BUKTI KORESPONDENSI

ARTIKEL JURNAL INTERNASIONAL BEREPUTASI

Judul artikel : Sentiment Analysis of Public Opinion on Handling Stunting In Indonesia Using Random Forest

Jurnal : Jurnal Ilmiah Teori dan Aplikasi Statistika, 2024, Volume 17 (1), 645 – 654

Penulis : Ariska Fitriyana Ningrum

No	Perihal	Tanggal
1	Bukti konfirmasi submit artikel dan artikel yang disubmit	8 Juni 2024
2	Bukti konfirmasi review dan hasil review pertama	19 Juni 2024
3	Bukti konfirmasi submit revisi pertama, respon kepada reviewer dan artikel yang di resubmit	20 Juni 2024
4	Bukti konfirmasi artikel accepted	08 Juli 2024

Sentiment Analysis Of Public Opinion On Handling Stunting In Indonesia Using Random Forest

Ariska Fitriyana Ningrum (1), Ihsan Fathoni Amri (2)

[1,2]Program Studi Sains Data, Universitas Muhammadiyah Semarang

Jln. Kedungmundu No 18, Kedungmundu, Kec. Tembalang, Kota Semarang
e-mail: ariskafitriyana@unimus.ac.id(1), ihsanfathoni@unimus.ac.id(2)

ABSTRAK

Masalah stunting penting untuk diselesaikan, karena berpotensi mengganggu potensi sumber daya manusia dan berkaitan dengan tingkat kesehatan, bahkan kematian anak. Pemerintah Indonesia menargetkan angka stunting turun menjadi 14 persen pada tahun 2024 melalui program percepatan penurunan stunting sebagai upaya meningkatkan status gizi masyarakat dan juga menurunkan prevalensi stunting atau balita pendek. Memahami sentimen publik terhadap inisiatif stunting sangat penting bagi para pembuat kebijakan dan pemangku kepentingan untuk merancang intervensi yang efektif dan mengalokasikan sumber daya secara efisien. Pada penelitian ini dilakukan klasifikasi pada sentiment positif dan negatif menggunakan algoritma random forest. Data yang digunakan adalah data komentar pada salah satu laman media sosial yaitu twitter mengenai sentiment masyarakat terhadap penanganan kasus stunting di Indonesia. Tahapan pertama pada penelitian ini setelah didapatkan sebuah data yaitu dilakukan prepocessing data. Tahapan preprocessing data dalam analisis sentimen berguna untuk membersihkan dan menormalkan teks, menghilangkan kata-kata tidak relevan, serta mempersiapkan data agar algoritma dapat menganalisis sentimen dengan lebih akurat dan efisien. Selanjutnya hasil data yang sudah di prepocessing diberikan label 0 untuk positif dan 1 untuk label negatif. Klasifikasi terhadap sentiment positif dan negatif ini dilakukan menggunakan random forest dan menghasilkan nilai akurasi sebesar 97,5%. Model ini sudah baik, namun kami menyarankan untuk mencoba algoritma lain dalam penelitian selanjutnya.

Kata kunci: Analisis Sentiment, Random Forest, Stunting

ABSTRACT

The problem of stunting is important to solve, as it has the potential to disrupt human resource potential and is linked to health outcomes and even child mortality. The Indonesian government targets the stunting rate to drop to 14 percent by 2024 through an accelerated stunting reduction program as an effort to improve the nutritional status of the community and also reduce the prevalence of stunting or short toddlers. Understanding public sentiment towards stunting initiatives is essential for policy makers and stakeholders to design effective interventions and allocate resources efficiently. In this research, classification of positive and negative sentiment is carried out using the random forest algorithm. The data used is comment data on one of the social media pages, namely Twitter, regarding public sentiment towards handling stunting cases in Indonesia. The first stage in this research after obtaining a data is data preprocessing. The data preprocessing stage in sentiment analysis is useful for cleaning and normalizing text, removing irrelevant words, and preparing data so that algorithms can analyze sentiment more accurately and efficiently. Furthermore, the results of the preprocessed data are labeled 0 for positive and 1 for negative labels. The classification of positive and negative sentiment was done using random forest and resulted in an accuracy value of 97.5%. This model is good, but we suggest trying other algorithms in future research.

Keywords: *Sentiment Analyst, Random Forest, Stunting*

INTRODUCTION

Stunting is a condition in children who experience growth disorders, so that the height and weight of children are not normal due to problems of nutritional deficiencies for a long time [1]. The problem of stunting in Indonesia is still quite large in the health sector today. According to the World Health Organization (WHO), as many as 22% or around 149.2 million children in the world under the age of five were recorded as stunted in 2020. Indonesia's position on the prevalence of stunting in the world is ranked 115 out of 151 countries. Meanwhile, in Southeast Asia, Indonesia is ranked second at 31.8% after Timor Leste at 48.8%. The third is Laos at 30.2%, the fourth is Cambodia at 29.9%, and the fifth is the Philippines at 29.9% [2] Stunting is caused by health problems, environmental factors and health services received by children. Genetic factors do not significantly affect stunting. Lack of nutrition in the fetus is the biggest cause of stunting in children. The first 1000 days of a child's life (1000 HPK) is the starting point for making important conclusions on long-term growth Thus, ineffective parenting and diet can increase the chance of stunting. Mental disorders and hypertension in mothers also affect the behavior and practices of nutrition in children. Limited access to health and sanitation services exacerbates the stunting conditions that occur in Indonesia such as lack of clean water, unclean latrines, and so on [3] Stunting in Indonesia is a deep-rooted problem. The problem of stunting is important to solve, because it has the potential to disrupt human resource potential and is related to health levels, even child mortality. In early 2021, the Indonesian government targeted the stunting rate to drop to 14 percent by 2024 through the accelerated stunting reduction program as an effort to improve the nutritional status of the community and also to reduce the prevalence of stunting or short toddlers [4].

Stunting in Indonesia is a deep-rooted problem. The problem of stunting is important to solve, because it has the potential to disrupt human resource potential and is related to health levels, even child mortality. In early 2021, the Indonesian government targeted the stunting rate to drop to 14 percent by 2024 through the accelerated stunting reduction program as an effort to improve the nutritional status of the community and also to reduce the prevalence of stunting or short toddlers[5] Understanding public sentiment towards stunting initiatives is crucial for policymakers and stakeholders to design effective interventions and allocate resources efficiently. Several previous studies have analyzed stunting predictions using the random forest algorithm which resulted in a classification accuracy value of 90.7%. [6]. Another study on social media analysis with the topic of stunting in Indonesia was conducted where the results showed that negative sentiment dominated by 60.6%, positive sentiment by 31.5%, and neutral by 7.9% [7]. In addition, this study shows that 'children', 'decline', 'numbers', 'prevention', and 'nutrition' are words that often appear in stunting [7]. Another study comparing SVM and random forest algorithms for the classification of stunting disease. the results show that the random forest algorithm provides higher accuracy of 88.2% compared to SVM of 65.6% [8].

