

Analysis Sentiment to Religious Based Issues Netizen comments on President Jokowi 's Instagram using the Naive Bayes Classifier Method

Lingga Kurnia Ramadhani ^{a,1,*}, Bajeng Nurul Widyaningrum ^{b,2}

^a Digital Business, Faculty of Science and Technology, IVET University, Indonesia

^b Medical Records and Health Information, Bina Trada Polytechnic, Indonesia

¹ linggakurnia@ivet.ac.id *, ² bnwidyani@gmail.com

*corresponding author

ARTICLE INFO

Article history:
Accepted

Keywords:
Data Mining
Naïve Bayes Classifier
Sentiment Analysis
Instagram

ABSTRACT

The presence of Instagram is one of the most popular social media platforms. Make people make a habit of posting photo and video content. In this way, Instagram can become a source of public opinion and sentiment data, which data can be used for social studies. Social studies are still a crucial issue for countries today, including Indonesia, is a matter of religion. Apart from problems regarding religion, public sentiment can be researched from Instagram. The sentiment dataset related to religion is taken from public comments from President Joko Widodo's content regarding Merry Christmas Greetings and Visits to Churches. The analysis system will use sentiment analysis using a method from machine learning, namely the naive Bayes classifier to determine positive & negative sentiments. The comment dataset is 2,600 in which the positive and negative sentiment class datasets are determined manually by experts in the field of linguistics or communication totaling 1,309 to be used as training data. the results of testing the Naïve Bayes classifier training data with a total of 1309 obtained an accuracy prediction of 97.63% and test data where positive & negative sentiment classes had not been determined before hand amounted to 1291 in sentiment analysis obtained positive sentiment 70.56% & negative sentiment 29.44%.

Copyright © 2024 by the Authors

I. Introduction

Instagram is social media based photos and videos. Instagram originates from the words instant and telegram [1]. Instant said or install is the word from Polaroid cameras were very famous in ancient times. The camera is also named with camera instant. The word telegram refers to a working tool for send information to others with fast. Instagram is a social media with user most in the world. Posts and comments from netizens milling around on Instagram own traits that don't structured and owned lots of noise [2]. Although so, that data can processed with analysis sentiment.

Analysis sentiment is the process of understanding, extracting and processing textual data in a way automatic for get information the sentiment contained in something sentence. Analysis sentiment considered as field science that integrates natural language processing, computing linguistics, and analysis purposeful text for identify opinion about something products delivered by the community [3]. In terms of This popular method for finish analysis sentiment is the Naïve Bayes Classifier.

Naïve Bayes Classifier is one of them algorithm used for classification text as well as use Machine Learning method with calculation probability and statistics proposed by Thomas Bayes [4]. This method can classify opinion positive and negative.



Meanwhile, natural language processing is application knowledge computers, esp linguistics computational (computational linguistics), for study interaction between computer with human (natural) language. NLP works For solve problem For understand Language experience human, with all rule grammatical and semantic, and changing Language the become formal representation that can be processed by a computer.

Study This will study religious based issues netizen comments in post Photo President Joko Widodo is related saying Merry Christmas and visiting to church in frame day Christmas based on class positive and negative For determine sentiment use Naïve Bayes Classifier method.

II. Method

Research methods This consists of five stages [5] . Stages the can see in Figure 1.

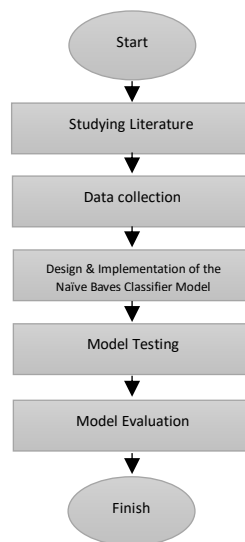


Fig. 1. Stages Study

Following description of each of the steps involved in Figure 1:

- Study Literature
Literature study carried out with learn references and literature related the problems studied.
- Data collection
Data collection techniques used in study This is crawling. The crawling process is carried out with use application based on the Phantombuster website. Data taken is comment post President Joko Widodo is related Photo Merry Christmas greetings and visits to Church in frame day Christmas.
- Design and Implementation of the Naïve Bayes Classifier Model [6]
Designed model will classify sentiment become class positive and negative. Stages next transform raw data into easy data understandable called with Preprocessing stage. By The entire preprocessing utilizes the natural language toolkit (NLTK) library which is part of natural language processing (NLP) Preprocessing steps can see in Figure 2.

