.

![Figure 2. Statistical information on data testing.](image2)

To find out the level of analysis of internet user sentiment towards public services by the Pandeglang Regency government, the sentiment analysis model was built based on the training data prepared. The model that has been built is then tested on data testing to determine the sentiment of internet users on government public services. In this research the analyzed text is Indonesian text or terminology in Indonesian vocabulary. The results of sentiment testing data analysis of training data are shown in table 1.
Table 1. Results of sentiment analysis with Naïve Bayes.

| No | Sentiment | Sentiment Prediction | Confident (neutral) | Confident (negative) | Confident (positive) | Comment (in Bahasa) |
|----|-----------|----------------------|---------------------|----------------------|----------------------|---------------------|
| 1  | negative  | negative             | 0.0072              | 0.9856               | 0.0071               | Edan, klo iya bener seorang pemimpin ngomongnya kek gt... Sama skali gk punya empati pd rakyatnya...Ngerakenn urang kulon we sira |
| 2  | negative  | negative             | 0.0072              | 0.9856               | 0.0071               | Jirr ngakak gw!! Lo pikir dia setan apa, senangnya tinggal di WC.. lo kali yg setan, secara lo shg pemimpinnya disana |
| 3  | negative  | negative             | 0.0072              | 0.9856               | 0.0071               | Bupatinya berhati binatang |
| 4  | neutral   | neutral              | 0.9853              | 0.0073               | 0.0072               | Astaghfirullahh... Kok bisa2nya jawab spt itu. Mana ada orang yg nyaman tinggal di WC bu?? Ingin rasanya berkata kasar |
| 5  | negative  | negative             | 0.0072              | 0.9856               | 0.0071               | Ada ya perempuan yg ga punya empati kek gini .. Kayanya ibu bupati ini hirs ngalamin dl biar tau rasanya. Mau bu?? |
| 6  | neutral   | neutral              | 0.9853              | 0.0073               | 0.0072               | Miris !!! Nasib guru Honorer di negeri Plongo Land |
| 7  | negative  | negative             | 0.0072              | 0.9856               | 0.0071               | Apa ga sebaiknya but bupati kantornya pindah ke WC Aja ? |
| 8  | positive  | positive             | 0.0073              | 0.0074               | 0.9851               | Ikhlasan pengabdianannya karna sesungguhnya Alloh pemberi rezeki yang sesungguhnya |
| 9  | positive  | positive             | 0.0073              | 0.0074               | 0.9852               | Dana pendidikan aja d pabgkas sampe 1T, jadi bu tetap senang ya menciptakan generasi bangsa yang hebat...sukses untuk ibu guru honorernya |
| 10 | positive  | positive             | 0.0073              | 0.0074               | 0.9852               | Guru tetap guru bukan status, bangsa mau maju muliakan guru |

Figure 3. Statistical information data testing results of sentiment analysis.

The results of sentiment analysis showed that there were 16 Twitter posts that indicated the target class was neutral, 19 posts indicated a positive class and the remaining 16 posts indicated a negative class. Figure 3 shows statistical information on the results of sentiment analysis using Naïve Bayes.
Figure 4. Accuracy of sentiment measurement results with Naïve Bayes.

Figure 4 shows true negative results get 72.73% precision, this is due to negative false prediction of 18.18% on neutral and 9.9% on positive. While neutral prediction and positive prediction get 100% precision, this is because there is no false result in prediction. So that the recall results in each class get 100% true negative, 75% true neutral, and 89.47% true positive. Therefore, the results on the testing data that do accuracy testing get as much as 88.24%. The results that have been classified, then taken action by the local government official Pandeglang according to the results of the accuracy testing.

The use of sections to divide the text of the paper is optional and left as a decision for the author. Where the author wishes to divide the paper into sections the formatting shown in table 2 should be used.

5. Conclusion
The research carried out is capable of mapping public opinion in cyberspace with the sentiment analysis approach with the Naïve Bayes method. The results of sentiment analysis show there are 16 Twitter posts that indicate the target class is neutral, 19 posts indicate a positive class and the remaining 16 posts indicate a negative class. The level of accuracy generated from the sentiment analysis model formed is 88.24%. Thus it can be said that basically the people of Pandeglang Regency still have a fairly good level of trust in their government despite the outstanding cases being discussed in this research. This can be seen from the number of negative opinions that are still below positive opinions even negative opinions having the same amount as neutral opinions.

References
[1] Brogan C 2010 Social media 101: Tactics and tips to develop your business online (John Wiley & Sons)
[2] Lopamudra D and Sanjay C 2016 Sentiment Analysis of Review Datasets using Naïve Bayes and K-NN Classifier Journal MESC 4 54–62
[3] Saleh 2015 Classification of Symptoms of Depression in Humans with the Naïve Bayes Method Using Java (Yogyakarta) p 9
[4] Arfiana 2014 Classification of Four-Wheel Vehicles Using the Naïve Bayes Method (Bandung) p 1
[5] Buket E, Ozlem A, Kiling D, and Erhasin M 2019 A Hybrid Sentiment Analysis Method for Turkish Journal TUBITAK 27 1780 - 1793
[6] Abo M E M, Raj R G, Qazi A and Zakari A 2019 Sentiment Analysis for Arabic in Social Media Network: A Systematic Mapping Study arXiv preprint arXiv:1911.05483
[7] Aberto P F, Messina F E and Bing E L 2017 Sentiment Analysis in Social Network (Cambridge) pp 91-111
[8] Tan C, Chen G and Song Q 2015 Sentiment Mining Analysis Based on Network Comments Metallurgical and Mining Industry (5) 261
[9] Boudad N, Faizi R, Thami R O H and Chiheb R 2018 Sentiment analysis in Arabic: A review of the literature Ain Shams Engineering Journal 9(4) 2479-2490
[10] Aulia S and Much M A 2015 use of Naïve Bayes to respond to emotions from Indonesian sentences 4(4) 148 - 156
[11] Rennie J D, Shih L, Teevan J and Karger D R 2003 Tackling the poor assumptions of naive bayes text classifiers In Proceedings of the 20th international conference on machine learning (ICML-03) pp 616-623