International Journal of Informatics, Economics, Management and Science (IJIEMS)
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i1.1237
Volume 3, Issue 1, January 2024, pp. 102-107

Public Sentiment towards the Medan-Binjai Water Pipeline Excavation using the Naïve Bayes Method

Liana Tasya Turnip[1], * Hotler Manurung [2], I Gusti Prahmana [3]

[1,2,3] Information System, STMIK Kaputama, Binjai Indonesia

E-mail address :
lianatt24@gmail.com (Liana Tasya Turnip), hotlermanurungskommkom@yahoo.com (Hotler Manurung), igustiprahmana4@gmail.com (I Gusti Prahmana)

*Corresponding author : lianatt24@gmail.com

Received: July 22, 2024; **Accepted**: December 20, 2023 ; **Published**: Janaury 23, 2024

Abstract: The work on this project received sharp scrutiny from Non-Governmental Organizations (NGOs) because it was suspected that it was not in accordance with the Work Operation Standards (SOP) and not in accordance with the Technical Instructions (Guidelines for Implementation) and Juknis (Technical Instructions) .Naive Bayes is based on a simplified estimate that attribute values are conditionally independent of each other when given output values. Obtain an accuracy value of 0.75, with a negative precision classification of 0.75, negative recall1.00 f1-score negative0.86 and a negative3 support value. Precision classification Positive 0.00 positive recall 0.00 f1-score positive 0.00 and positive support value 1. This means that the performance of the system's success in retrieving information that has a positive value in the document is very low. -Binjai using the Naive Bayes algorithm method, it can be concluded as follows: 1) The accuracy level produced by the naïve Bayes classifier is 75%. 2) The advantages of this study are that it has good accuracy, precision and recall values, so that it is enough to be used in a system. 3) The deficiency in this research is in the performance of the system in finding the success of the system to find back information in the positive class, which is equal to 0.00%. This is because the amount of training data in the positive class is less than the negative class or it can be said that the training data used in this study is not balanced.

Keywords:Text Mining, Naïve Bayes, Analisis Sentimen

1. Introduction

The execution of this project has drawn sharp attention from Non-Governmental Organizations (NGOs) due to alleged deviations from the Standard Operating Procedures (SOP) and inconsistencies with the Implementation Guidelines (Juklak) and Technical Guidelines

DOI: 10.52362/ijiems.v3i1.1237
IJIEMS This work is licensed under a Creative Commons Attribution 4.0 International License.

International Journal of Informatics, Economics, Management and Science (IJIEMS)
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i1.1237
Volume 3, Issue 1, January 2024, pp. 102-107

(Juknis). Several contractor partners have also been working in non-compliance with the contract terms and have violated several agreements (sumut.tintarakyat, 2023). However, it turns out that it's not only the NGOs that have responded to the project; many members of the public have expressed their opinions through social media, one of which is Instagram. These comments will be taken as the subject of the research. These comments can be utilized as references for assessment, as a medium for observing public opinions regarding the related project, as well as an information source for performance enhancement. By harnessing public sentiments expressed through these comments, sentiment analysis is conducted.

Based on the aforementioned background, the author is interested in addressing a thesis with the title "Public Sentiment towards the Medan-Binjai Water Pipeline Excavation using the Naïve Bayes Method.

2. Literature

In this study, several references were taken from the literature from previous studies which were useful as a comparison of reference materials in completing this research.

This research is supported by (Harfian, 2021) with the title "Sentiment Classification of Digital Wallet Application Comments on Instagram Using Naive Bayes Classifier," which discusses the implementation of the Naive Bayes Classifier method for classifying positive and negative sentiments towards the Dana application based on comments on Instagram. The comment classification using the Naive Bayes Classifier method achieved the highest accuracy rate of 93.33% on the training data and a 90% : 10% ratio for test data.

Additionally, this study is also supported by (Pandunata et al., 2022) with the title "Sentiment Analysis of Public Opinions on the Covid-19 Vaccination Program in Indonesia on Twitter Using Naive Bayes Classifier." This study aims to identify the reasons for the public's pros and cons responses to this vaccination program. Based on the results of testing using the naive bayes method, varying accuracy results were obtained in the four experiments, with an average accuracy value in the 80s percentage range. The highest accuracy value was obtained in the fourth experiment, reaching 86% with 90% training data and 10% testing data.

This research is further supported by (Sidiq et al., 2020) in the paper titled "Sentiment Analysis of Toxic Comments in an Online Game Facebook Group Using Naïve Bayes Classification." Based on the research conducted, it can be concluded that the sentiment classification process of toxic comments in the AOV community group from a validated set of 1500 comments consisted of 1237 non-toxic comments and 263 toxic comments. The result indicated a tendency towards non-toxic sentiments due to the higher frequency of non-toxic words in the word cloud and the larger size of non-toxic words compared to toxic words.