The background of this study is based on the need to analyze public sentiment regarding the handling of stunting in Indonesia, which is a significant public health problem. The data used are the results of positive and negative reviews from the public on social media such as twitter regarding the handling of stunting cases in Indonesia. Some previous studies have also analyzed using comment data on twitter such as research on Sentiment Analysis of Twitter Netizens on the News of VAT on Basic Food and Education Services with Social Network Analysis and Naive

Bayes Classifier Approaches with data obtained as many as 4090 tweets [9]. While other studies have also conducted sentiment analysis of twitter users regarding online transportation service users [10].

The method used in this research is Random Forest. Random Forest was chosen because of its superior ability to handle complex and varied data, and provide accurate results in classification and prediction. This method is suitable for sentiment analysis because it is able to overcome overfitting, works well on large and irregular datasets, and provides a good interpretation of the features that affect the results. Random Forest is one of the state-of-the-art methods in machine learning that consists of a number of independently trained decision trees and the results are combined to produce more accurate and stable predictions. In the context of sentiment analysis, Random Forest can handle variations in language expression and identify relevant patterns from unstructured text data. In addition, Random Forest's ability to handle data with many features is very useful in sentiment analysis involving various emotional aspects and public opinions related to stunting treatment [11].

The state of the art of the Random Forest method shows that this technique has been successfully applied in various domains, including text and sentiment analysis, with satisfactory results. Previous studies have shown that Random Forest often outperforms other methods such as logistic regression and Support Vector Machines (SVM) in terms of accuracy and robustness to noise in the data. This makes Random Forest an appropriate choice for this research in an effort to understand and measure public sentiment towards stunting response efforts in Indonesia.

METHOD

In this study, a sentiment analysis about stunting in Indonesia was conducted. The remaining data will be analyzed using the random forest algorithm to check the accuracy of the sentiment results. The following are the stages of sentiment analysis research on stunting using the random forest algorithm

1. Data Collection

The data used in this study is the result of crawling Twitter data related to positive and negative responses to stunting conditions in Indonesia. The data collection process was carried out over a one-month period, starting from January 1, 2024 to January 31, 2024, to get a current picture of public sentiment on the issue of stunting. In the crawling process, certain filters and keywords related to infant stunting were used. The keywords used included "stunting", "child growth", and "child nutrition", as well as other keyword variations related to stunting and child health in Indonesia. In applying a series of filters, including language settings (Bahasa Indonesia) and geographical location (Indonesia) were used to ensure that the data captured was specifically related to responses to stunting conditions in Indonesia. This resulted in a total of 4601 comments showing both positive and negative opinions. Labeling the data into positive and negative categories was done through a manual process by the researcher, where each opinion was classified based on the sentiment expressed towards stunting conditions. This assessment is based on the context within each opinion, with the aim of gaining an accurate understanding of public sentiment.

2. Prepocessing Data

The first stage of the system is preprocessing. This stage involves several processes including Case Folding, Tokenization, Normalization, and Stemming. Case Folding is a task of splitting review text into smaller units called tokens or terms [12]. For infant stunting cases, what is done before and after case folding is, for example, "Breastfeeding mothers must have good nutrition" becomes "breastfeeding mothers must have good nutrition". Next is Tokenizing, in this process the separation is carried out on each word that makes up a document. In general, each word is identified or separated from other words by space characters, so the tokenizing process relies on space characters in the document to perform word separation [13] In sentiment analysis of stunting cases, what is done is to present the number of tokens generated from a review or comment. For example, from the sentence "breastfeeding mothers should have good nutrition", the tokens generated are "mother", "breastfeeding", "should", "nutrition", "which", "good". Normalization (Stopword Removal) process Removes special characters, numbers, and stopwords (common words) from each token. In the case of sentiment analysis, it shows a list of stopwords used and examples of text before and after stopword removal. For example, from "mother", "breastfeeding", "should", "nutrition", "which", "good", after removing the stopwords "should", "which", then "mother", "breastfeeding", "nutrition", "good" remains. This research also uses Stemming techniques which aim to find the base word, by removing all affixes that are fused to the word.[14] In Indonesian, this usually involves the removal of prefixes, suffixes or infixes. As an example of words before and after the stemming process, for example, "menyusui" can be reduced to "susu".

3. Sentiment Analyst Using Random Forest

The last stage is sentiment classification. Each review will be classified into positive or negative category. In this study, we employ random forest for the classification task. Random forest algorithm is a supervised classification algorithm. It is an ensemble learning technique based on decision tree algorithm [15]. Random Forest Algorithm is the advancement of Classification and Regression Tree (CART) method with the implementation of bootstrap aggregating (bagging) and random feature selection. Procedure of random forest algorithm on the data of n observations and p predictor [16]

a. Random samples of size n are drawn with the possibility of obtaining the same data (with replacement). This phase is called bootstrap.
b. Using the bootstrap samples, the tree is grown until the maximum size is reached, which is done without pruning. At each node, the random feature selection is used to determine the split, which m number of variables randomly sampled as candidates at each split must be m << p, at which point, the best node will be chosen based on m number of variables available for splitting [17]
c. Repeat stage 1 and 2 for k times to generate a forest that consists of k trees.
Breiman and Cutler suggests to observe the error OOB when

$$m = \left(\frac{1}{2}\left|\sqrt{p}\right|, \left|\sqrt{p}\right|, 2\left|\sqrt{p}\right|\right) \tag{1}$$

where *p* is the total variable and the number of *k* is small, then m with the smallest error OOB will be chosen.

In order to determine the split used as root node/node, Gini Index is used in Random Forest method. The formula of Gini index can be described as following:

$$\boldsymbol{Gini = 1 - \sum_{i=1}^{k} p_i^2} \tag{2}$$

where:

p_i : probability of an attribute being classified to class i.

k : total number of attributes being classified to a particular class.

The number of k suggested to apply in bagging is k = 50 which will provide satisfied results for classification [18]

RESULT AND DISCUSSION

In this study, the process of labeling public opinion data on stunting in Indonesia, obtained from social media through crawling, was carried out using a manual approach. Initially, a portion of the data was manually labeled by the research team to form the training dataset. Each opinion was classified into two categories: 0 for negative sentiment and 1 for positive sentiment. This manual labeling process is important to ensure that the data used accurately reflects public sentiment towards stunting and to generate a valid classification model.

```
Label                                     komen bersih
    0  dengar juta anak indonesia derita terapi efisi...
    0  bilang sisa rakyat komen lucu lucu gagal dulan...
    0  anak susah makan momok banget butuh nutrisi an...
    0                            kuwu tatangga desa nu
    0  rakyat indonesia sehat ongkos juta kondisi sak...
  ...                                              ...
    0  hamil minum puskesmas kasih gratis takut anemi...
    0  iya cupu cupu paham google search baca jurnal ...
    0  banteng jawa barat komitmen entas pdiperjuanga...
    0                                  bahan becandaan
    0  tingkat sabar setip tissue hadap anak emosi mu...
```

Figure 1. Review data that has been labelled

After obtaining a dataset of community reviews on stunting handling, data preprocessing is carried out with various stages such as *case folding, tokenizing, normalization, and stemming*. Data preprocessing is important because it helps to optimally clean, prepare, and organize data before analysis, thereby improving the accuracy, reliability, and interpretation of results from the model or analysis technique used.