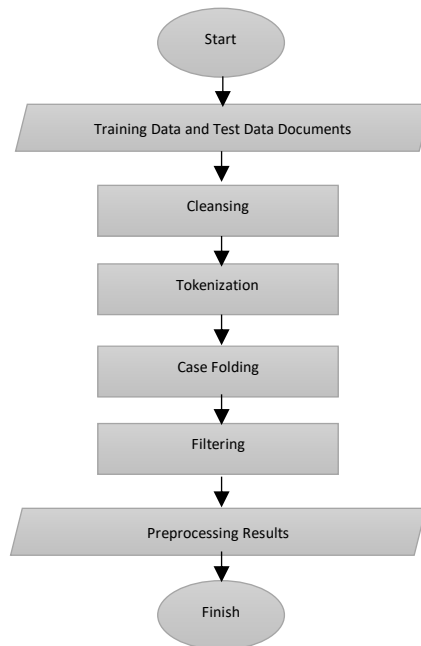


Fig. 2. Preprocessing steps

- Cleansing, For clean text from links, symbols, numbers, and signs read.
- Tokenization For separate sentence into words.
- Case folding, for change text becomes lower case.
- Filtering, removing unnecessary stop words own influence in analysis.

After preprocessing stage is carried out stages Research word weighting This use Term Frequency Inverse Document Frequency (TF-IDF) method. TF-IDF is A method which is integration between term frequency (TF), and inverse document frequency (IDF). Term Frequency is calculated use Equality with the i th term frequency is frequency the appearance of the i th term in document to $-j$. Inverse Document Frequency (IDF) is logarithm from ratio amount all over document in corpus with amount documents that contain the term in question as written in a way mathematically in Eq. Value obtained with multiply both of which are formulated in Eq.

$$tf_i = \frac{freq_i(d_j)}{\sum_{i=1}^k freq_i(d_j)} \quad (1)$$

$$idf_i = \log \frac{|D|}{|\{d:t_i \in d\}|} \quad (2)$$

$$(tf - idf)_{ij} = tf_i(d_j) * idf_i \quad (3)$$

Function the TF-IDF method is For look for representation mark from each document from something where is the training data set ? later formed something vector Between documents with the words (documents with terms) later For similarity between document with the cluster will determined by a prototype vector which is also called the cluster centroid.

- Naïve Bayes Classifier

After stage word weighting, next done stage classification For grouping sentiments in comments Instagram about internal religious issues post President Joko Widodo became sentiment negativity and sentiment positive. Method used For classification sentiment on

research This is the Naïve Bayes Classifier [7] . The Naïve Bayes method is based on an algorithm with technique classification that has been through efficiency and effectiveness tests in a database with large amounts of data, Naïve Bayes has level high accuracy and efficiency. Besides that, Naïve Bayes performance has time short classifications that speed up system processes analysis sentiment.

Naïve Bayes with draft base theorem conditional Bayes can see in the equation

$$P(B) = P(A) P(A) \quad (4)$$

Opportunity event A conditional on B being determined from probability A and probability B conditional A. equation This developed become equality following:

$$P(B) = \frac{P(A) P(A)}{P(B)} \quad (5)$$

Application Bayes' theorem is based on the equation above can changed to inform equality following:

$$P(D) = \frac{P(Ci) P(Ci)}{P(D)} \quad (6)$$

The Naïve Bayes method assumes every variable stand free One each other and not There is connection with other variables, so A document will be considered as a collection of words that make up document that, and no notice order appearance of the word in the document. Calculation probability can be considered as results multiplication from probability appearance of words in the document.

$$P(Ci) = \frac{fD(Ci)}{|D|} \quad (7)$$

With $P(Ci)$ is probability from something category documents, meanwhile $fD(Ci)$ is frequency documents that have categories Ci and $|D|$ is amount all over document train.

$$P(Ci) = \frac{fD(Ci) + 1}{f(Ci) + |W|} \quad (8)$$

With $P(Ci)$ is probability emergence Wkj on a document with category class Ci . Wkj is frequency of the k th word in each category. While W is the number of words in the test document and $f(Ci)$ is frequency document categorized Ci class.

Equality This there is something addition the number 1 in the numerator For anticipate If there is a word in the test document that is valuable zero (0) because No contained in the document train.

- Model Evaluation

next step is do evaluation_ aim For know to what extent accuracy results data classification using Naive Bayes method on training data and knowing results analysis of positive and negative sentiment on test data.