Testing the classification model using an 80:20 data split ratio, where the model was tested without TF-IDF, with TF-IDF, and with the inclusion of Information Gain on TF-IDF, yielded the highest accuracy when using TF-IDF, with an accuracy of 75%, precision of 63%, recall of 67%, and F-measure of 64%.

DOI: 10.52362/ijiems.v3i1.1237
IJIEMS This work is licensed under a Creative Commons Attribution 4.0 International License.

International Journal of Informatics, Economics, Management and Science (IJIEMS)
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i1.1237
Volume 3, Issue 1, January 2024, pp. 102-107

3. Method

3.1. Definition of Sentiment Analysis

Sentiment analysis, also referred to as opinion mining, is a technique used to identify how a topic or activity is expressed using text and how it can be categorized as either a positive or negative sentiment (Julianto et al., 2022).

The sentiment analysis rules that can be used as follows:

1. **Text pre-processing**

   Text Preprocessing is a process to refine textual data and make it more structured. It involves several stages where the application selects data for processing in each document. The following are the steps of text preprocessing:

   a. *Case Folding :* Casing normalization is a stage used to standardize the usage of capital letters, converting them to lowercase.
   b. Cleaning : It is the process of cleaning the data from irrelevant words, such as characters, numbers, symbols, and emoticons.
   c. Tokenizing : It is the step of dividing the text within a document into tokens.
   d. Stemming : Lemmatization is the process that occurs after tokenizing. In this stage, words are transformed back to their base form.
   e. Normalization : is the step of altering the writing form of each word according to KBBI (Indonesian Dictionary).
   f. Filtering : is the process of eliminating irrelevant words obtained from tokenization.

2. **TF-IDF**

   TF-IDF weighting is used to assign weight values to words. Term Frequency (TF) represents the count of a specific word or term within a document. Inverse Document Frequency (IDF) is the frequency of appearance of a word or term across all documents.

   $$idf(t, D) = \log(\frac{N+1}{df(t)+1}) + 1$$

3. **NAÏVE BAYES CLASSIFIER**

   The equation of the Naïve Bayes theorem:
   P(C|X) = (P(X|C) * P(C)) / P(X)
   Where:
   X : Data with an unknown class.
   C : Hypothesis that the data belongs to a specific class.
   P(C|X) : The probability of hypothesis C given condition X (posterior probability).
   P(C) : The probability of hypothesis C (prior probability).
   P(X|C) : The probability of X given the condition in the hypothesis.
   P(X) : The probability of X.

3.2. Completetion of Naïve Bayes Cases

DOI: 10.52362/ijiems.v3i1.1237
IJIEMS This work is licensed under a Creative Commons Attribution 4.0 International License.

International Journal of Informatics, Economics, Management and Science (IJIEMS)
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i1.1237
Volume 3, Issue 1, January 2024, pp. 102-107

$$P\left(\frac{positif}{negatif}\right) = \frac{x(\frac{positif}{negatif})}{|c|} \dots$$

$$P(positif) = \frac{fx(positif)}{|c|} = \frac{1}{6} = 0.166$$

$$P(negatif) = \frac{fx(negatif)}{|c|} = \frac{1}{8} = 0.125$$

Table of Word Probabilities for Each Word

NO	KATA	POSITIF	NEGATIF
1	Pasang	0.1	0.4
2	Cctv	0.1	0.4
3	Pantau	0.1	0.4
4	kinerja	0.05	0.09
5	buruk	0.05	0.09
6	Lambat	0.05	0.09
7	Kerja	0.1	0.4
8	Malam	0.1	0.4
9	Setuju	0.1	0.4
10	Macet	0.05	0.09
11	Pagi	0.05	0.09
12	telat	0.05	0.09
13	proyek	0.05	0.09
14	nimbun	0.05	0.09

4. Result and Discussion

Table of Word Probabilities in the Test Set

Kata	Positif	Negatif
Macet	0.05	0.09
lambat	0.05	0.09

Then, to find the prediction values for each class, posterior probability is used.
Positif → log (0.16) + log(0.05) + log(0.05) = -3.39794000867

DOI: 10.52362/ijiems.v3i1.1237
IJIEMS This work is licensed under a Creative Commons Attribution 4.0 International License.

International Journal of Informatics, Economics, Management and Science (IJIEMS)
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i1.1237
Volume 3, Issue 1, January 2024, pp. 102-107

Negatif → log (0.12) + log(0.09) + log(0.09) = -3.01233373507

Table of Classification Prediction Values

Kelas	Prediksi
Positif	-339.794.000.867
Negatif	-301.233.373.507

Based on the calculations it is evident that the negative class has a higher value compared to the positive class value. Therefore, the result of the
classification for the above data is negative.