The first stage, namely preprocessing, aims to homogenize all text into lowercase letters. The results of the case folding process are as follows.

2 Data Pre-Processing¶

- Tahap Case Holding => pada tahapan ini bertujuan untuk menyeragamkan seluruh teks kedalam huruf kecil (lowercase)

```
# Case Folding: Mengubah semua teks menjadi huruf kecil
data['komen case folding'] = data['clean_teks'].str.lower()

print("\n Data Setelah Case Folding:")
(data.head(10))
```

clean_teks	komen case folding
dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...
bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...
anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...
kuwu tatangga desa nu	kuwu tatangga desa nu
rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...

Figure 2. Case Holding Stages Result

Next, the tokenizing process is carried out. At this stage, the sentence in the comment will be broken down into words. The results of the tokenizing process are presented as in the following figure.

- Tokenizing proses dilakukannya pemecahan kata pada kalimat. Pada tahapan ini akan dilakukannya pemecahan dari kalimat di komentar menjadi kata per kata.

```
import nltk
from nltk.tokenize import word_tokenize
nltk.download('punkt')
data = data.dropna(subset=['komen case folding'])
data['komen tokenized'] = data['komen case folding'].apply(word_tokenize)
print("\nData Setelah Tokenizing:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]
0	rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...	[rakyat, indonesia, sehat, ongkos, juta, kondi...

Figure 3. Tokenizing Stages Result

The next preprocessing is to perform normalization to change the values of a dataset so that they have a uniform scale. The main purpose is to ensure that variables with different value ranges have equal influence when used in the analysis. The results of data normalization are as follows.

3. Normalization(Stopword Removal) proses Menghapus karakter khusus, angka, dan stopwords (kata-kata umum) pada setiap token.

```
from nltk.corpus import stopwords
import re
nltk.download('stopwords')

stop_words = set(stopwords.words('indonesian'))

data['komen normalized'] = data['komen tokenized'].apply(lambda tokens: [word for word in tokens if word.isalpha() and word not in stop_words]

print("\nData Setelah Normalization:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized	komen normalized
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...	[anak, susah, makan, momok, banget, butuh, nut...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]