III. Results and Discussion

Relevant data retrieved from netizen comments on his Instagram President Joko Widodo is related internal religious issues post Photo saying Merry Christmas and visiting to church in celebration day Christmas with 2600 comments. Data retrieved with use appropriate keywords with case. Then the data is labeled manually become positive and negative as many as 1309 as training data used in study This.

Table 1. Classification of Data Classes

Comment	Class
<i>saya ada di misa pagi itu ketika bapak berkunjung wah suatu kejutan karena baru pertama kali saya lihat pak presiden dari dekat terima kasih atas kunjungannya pak semoga pak jokowi sekeluarga selalu diberi kesehatan</i>	Positive

Comment	Class
<i>ga takut ntar dibilang murtad pak</i>	Negative
<i>presidenku mantap toleransi yes sehat selalu pak presiden tuhan memberkatimu</i>	Positive
<i>barang siapa mengucapkan selamat natal maka ia telah kafir</i>	Negative

Crawling process on netizen comment data from source Instagram via application website based ie Phantombusters in CSV format. Then enforced dividing data into training data and test data. Where 1309 comment data into training data and the rest as test data yet determined class positive and negative. Data distribution can be done seen in Table 2.

Table 2. Distribution of Training Data & Test Data

	Training Data	Test Data
Amount of data	1309	1291

These data will through The preprocessing stage is carried out in the application rapidminer 10.0. The results of each step from preprocessing stage can see in Table 3 – Table 6.

Table 3. Cleansing

Before Cleansing	After Cleansing
<i>Saya ada di misa pagi itu ketika bapak berkunjung... wah suatu kejutan karena baru pertama kali saya lihat pak presiden dari dekat.. terima kasih atas kunjungannya pak. Semoga pak @jokowi sekeluarga selalu diberi kesehatan ḍŸ, ḍŸ, ḍŸ,</i>	<i>Saya ada di misa pagi itu ketika bapak berkunjung wah suatu kejutan karena baru pertama kali saya lihat pak presiden dari dekat terima kasih atas kunjungannya pak Semoga pak jokowi sekeluarga selalu diberi kesehatan</i>
<i>Ga takut ntar dibilang murtad pak ?</i>	<i>Ga takut ntar dibilang murtad pak</i>
<i>â☐☐, ☐ḍŸ☐ Presidenku Mantap, Toleransi Yes.. Sehat selalu Pak Presiden Tuhan Memberkatimu ḍŸ™☐</i>	<i>Presidenku Mantap Toleransi Yes Sehat selalu Pak Presiden Tuhan Memberkatimu</i>
<i>Barang siapa mengucapkan selamat natal maka ia telah kafir!</i>	<i>Barang siapa mengucapkan selamat natal maka ia telah kafir</i>

Table 4. Tokenization

Before Tokenization	After Tokenization
<i>Saya ada di misa pagi itu ketika bapak berkunjung wah suatu kejutan karena baru pertama kali saya lihat pak presiden dari dekat terima kasih atas kunjungannya pak Semoga pak jokowi sekeluarga selalu diberi kesehatan</i>	<i>['Saya', 'ada', 'di', 'misa', 'pagi', 'itu', 'ketika', 'bapak', 'berkunjung', 'wah', 'suatu', 'kejutan', 'karena', 'baru', 'pertama', 'kali', 'saya', 'lihat', 'pak', 'presiden', 'dari', 'dekat', 'terima', 'kasih', 'atas', 'kunjungannya', 'pak', 'Semoga', 'pak', 'jokowi', 'sekeluarga', 'selalu', 'diberi', 'kesehatan']</i>
<i>Ga takut ntar dibilang murtad pak</i>	<i>['Ga', 'takut', 'ntar', 'dibilang', 'murtad', 'pak']</i>
<i>Presidenku Mantap Toleransi Yes Sehat selalu Pak Presiden Tuhan Memberkatimu</i>	<i>['Presidenku', 'Mantap', 'Toleransi', 'Yes', 'Sehat', 'selalu', 'Pak', 'Presiden', 'Tuhan', 'Memberkatimu']</i>

Barang siapa mengucapkan selamat natal maka ia telah kafir [‘Barang’, ‘siapa’, ‘mengucapkan’, ‘selamat’, ‘natal’, ‘maka’, ‘ia’, ‘telah’, ‘kafir’]