5. Conclusion

The conclusions drawn from this can be summarized as follows:
1) The accuracy level yielded by the naive bayes classifier classification is 75%.
2) The strength of this research lies in achieving good values of accuracy, precision, and recall, making it suitable for implementation within a system.
3) The limitation of this study lies in the system's performance in retrieving successful outcomes in locating positive-class information, which stands at 0.00%. This is attributed to an imbalance in the number of training data between the positive and negative classes, indicating that the data used in this research is not balanced.

References

[1] Agianto, R., Setiawati, A., & Firmansyah, R. (2020). "The Influence of Instagram Social Media on Adolescent Lifestyle and Ethics. TEMATIK-Journal of Information and Communication Technology, 7(2 SE-Articles), 130–139.
[2] Harfian, Y. (2021). "Sentiment Classification of Dana Digital Wallet Application Based on Comments on Instagram Using Naive Bayes Classifier" (Doctoral dissertation, State Islamic University of Sultan Syarif Kasim Riau).
[3] Hermanto, H., Mustopa, A., & Kuntoro, A. Y. (2020). "Naive Bayes and Support Vector Machine Classification Algorithms in Student Complaint Services." JITK (Journal of Computer Science and Technology), 5(2), 211-220.
[4] Hidayatullah, R., Rosyidah, U. A., & Abdurrahman, G. (2022). "Sentiment Analysis of Public towards Eiger Product Comments on Instagram Using Naive Bayes Classifier Method." National Multidisciplinary Sciences, 1(6), 770-777.
[5] Husada, H. C., & Paramita, A. S. (2021). "Sentiment Analysis on Airline Companies on Twitter Platform Using Support Vector Machine (SVM) Algorithm." Teknika, 10(1), 18-26.

DOI: 10.52362/ijiems.v3i1.1237
IJIEMS This work is licensed under a Creative Commons Attribution 4.0 International License.

International Journal of Informatics, Economics, Management and Science (IJIEMS)
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i1.1237
Volume 3, Issue 1, January 2024, pp. 102-107

[6] Julianto, Y., Setiabudi, D. H., & Rostianingsih, S. (2022). "Sentiment Analysis of Restaurant Reviews using Support Vector Machine Method." Jurnal Infra, 10(1), 1-7.

[7 Pandunata, P., Nurdiansyah, Y., & El Maidah, N. (2022). "Sentiment Analysis of Public Opinions on the Covid-19 Vaccination Program in Indonesia on Twitter Using Naive Bayes Classifier Method." INFORMAL: Informatics Journal, 7(3), 246-257.

[8] Ponomban, K., Jansen, T., & Supit, C. J. (2021). "Planning of Clean Water Supply System in Eris Village, Eris District, Minahasa Regency." JOURNAL SIPIL STATIK, 9(4).

[9] Ridwansyah, T. (2022). "Implementation of Text Mining for Sentiment Analysis of Global Public on Twitter towards the City of Medan Using K-Fold Cross Validation and Naïve Bayes Classifier." KLIK: Journal of Informatics and Computer Scientific Studies, 2(5), 178-185.

[10] Safi'i, A. M. (2021). "Text Mining Classification of Smartphone Product Reviews Using Multinomial Naive Bayes (Case Study: Asus ROG Smartphone Reviews)" (Doctoral dissertation, Kalimantan Institute of Technology).

[11] Sidiq, R. P., Dermawan, B. A., & Umaidah, Y. (2020). "Sentiment Analysis of Toxic Comments in Online Game Facebook Groups Using Naïve Bayes Classification." Journal of Information, Pamulang University, 5(3), 356.

[12] Susianti, I., Ningsih, S. S., Al Haris, M., & Utami, T. W. (2020). "Sentiment Analysis on Twitter Regarding the New Normal using Naïve Bayes Classifier Method." EDUSAINTEK, 4.

[13] Verawati, I., & Audit, B. S. (2022). "Naïve Bayes Classifier Algorithm for Sentiment Analysis of Twitter Users towards Provider By.u." Jurnal Media Informatika Budidarma, 6(3), 1411-1417.

[14] Yohana, P., Agustian, S., & Gusti, S. K. (Classification of Public Sentiment towards Covid-19 Vaccine Policy on Twitter with Imbalance Classes Using Naive Bayes. In National Seminar on Information Technology, Communication, and Industry (pp. 69-80).

DOI: 10.52362/ijiems.v3i1.1237
IJIEMS This work is licensed under a Creative Commons Attribution 4.0 International License.