Figure 4. Normalization Stages Result

The last step in data preprocessing is stemming. This aims to find the base word, by removing all affixes that are fused to the word. The results of stemming performed for sentiment analysis are as follows.

```python
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()
stemmer = factory.create_stemmer()
data['komen stemming'] = data['komen normalized'].apply(lambda tokens: [stemmer.stem(word) for word in tokens])

print("\nData Setelah Stemming:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized	komen normalized	komen stemming
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...	[anak, susah, makan, momok, banget, butuh, nut...	[anak, susah, makan, momok, banget, butuh, nut...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]
0	rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...	[rakyat, indonesia, sehat, ongkos, juta, kondi...	[rakyat, indonesia, sehat, ongkos, juta, kondi...	[rakyat, indonesia, sehat, ongkos, juta, kondi...

Figure 5 Stemming Performed for sentiment analysis

Furthermore, the results of data preprocessing that have been carried out can be seen visually regarding positive and negative opinions. Visualization aims to display the words that appear most or most often in a sentiment. Wordcloud this time describes each sentiment, the more often a word is used when giving a review, the larger the size of the word displayed on the wordcloud visualization. The following figure shows the visualization results for positive and negative sentiments

Figure 6. Visualization Positive Sentiment

Based on the figure above, it can be seen that in the positive sentiment there are several words that stand out such as, "balanced nutrition,", "helps reduce", "Welfare Growth" and several other words which indicate that the public's response to handling stunting in Indonesia has helped reduce stunting rates, provide balanced nutrition to children and can foster community welfare.

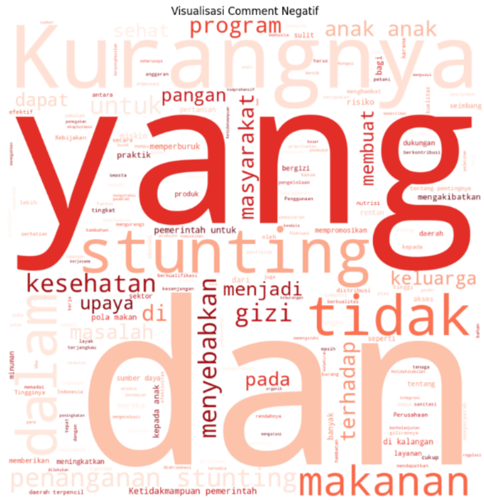

Figure 7. Visualization Negative Sentiment

Based on the figure above, it can be seen that in the negative sentiment regarding the handling of stunting in Indonesia, there are prominent words such as "lack of", "government inability", "gap" and several other words that indicate that the handling of stunting is still lacking, there is a social gap and the lack of government in handling cases of stunting in the society. Next, Classification on positive and negative comment data that has been done feature extraction using TF-IDF function is to convert text into a numerical vector representation based on the frequency of occurrence of words in it, taking into account the TF-IDF weight of each word. The classification metric evaluation calculations that we use are confution matrix, F1 measure and accuracy. The following are the results of the random forest performance evaluation

Table 1 Confution Matrix

		Actual	
		Negative	Positive
Prediction	Negative	893	0
	Positive	23	5

Based on table 1, which is related to the confusion matrix of prediction results, TP = 5, FP = 23, TN = 893, FN = 0, Total data = 258. The calculation results are obtained from the following calculations:

Table 2. Precision, recall, and f1-score

	Precision	Recall	f1-score
Negative	0.97	1.00	0.99
Positive	1.00	0.18	0.30

The performance generated by the random forest algorithm provides considerable accuracy, which is 97.50%, indicating that this model can classify data has a very good indication, and produces precision on Label 0 (negative comments) of 97% and recall of 100%, the results obtained are very high, and F1 score of 99%, indicating a high balance of precision and Recall. Meanwhile for precision on Label 1 (Positive comments) of 100%, and recall of 18% and the result for f1-score is 30%.

CONCLUSION

This study shows that public sentiment towards the handling of stunting cases in Indonesia can be divided into positive and negative based on the analysis of 4601 comments from Twitter social media. The results show that positive responses include the view that the handling of stunting has succeeded in reducing stunting rates, providing balanced nutrition to children, and potentially improving the general welfare of society. On the other hand, negative responses include dissatisfaction with the effectiveness of stunting handling, the existence of unresolved social inequalities, and the lack of effort from the government in handling stunting cases in the community.

This research continued with the classification of comment data based on sentiment using TF-IDF feature extraction. This method is important because it converts text into a numerical vector representation, where the TF-IDF weight of each word gives an idea of the importance of the word in determining positive or negative sentiment. Through this classification, it is possible to identify and categorize the sentiments present in the text data, enabling a deeper understanding of the public's views on stunting in Indonesia.

We performed sentiment analysis using random forest algorithm and achieved about 97.5% accuracy. I would recommend you to try using some other machine learning algorithms such as LSTM or KNN and see if you can get better results.

ACKNOWLEDGMENTS

This work was supported by Data Science Program, Faculty of Science and Agricultural Teknologi, Univercity of Muhammadiyah Semarang.

REFERENCE

[1] H. Rahman, M. Rahmah, and Nur Saribulan, "UPAYA PENANGANAN STUNTING DI INDONESIA," *Jurnal Ilmu Pemerintahan Suara Khatulistiwa (JIPSK)*, vol. VIII, no. 01, pp. 44–59, 2023.

[2] W. H. S. W. H. Organization, " Monitoring Health for the SDGs, World Health Organization," 2022.

[3] A. Muhaimin *et al.*, "Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia Analisis Sosial Media dan Pemodelan Topik: Kasus Studi Stunting di Indonesia," *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.

[4] R. Rizqi Robbi Arisandi, B. Warsito, and A. Rachman Hakim, "Aplikasi Naive Bayes Classifier (NBC) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *Jurnal Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[5] U. R. Gurning, S. F. Octavia, D. R. Andriyani, N. Nurainun, and I. Permana, "Prediksi Risiko Stunting pada Keluarga Menggunakan Naïve Bayes Classifier dan Chi-Square," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 172–180, Jan. 2024, doi: 10.57152/malcom.v4i1.1074.

[6] A. A. Reza1 and M. S. Rohman2, "JITE (Journal of Informatics and Telecommunication Engineering) Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization," *JITE*, vol. 7, no. 2, 2024, doi: 10.31289/jite.v7i2.10628.

[7] A. Muhaimin *et al.*, "Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia Analisis Sosial Media dan Pemodelan Topik: Kasus Studi Stunting di Indonesia," *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.

[8] M. Banurea, D. Betaria Hutagaol, and O. Sihombing, "KLASIFIKASI PENYAKIT STUNTING DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST," *Jurnal TEKINKOM*, vol. 6, no. 2, 2023, doi: 10.37600/tekinkom.v6i2.927.

[9] J. A. Nursiyono and C. Chotimah, "Analisis Sentimen Netizen Twitter terhadap Pemberitaan PPN Sembako dan Jasa Pendidikan dengan Pendekatan Social Network Analysis dan Naive Bayes Classifier," *Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 14, pp. 52–58, 2021.

[10] "Penerapan Text Mining pada Analisis Sentimen Pengguna Twitter Layanan Transportasi Online Menggunakan Metode Density Based Spatial Clustering of Applications With Noise (DBSCAN) dan K-Means," *Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 15, pp. 184–194, 2022.

[11] Stephenie, B. Warsito, and A. Prahutama, "Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method," in *E3S Web of Conferences*, EDP Sciences, Nov. 2020. doi: 10.1051/e3sconf/202020216006.

[12] M. Ali Fauzi, A. Z. Arifin, S. C. Gosaria, and I. S. Prabowo, "Indonesian news classification using naïve bayes and two-phase feature selection model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 8, no. 3, pp. 610–615, Dec. 2017, doi: 10.11591/ijeecs.v8.i3.pp610-615.

[13] Bahrawi, "Sentiment Analysis Using Random Forest Algorithm Online Social Media Based," Makassar, Dec. 2019.

[14] G. Khanvilkar and D. Vora, "Sentiment analysis for product recommendation using random forest," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 3, pp. 87–89, 2018, doi: 10.14419/ijet.v7i3.3.14492.

[15] L. Breiman, "Random Forests," 2001.

[16] Leo Breiman and A. Cutler, " Manual on Setting Up, Using, and Understanding Random Forest 4.0," 2003.

[17] M. A. Fauzi, "Random forest approach fo sentiment analysis in Indonesian language," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 1, pp. 46–50, Oct. 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.

[18] I. Afdhal *et al.*, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.

KOMENTAR REVIEWER

Sentiment Analysis Of Public Opinion On Handling Stunting In Indonesia Using Random Forest

ABSTRAK

Masalah stunting penting untuk diselesaikan, karena berpotensi mengganggu potensi sumber daya manusia dan berkaitan dengan tingkat kesehatan, bahkan kematian anak. Pemerintah Indonesia menargetkan angka stunting turun menjadi 14 persen pada tahun 2024 melalui program percepatan penurunan stunting sebagai upaya meningkatkan status gizi masyarakat dan juga menurunkan prevalensi stunting atau balita pendek. Memahami sentimen publik terhadap inisiatif stunting sangat penting bagi para pembuat kebijakan dan pemangku kepentingan untuk merancang intervensi yang efektif dan mengalokasikan sumber daya secara efisien. Dalam penelitian ini dilakukan analisis sentiment dari hasil crawling data twitter yang menunjukkan sentiment positif dan negatif masyarakat mengenai penanganan stunting di indonesia. selanjutnya dilakukan analisis klasifikasi menggunakan random forest dan menghasilkan nilai akurasi sebesar 97,5%. Model ini sudah cukup baik, namun kami menyarankan untuk mencoba algoritma lain dalam penelitian selanjutnya.

Kata kunci: Analisis Sentiment, Random Forest, Stunting

ABSTRACT

The issue of stunting is important to address, as it has the potential to affect the human resource potential and is related to health levels, and even child mortality. The Indonesian government targets to reduce the stunting rate to 14 percent by 2024 through an accelerated stunting reduction program as an effort to improve the nutritional status of the society and also reduce the prevalence of stunting or stunted children. Understanding public sentiment towards the stunting initiative is crucial for policymakers and stakeholders to design effective interventions and allocate resources efficiently. This study aims to analyze public sentiment related to stunting in Indonesia, which impacts children's growth and development. Through the use of sentiment analysis techniques, this study aims to understand public perceptions and attitudes towards the issue of stunting, evaluating whether the general sentiment is positive, negative or neutral. The results of this analysis are expected to provide useful insights for policymakers and health practitioners in designing and implementing more effective strategies to address the issue of stunting. This study conducted sentiment analysis from crawled Twitter data, showing positive and negative sentiments of the public regarding stunting handling in Indonesia. Furthermore, classification analysis using random forest was conducted and resulted in an accuracy score of 97.5%. The model is good enough but, we suggest trying other algorithms in further research