Table 5. Case Folding

Before Case Folding	After Case Folding
Saya ada di misa pagi itu ketika bapak berkunjung wah suatu kejutan karena baru pertama kali saya lihat pak presiden dari dekat terima kasih atas kunjungannya pak Semoga pak jokowi sekeluarga selalu diberi kesehatan	saya ada di misa pagi itu ketika bapak berkunjung wah suatu kejutan karena baru pertama kali saya lihat pak presiden dari dekat terima kasih atas kunjungannya pak semoga pak jokowi sekeluarga selalu diberi kesehatan
Ga takut ntar dibilang murtad pak	ga takut ntar dibilang murtad pak
Presidenku Mantap Toleransi Yes Sehat selalu Pak Presiden Tuhan Memberkatimu	presidenku mantap toleransi yes sehat selalu pak presiden tuhan memberkatimu
Barang siapa mengucapkan selamat natal maka ia telah kafir	barang siapa mengucapkan selamat natal maka ia telah kafir

Table 6. Filtering

Before Filtering	After Filtering
saya ada di misa pagi itu ketika bapak berkunjung wah suatu kejutan karena baru pertama kali saya lihat pak presiden dari dekat terima kasih atas kunjungannya pak semoga pak jokowi sekeluarga selalu diberi kesehatan	saya ada misa pagi ketika bapak berkunjung suatu kejutan karena baru pertama kali saya lihat pak presiden dari dekat terima kasih atas kunjungannya pak semoga pak jokowi sekeluarga selalu diberi kesehatan
ga takut ntar dibilang murtad pak	takut ntar dibilang murtad pak
presidenku mantap toleransi yes sehat selalu pak presiden tuhan memberkatimu	presidenku mantap toleransi sehat selalu pak presiden tuhan memberkatimu
barang siapa mengucapkan selamat natal maka ia telah barang siapa mengucapkan selamat natal maka telah kafir	barang siapa mengucapkan selamat natal maka telah kafir

Following is preprocessing stage of data in the application Rapidminer in Figure 3.

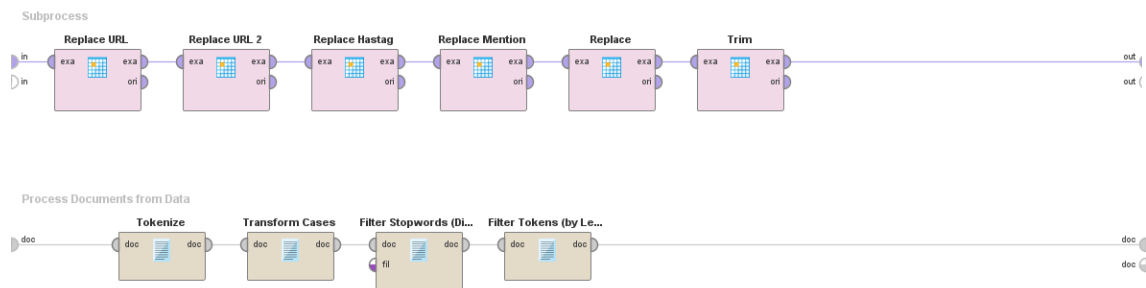


Fig. 3. Preprocessing stage

Furthermore, will done word weighting with TF-IDF for produce a naïve Bayes classifier model. The following is the word weighting process with TF-IDF in Figure 4.

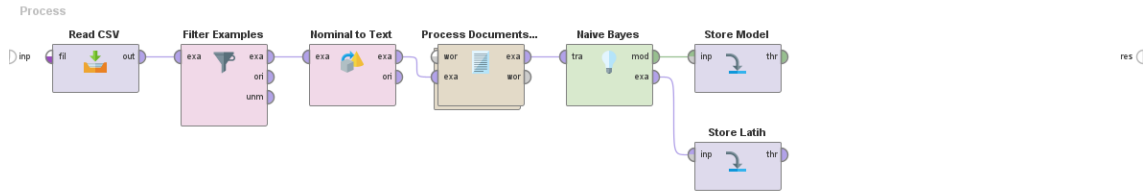


Fig. 4. Naïve Bayes Classifier Model Process

Based on the Naïve Bayes classifier model process. Distributing models for sentiment labels attribute can see in table 7.

Table 7. Distribution of Attribute Sentiment Label Models

Regular Attribute	Class	
	Positive	Negative
2083	0.687	0.313

After That done stage analysis sentiment For label data that has not There is class positive and negative using the naïve Bayes classifier model that has been built and evaluated of the model on training data. Here's the process for analysis sentiment in Figure 5.