Keywords: *Sentiment Analyst, Random Forest, Stunting*

Commented [M.A1]: Please address the method used in this research and also give a summary how this research was done

INTRODUCTION

Stunting is a condition in children who experience growth disorders, so that the height and weight of children are not normal due to problems of nutritional deficiencies for a long time [1]. The problem of stunting in Indonesia is still quite large in the health sector today. According to the

Commented [M.A2]: You need to address some explanation about the method used, why this method is chosen, why this method is suitable related to the problem and also the state of the art from that

World Health Organization (WHO), as many as 22% or around 149.2 million children in the world under the age of five were recorded as stunted in 2020. Indonesia's position on the prevalence of stunting in the world is ranked 115 out of 151 countries. Meanwhile, in Southeast Asia, Indonesia is ranked second at 31.8% after Timor Leste at 48.8%. The third is Laos at 30.2%, the fourth is Cambodia at 29.9%, and the fifth is the Philippines at 29.9% [2] Stunting is caused by health problems, environmental factors and health services received by children. Genetic factors do not significantly affect stunting. Lack of nutrition in the fetus is the biggest cause of stunting in children. The first 1000 days of a child's life (1000 HPK) is the starting point for making important conclusions on long-term growth Thus, ineffective parenting and diet can increase the chance of stunting. Mental disorders and hypertension in mothers also affect the behavior and practices of nutrition in children. Limited access to health and sanitation services exacerbates the stunting conditions that occur in Indonesia such as lack of clean water, unclean latrines, and so on [3] Stunting in Indonesia is a deep-rooted problem. The problem of stunting is important to solve, because it has the potential to disrupt human resource potential and is related to health levels, even child mortality. In early 2021, the Indonesian government targeted the stunting rate to drop to 14 percent by 2024 through the accelerated stunting reduction program as an effort to improve the nutritional status of the community and also to reduce the prevalence of stunting or short toddlers [4].

Stunting in Indonesia is a deep-rooted problem. The problem of stunting is important to solve, because it has the potential to disrupt human resource potential and is related to health levels, even child mortality. In early 2021, the Indonesian government targeted the stunting rate to drop to 14 percent by 2024 through the accelerated stunting reduction program as an effort to improve the nutritional status of the community and also to reduce the prevalence of stunting or short toddlers[5] Understanding public sentiment towards stunting initiatives is crucial for policymakers and stakeholders to design effective interventions and allocate resources efficiently. Several previous studies have analyzed stunting predictions using the random forest algorithm which resulted in a classification accuracy value of 90.7%. [6]. Another study on social media analysis with the topic of stunting in Indonesia was conducted where the results showed that negative sentiment dominated by 60.6%, positive sentiment by 31.5%, and neutral by 7.9% [7]. In addition, this study shows that 'children', 'decline', 'numbers', 'prevention', and 'nutrition' are words that often appear in stunting [7]. Another study comparing SVM and random forest algorithms for the classification of stunting disease. the results show that the random forest algorithm provides higher accuracy of 88.2% compared to SVM of 65.6% [8].

Thus, this study will analyze public sentiment related to stunting in Indonesia using sentiment analysis techniques. This study aims to analyze public sentiment related to stunting in Indonesia, which impacts children's growth and development. Through the use of sentiment analysis techniques, this study aims to understand public perceptions and attitudes towards the issue of stunting, evaluating whether the general sentiment is positive, negative or neutral. The results of this analysis are expected to provide useful insights for policymakers and health practitioners in designing and implementing more effective strategies to address the issue of stunting. The techniques used in this research focus on the application of. Random Forest, a machine learning algorithm known for its robustness and accuracy in classification [9] . By utilizing a dataset consisting of textual data extracted from social media platforms, we seek to analyze the sentiments expressed towards stunting response efforts, identify key themes and topics associated with positive and negative sentiments, and assess the effectiveness of various strategies in addressing public concerns and perceptions.

METHOD

In this study, a sentiment analysis about stunting in Indonesia was conducted. The remaining data will be analyzed using the random forest algorithm to check the accuracy of the sentiment results. The following are the stages of sentiment analysis research on stunting using the random forest algorithm

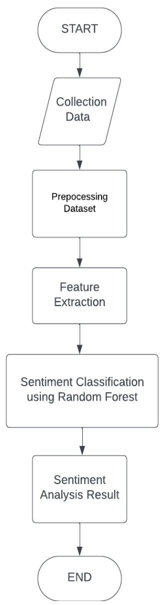

Figure 1 Flowchart Sentiment Analysis using Random Forest

1. Data Collection

The data used in this study is the result of crawling Twitter data related to positive and negative responses to stunting conditions in Indonesia. The data collection process was carried out over a one-month period, starting from January 1, 2024 to January 31, 2024, to get a current picture of public sentiment on the issue of stunting. In the crawling process, certain filters and keywords related to infant stunting were used. The keywords used included "stunting", "child growth", and "child nutrition", as well as other keyword variations related to stunting and child health in Indonesia. In applying a series of filters, including language settings (Bahasa Indonesia) and geographical location (Indonesia) were used to ensure that the data captured was specifically related to responses to stunting conditions in Indonesia. This resulted in a total of 4601 comments showing both positive and negative opinions. Labeling the data into positive and negative categories was done through a manual process by the researcher, where each opinion was classified based on the sentiment expressed towards stunting conditions. This assessment is based on the context within each opinion, with the aim of gaining an accurate understanding of public sentiment.

2. Prepocessing Data

The first stage of the system is preprocessing. This stage involves several processes including Case Folding, Tokenization, Normalization, and Stemming. Case Folding is a task of splitting review text into smaller units called tokens or terms [10]. For infant stunting cases, what is done before and after case folding is, for example, "Breastfeeding mothers must have good nutrition" becomes "breastfeeding mothers must have good nutrition". Next is Tokenizing, in this process the separation is carried out on each word that makes up a document. In general, each word is identified or separated from other words by space characters, so the tokenizing process relies on space characters in the document to perform word separation [11] In sentiment analysis of stunting cases, what is done is to present the number of tokens generated from a review or comment. For example, from the sentence "breastfeeding mothers should have good nutrition", the tokens generated are "mother", "breastfeeding", "should", "nutrition", "which", "good". Normalization (Stopword Removal) process Removes special characters, numbers, and stopwords (common words) from each token. In the case of sentiment analysis, it shows a list of stopwords used and examples of text before and after stopword removal. For example, from "mother", "breastfeeding", "should", "nutrition", "which", "good", after removing the stopwords "should", "which", then "mother", "breastfeeding", "nutrition", "good" remains. This research also uses Stemming techniques which aim to find the base word, by removing all affixes that are fused to the word.[12] In Indonesian, this usually involves the removal of prefixes, suffixes or infixes. As an example of words before and after the stemming process, for example, "menyusui" can be reduced to "susu".

3. Sentiment Analyst Using Random Forest

The last stage is sentiment classification. Each review will be classified into positive or negative category. In this study, we employ random forest for the classification task. Random forest algorithm is a supervised classification algorithm. It is an ensemble learning technique based on decision tree algorithm [13]. Random Forest Algorithm is the advancement of Classification and Regression Tree (CART) method with the implementation of bootstrap aggregating (bagging) and random feature selection. Procedure of random forest algorithm on the data of n observations and p predictor [14]

1. Random samples of size n are drawn with the possibility of obtaining the same data (with replacement). This phase is called bootstrap.
2. Using the bootstrap samples, the tree is grown until the maximum size is reached, which is done without pruning. At each node, the random feature selection is used to determine the split, which m number of variables randomly sampled as candidates at each split must be m << p, at which point, the best node will be chosen based on m number of variables available for splitting [15]
3. Repeat stage 1 and 2 for k times to generate a forest that consists of k trees.
   Breiman and Cutler suggests to observe the error OOB when

$$m = \left(\frac{1}{2}\left|\sqrt{p}\right|, \left|\sqrt{p}\right|, 2\left|\sqrt{p}\right|\right) \quad (1)$$

where p is the total variable and the number of k is small, then m with the smallest error OOB will be chosen.

In order to determine the split used as root node/node, Gini Index is used in Random Forest method. The formula of Gini index can be described as following:

$$Gini = 1 - \sum_{i=1}^{k} p_i^2 \quad (2)$$

where:

p_i : probability of an attribute being classified to class i.

k : total number of attributes being classified to a particular class.

The number of k suggested to apply in bagging is k = 50 which will provide satisfied results for classification [16]

RESULT AND DISCUSSION

In this study, the process of labeling public opinion data on stunting in Indonesia, obtained from social media through crawling, was carried out using a manual approach. Initially, a portion of the data was manually labeled by the research team to form the training dataset. Each opinion was classified into two categories: 0 for negative sentiment and 1 for positive sentiment. This manual labeling process is important to ensure that the data used accurately reflects public sentiment towards stunting and to generate a valid classification model.

```
Label                                          komen bersih
    0  dengar juta anak indonesia derita terapi efisi...
    0  bilang sisa rakyat komen lucu lucu gagal dulan...
    0  anak susah makan momok banget butuh nutrisi an...
    0                               kuwu tatangga desa nu
    0  rakyat indonesia sehat ongkos juta kondisi sak...
  ...                                                 ...
    0  hamil minum puskesmas kasih gratis takut anemi...
    0  iya cupu cupu paham google search baca jurnal ...
    0  banteng jawa barat komitmen entas pdiperjuanga...
    0                                      bahan becandaan
    0  tingkat sabar setip tissue hadap anak emosi mu...
```

Figure 2. Review data that has been labelled

After obtaining a dataset of community reviews on stunting handling, data preprocessing is carried out with various stages such as *case folding, tokenizing, normalization, and stemming*. The first stage, namely preprocessing, aims to homogenize all text into lowercase letters. The results of the case folding process are as follows.

Commented [M.A3]: Please explain this process and why this is matter

2 Data Pre-Processing¶

- Tahap Case Holding => pada tahapan ini bertujuan untuk menyeragamkan seluruh teks kedalam huruf kecil (lowercase)

```
# Case Folding: Mengubah semua teks menjadi huruf kecil
data['komen case folding'] = data['clean_teks'].str.lower()

print("\n Data Setelah Case Folding:")
(data.head(10))
```

clean_teks	komen case folding
dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...
bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...
anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...
kuwu tatangga desa nu	kuwu tatangga desa nu
rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...

Figure 3. Case Holding Stages Result

Next, the tokenizing process is carried out. At this stage, the sentence in the comment will be broken down into words. The results of the tokenizing process are presented as in the following figure.

- Tokenizing proses dilakukannya pemecahan kata pada kalimat. Pada tahapan ini akan dilakukannya pemecahan dari kalimat di komentar menjadi kata per kata.

```
import nltk
from nltk.tokenize import word_tokenize
nltk.download('punkt')
data = data.dropna(subset=['komen case folding'])
data['komen tokenized'] = data['komen case folding'].apply(word_tokenize)
print("\nData Setelah Tokenizing:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]
0	rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...	[rakyat, indonesia, sehat, ongkos, juta, kondi...

Figure 4. Tokenizing Stages Result

The next preprocessing is to perform normalization to change the values of a dataset so that they have a uniform scale. The main purpose is to ensure that variables with different value ranges have equal influence when used in the analysis. The results of data normalization are as follows.