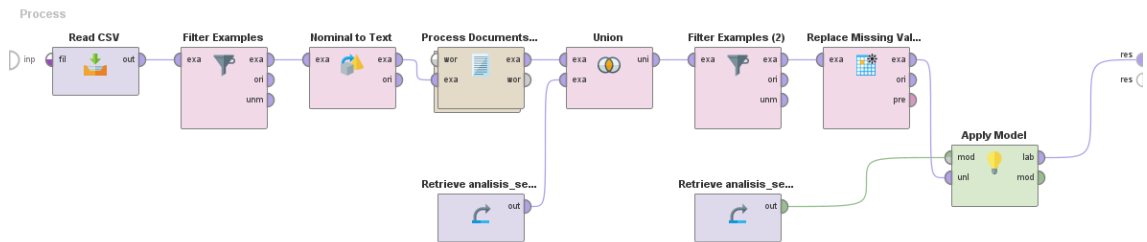


Fig. 5. Analysis Process Sentiment

Results of the analysis process sentiment from the Naïve Bayes Classifier model based on existing training data determined class positive and negative manually beforehand with number 1309 then obtained for test data viz sentiment positive 70.56% and sentiment negative 29.44%. Full results can see in Table 8.

Table 8. Sentiment Analysis Results

Number of Test data	Class	
	Positive	Negative
1291	911 (70.56%)	380 (29.44%)

Furthermore For results classification evaluation from the Naïve Bayes Classifier model based on training data obtained level accuracy 97.63%. Full results can see in Table 9.

Table 9. Classification Results of Naïve Bayes Classifier Evaluation

	True Positive	True Negative	Precision
Pred. Positive	868	0	100%
Pred. Negative	31	410	92.97%
Recall	96.55%	100%	

IV. Conclusion

Study This retrieve netizen comment data from President Joko Widodo's Instagram is related religious issues in posts saying Merry Christmas and visiting to church in celebration day Christmas. The data obtained as many as 2600 where for class datasets sentiment positive and negative determined manually by experts in linguistic field or communication totaling 1309 for used as training data. results testing training data with the number 1309 is obtained predictions accuracy Naïve Bayes Classifier method and collaborated with TF-IDF, namely 97.63%. For test data where class sentiment positive & negative Not yet determined previously totaling 1291 in analysis sentiment obtained sentiment positive 70.56% & sentiment negative 29.44%. This study produce level trend public to religious issues are positive matter This caused Because public looking that tolerance religious that's very important Because in accordance practice, please to 1 and above guard unity and wholeness between people religious.

References

- [1] V. Nainggolan, SA Randonuwu, and GJ Waleleng, "The Role of Instagram Social Media in Social Interaction Between Students at the Faculty of Social and Political Sciences, Unsrat Manado," *J. Acta Diurna*, vol. 7, no. 4, pp. 1–15, 2018, [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/actadiurnakomunikasi/article/view/22022>.
- [2] AN Rohim and AR Pratama, "Analysis of Public Sentiment on Instagram Social Media on President Joko Widodo's Performance," *AUTOMATA Journal*, vol. 3, no. 1, pp. 1–5, 2022.
- [3] A. Anggara, S. Widiono, A. Tri Hidayat, and S. Sutarman, "Analysis of Netizen Comments Sentiment on Public Official Statements on Instagram Social Media Accounts," *Int. J. Adv. Data Inf. Syst.*, vol. 3, no. 2, pp. 87–97, 2022, doi: 10.25008/ijadis.v3i2.1244.
- [4] MK Al Fath, A. Arini, and N. Hakiem, "Sentiment Analysis Of Full Day School Policy Comment Using Naïve Bayes Classifier Algorithm," *Sinkron*, vol. 5, no. 1, pp. 107–114, 2020, doi: 10.33395/sinkron.v5i1.10564.
- [5] M. Ikhsan Yusuf, N. Ransi, A. Arman, A. Tenriawaru, LO Saidi, and LS La Surimi, "Sentiment Analysis of Instagram Netizen Comments on Racism in Indonesian Football Using the Naive Bayes Method," *J. Mat. Computation and Stat.*, vol. 2, no. 3, pp. 136–143, 2022, doi: 10.33772/jmks.v2i3.10.
- [6] A. Anggara and S. Widiono, "Classification of Netizen Comments on Public Figures on Instagram Using the Naive Bayes Classification Method," *Pros. Semin. Nas. Computer Science.*, vol. 2, no. January 2019, pp. 1–10, 2019.
- [7] MB Ismiati, "Detection of Negative Comments on Instagram Using the Naive Bayes Classifier Algorithm," *Pros. Semin. Nas. Science and Technology.*, pp. 243–248, 2018.