3. Normalization(Stopword Removal) proses Menghapus karakter khusus, angka, dan stopwords (kata-kata umum) pada setiap token.

```
[ ] from nltk.corpus import stopwords
    import re
    nltk.download('stopwords')

    stop_words = set(stopwords.words('indonesian'))

    data['komen normalized'] = data['komen tokenized'].apply(lambda tokens: [word for word in tokens if word.isalpha() and word not in stop_words])

    print("\nData Setelah Normalization:")
    (data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized	komen normalized
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...	[anak, susah, makan, momok, banget, butuh, nut...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]

Figure 5. Normalization Stages Result

The last step in data preprocessing is stemming. This aims to find the base word, by removing all affixes that are fused to the word. The results of stemming performed for sentiment analysis are as follows.

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()
stemmer = factory.create_stemmer()
data['komen stemming'] = data['komen normalized'].apply(lambda tokens: [stemmer.stem(word) for word in tokens])

print("\nData Setelah Stemming:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized	komen normalized	komen stemming
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...	[anak, susah, makan, momok, banget, butuh, nut...	[anak, susah, makan, momok, banget, butuh, nut...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]
0	rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...	[rakyat, indonesia, sehat, ongkos, juta, kondi...	[rakyat, indonesia, sehat, ongkos, juta, kondi...	[rakyat, indonesia, sehat, ongkos, juta, kondi...

Figure 6 Stemming Performed for sentiment analysis

Furthermore, the results of data preprocessing that have been carried out can be seen visually regarding positive and negative opinions. Visualization aims to display the words that appear most or most often in a sentiment. Wordcloud this time describes each sentiment, the more often a word is used when giving a review, the larger the size of the word displayed on the wordcloud visualization. The following figure shows the visualization results for positive and negative sentiments

Figure 7. Visualization Positive Sentimen

Based on the figure above, it can be seen that in the positive sentiment there are several words that stand out such as, "balanced nutrition,", "helps reduce", "Welfare Growth" and several other words which indicate that the public's response to handling stunting in Indonesia has helped reduce stunting rates, provide balanced nutrition to children and can foster community welfare.

Figure 8. Visualization Negative Sentiment

Based on the figure above, it can be seen that in the negative sentiment regarding the handling of stunting in Indonesia, there are prominent words such as "lack of", "government inability", "gap" and several other words that indicate that the handling of stunting is still lacking, there is a social gap and the lack of government in handling cases of stunting in the society.

Next, Classification on positive and negative comment data that has been done feature extraction using TF-IDF function is to convert text into a numerical vector representation based on the frequency of occurrence of words in it, taking into account the TF-IDF weight of each word. The classification metric evaluation calculations that we use are confution matrix, F1 measure and accuracy. The following are the results of the random forest performance evaluation

Table 1 Confution Matrix

		Actual	
		Negative	Positive
Prediction	Negative	893	0
	Positive	23	5

Based on table 1, which is related to the confusion matrix of prediction results, TP = 5, FP = 23, TN = 893, FN = 0, Total data = 258. The calculation results are obtained from the following calculations:

Table 2. Precision, recall, and f1-score

	Precision	Recall	f1-score
Negative	0.97	1.00	0.99
Positive	1.00	0.18	0.30

The performance generated by the random forest algorithm provides considerable accuracy, which is 97.50%, indicating that this model can classify data has a very good indication, and produces precision on Label 0 (negative comments) of 97% and recall of 100%, the results obtained are very high, and F1 score of 99%, indicating a high balance of precision and Recall. Meanwhile for precision on Label 1 (Positive comments) of 100%, and recall of 18% and the result for f1-score is 30%.

This section explain the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily. On each figure should be given a caption below the figure (Figure 1). The captions on the table are given above the table. Captions are written in lowercase letters except for the first character of each sentence. All figures should be numbered sequentially.

CONCLUSION

Commented [M.A4]: what the sentiment public goes to? Is it positif or negative? You need to explain

Sentiment analysis or opinion mining is a field of study that analyzes people's sentiments, attitudes, or emotions towards certain entities. this research discusses public sentiment towards handling stunting disease in Indonesia. A total of 4601 tweets were selected as the data used for this study. We performed sentiment analysis using random forest algorithm and achieved about

97.5% accuracy. I would recommend you to try using some other machine learning algorithms such as logistic regression, SVM, or KNN and see if you can get better results.

ACKNOWLEDGMENTS

This work was supported by Data Science Program, Faculty of Science and Agricultural Teknologi, Univercity of Muhammadiyah Semarang.

REFERENCE

Commented [M.A5]: add 2 reference from jstatistika

[1] H. Rahman, M. Rahmah, and Nur Saribulan, "UPAYA PENANGANAN STUNTING DI INDONESIA," *Jurnal Ilmu Pemerintahan Suara Khatulistiwa (JIPSK)*, vol. VIII, no. 01, pp. 44–59, 2023.

[2] W. H. S. W. H. Organization, " Monitoring Health for the SDGs, World Health Organization," 2022.

[3] A. Muhaimin *et al.*, "Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia Analisis Sosial Media dan Pemodelan Topik: Kasus Studi Stunting di Indonesia," *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.

[4] R. Rizqi Robbi Arisandi, B. Warsito, and A. Rachman Hakim, "Aplikasi Naive Bayes Classifier (NBC) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *Jurnal Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[5] U. R. Gurning, S. F. Octavia, D. R. Andriyani, N. Nurainun, and I. Permana, "Prediksi Risiko Stunting pada Keluarga Menggunakan Naïve Bayes Classifier dan Chi-Square," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 172–180, Jan. 2024, doi: 10.57152/malcom.v4i1.1074.

[6] A. A. Reza1 and M. S. Rohman2, "JITE (Journal of Informatics and Telecommunication Engineering) Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization," *JITE*, vol. 7, no. 2, 2024, doi: 10.31289/jite.v7i2.10628.

[7] A. Muhaimin *et al.*, "Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia Analisis Sosial Media dan Pemodelan Topik: Kasus Studi Stunting di Indonesia," *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.

[8] M. Banurea, D. Betaria Hutagaol, and O. Sihombing, "KLASIFIKASI PENYAKIT STUNTING DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST," *Jurnal TEKINKOM*, vol. 6, no. 2, 2023, doi: 10.37600/tekinkom.v6i2.927.

[9] Stephenie, B. Warsito, and A. Prahutama, "Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method," in *E3S Web of Conferences*, EDP Sciences, Nov. 2020. doi: 10.1051/e3sconf/202020216006.

[10] M. Ali Fauzi, A. Z. Arifin, S. C. Gosaria, and I. S. Prabowo, "Indonesian news classification using naïve bayes and two-phase feature selection model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 8, no. 3, pp. 610–615, Dec. 2017, doi: 10.11591/ijeecs.v8.i3.pp610-615.

[11] Bahrawi, "Sentiment Analysis Using Random Forest Algorithm Online Social Media Based," Makassar, Dec. 2019.

[12] G. Khanvilkar and D. Vora, "Sentiment analysis for product recommendation using random forest," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 3, pp. 87–89, 2018, doi: 10.14419/ijet.v7i3.3.14492.

[13] L. Breiman, “Random Forests,” 2001.
[14] Leo Breiman and A. Cutler, “ Manual on Setting Up, Using, and Understanding Random Forest 4.0,” 2003.
[15] M. A. Fauzi, “Random forest approach fo sentiment analysis in Indonesian language,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 1, pp. 46–50, Oct. 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
[16] I. Afdhal *et al.*, “Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.

JAWABAN KOREKSI

Sentiment Analysis Of Public Opinion On Handling Stunting In Indonesia Using Random Forest

ABSTRAK

Masalah stunting penting untuk diselesaikan, karena berpotensi mengganggu potensi sumber daya manusia dan berkaitan dengan tingkat kesehatan, bahkan kematian anak. Pemerintah Indonesia menargetkan angka stunting turun menjadi 14 persen pada tahun 2024 melalui program percepatan penurunan stunting sebagai upaya meningkatkan status gizi masyarakat dan juga menurunkan prevalensi stunting atau balita pendek. Memahami sentimen publik terhadap inisiatif stunting sangat penting bagi para pembuat kebijakan dan pemangku kepentingan untuk merancang intervensi yang efektif dan mengalokasikan sumber daya secara efisien. Dalam penelitian ini dilakukan analisis sentiment dari hasil crawling data twitter yang menunjukkan sentiment positif dan negatif masyarakat mengenai penanganan stunting di indonesia. selanjutnya dilakukan analisis klasifikasi menggunakan random forest dan menghasilkan nilai akurasi sebesar 97,5%. Model ini sudah cukup baik, namun kami menyarankan untuk mencoba algoritma lain dalam penelitian selanjutnya.

Kata kunci: Analisis Sentiment, Random Forest, Stunting

ABSTRACT

The issue of stunting is important to address, as it has the potential to affect the human resource potential and is related to health levels, and even child mortality. The Indonesian government targets to reduce the stunting rate to 14 percent by 2024 through an accelerated stunting reduction program as an effort to improve the nutritional status of the society and also reduce the prevalence of stunting or stunted children. Understanding public sentiment towards the stunting initiative is crucial for policymakers and stakeholders to design effective interventions and allocate resources efficiently. This study aims to analyze public sentiment related to stunting in Indonesia, which impacts children's growth and development. Through the use of sentiment analysis techniques, this study aims to understand public perceptions and attitudes towards the issue of stunting, evaluating whether the general sentiment is positive, negative or neutral. The results of this analysis are expected to provide useful insights for policymakers and health practitioners in designing and implementing more effective strategies to address the issue of stunting. This study conducted sentiment analysis from crawled Twitter data, showing positive and negative sentiments of the public regarding stunting handling in Indonesia. Furthermore, classification analysis using random forest was conducted and resulted in an accuracy score of 97.5%. The model is good enough but, we suggest trying other algorithms in further research

Keywords: *Sentiment Analyst, Random Forest, Stunting*

Commented [M.A1]: Please address the method used in this research and also give a summary how this research was done

Commented [AN2R1]: This research was conducted by scrapping review data on the handling of stunting cases in Indonesia with the random forest method.

INTRODUCTION

Stunting is a condition in children who experience growth disorders, so that the height and weight of children are not normal due to problems of nutritional deficiencies for a long time [1]. The problem of stunting in Indonesia is still quite large in the health sector today. According to the

Commented [M.A3]: You need to address some explanation about the method used, why this method is chosen, why this method is suitable related to the problem and also the state of the art from that

Commented [AN4R3]: I will briefly explain how the random forest algorithm works in classifying positive and negative sentiment on the results of a review of stunting handling in Indonesia.

World Health Organization (WHO), as many as 22% or around 149.2 million children in the world under the age of five were recorded as stunted in 2020. Indonesia's position on the prevalence of stunting in the world is ranked 115 out of 151 countries. Meanwhile, in Southeast Asia, Indonesia is ranked second at 31.8% after Timor Leste at 48.8%. The third is Laos at 30.2%, the fourth is Cambodia at 29.9%, and the fifth is the Philippines at 29.9% [2] Stunting is caused by health problems, environmental factors and health services received by children. Genetic factors do not significantly affect stunting. Lack of nutrition in the fetus is the biggest cause of stunting in children. The first 1000 days of a child's life (1000 HPK) is the starting point for making important conclusions on long-term growth Thus, ineffective parenting and diet can increase the chance of stunting. Mental disorders and hypertension in mothers also affect the behavior and practices of nutrition in children. Limited access to health and sanitation services exacerbates the stunting conditions that occur in Indonesia such as lack of clean water, unclean latrines, and so on [3] Stunting in Indonesia is a deep-rooted problem. The problem of stunting is important to solve, because it has the potential to disrupt human resource potential and is related to health levels, even child mortality. In early 2021, the Indonesian government targeted the stunting rate to drop to 14 percent by 2024 through the accelerated stunting reduction program as an effort to improve the nutritional status of the community and also to reduce the prevalence of stunting or short toddlers [4].

Stunting in Indonesia is a deep-rooted problem. The problem of stunting is important to solve, because it has the potential to disrupt human resource potential and is related to health levels, even child mortality. In early 2021, the Indonesian government targeted the stunting rate to drop to 14 percent by 2024 through the accelerated stunting reduction program as an effort to improve the nutritional status of the community and also to reduce the prevalence of stunting or short toddlers[5] Understanding public sentiment towards stunting initiatives is crucial for policymakers and stakeholders to design effective interventions and allocate resources efficiently. Several previous studies have analyzed stunting predictions using the random forest algorithm which resulted in a classification accuracy value of 90.7%. [6]. Another study on social media analysis with the topic of stunting in Indonesia was conducted where the results showed that negative sentiment dominated by 60.6%, positive sentiment by 31.5%, and neutral by 7.9% [7]. In addition, this study shows that 'children', 'decline', 'numbers', 'prevention', and 'nutrition' are words that often appear in stunting [7]. Another study comparing SVM and random forest algorithms for the classification of stunting disease. the results show that the random forest algorithm provides higher accuracy of 88.2% compared to SVM of 65.6% [8].

Thus, this study will analyze public sentiment related to stunting in Indonesia using sentiment analysis techniques. This study aims to analyze public sentiment related to stunting in Indonesia, which impacts children's growth and development. Through the use of sentiment analysis techniques, this study aims to understand public perceptions and attitudes towards the issue of stunting, evaluating whether the general sentiment is positive, negative or neutral. The results of this analysis are expected to provide useful insights for policymakers and health practitioners in designing and implementing more effective strategies to address the issue of stunting. The techniques used in this research focus on the application of. Random Forest, a machine learning algorithm known for its robustness and accuracy in classification [9] . By utilizing a dataset consisting of textual data extracted from social media platforms, we seek to analyze the sentiments expressed towards stunting response efforts, identify key themes and topics associated with positive and negative sentiments, and assess the effectiveness of various strategies in addressing public concerns and perceptions.

METHOD

In this study, a sentiment analysis about stunting in Indonesia was conducted. The remaining data will be analyzed using the random forest algorithm to check the accuracy of the sentiment results. The following are the stages of sentiment analysis research on stunting using the random forest algorithm

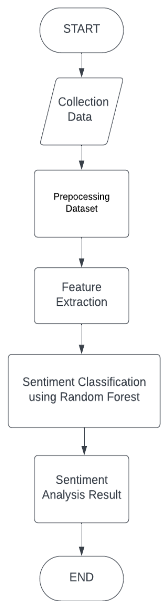

Figure 1 Flowchart Sentiment Analysis using Random Forest

1. Data Collection

The data used in this study is the result of crawling Twitter data related to positive and negative responses to stunting conditions in Indonesia. The data collection process was carried out over a one-month period, starting from January 1, 2024 to January 31, 2024, to get a current picture of public sentiment on the issue of stunting. In the crawling process, certain filters and keywords related to infant stunting were used. The keywords used included "stunting", "child growth", and "child nutrition", as well as other keyword variations related to stunting and child health in Indonesia. In applying a series of filters, including language settings (Bahasa Indonesia) and geographical location (Indonesia) were used to ensure that the data captured was specifically related to responses to stunting conditions in Indonesia. This resulted in a total of 4601 comments showing both positive and negative opinions. Labeling the data into positive and negative categories was done through a manual process by the researcher, where each opinion was classified based on the sentiment expressed towards stunting conditions. This assessment is based on the context within each opinion, with the aim of gaining an accurate understanding of public sentiment.

2. Prepocessing Data

The first stage of the system is preprocessing. This stage involves several processes including Case Folding, Tokenization, Normalization, and Stemming. Case Folding is a task of splitting review text into smaller units called tokens or terms [10]. For infant stunting cases, what is done before and after case folding is, for example, "Breastfeeding mothers must have good nutrition" becomes "breastfeeding mothers must have good nutrition". Next is Tokenizing, in this process the separation is carried out on each word that makes up a document. In general, each word is identified or separated from other words by space characters, so the tokenizing process relies on space characters in the document to perform word separation [11] In sentiment analysis of stunting cases, what is done is to present the number of tokens generated from a review or comment. For example, from the sentence "breastfeeding mothers should have good nutrition", the tokens generated are "mother", "breastfeeding", "should", "nutrition", "which", "good". Normalization (Stopword Removal) process Removes special characters, numbers, and stopwords (common words) from each token. In the case of sentiment analysis, it shows a list of stopwords used and examples of text before and after stopword removal. For example, from "mother", "breastfeeding", "should", "nutrition", "which", "good", after removing the stopwords "should", "which", then "mother", "breastfeeding", "nutrition", "good" remains. This research also uses Stemming techniques which aim to find the base word, by removing all affixes that are fused to the word.[12] In Indonesian, this usually involves the removal of prefixes, suffixes or infixes. As an example of words before and after the stemming process, for example, "menyusui" can be reduced to "susu".

3. Sentiment Analyst Using Random Forest

The last stage is sentiment classification. Each review will be classified into positive or negative category. In this study, we employ random forest for the classification task. Random forest algorithm is a supervised classification algorithm. It is an ensemble learning technique based on decision tree algorithm [13]. Random Forest Algorithm is the advancement of Classification and Regression Tree (CART) method with the implementation of bootstrap aggregating (bagging) and random feature selection. Procedure of random forest algorithm on the data of n observations and p predictor [14]

1. Random samples of size n are drawn with the possibility of obtaining the same data (with replacement). This phase is called bootstrap.
2. Using the bootstrap samples, the tree is grown until the maximum size is reached, which is done without pruning. At each node, the random feature selection is used to determine the split, which m number of variables randomly sampled as candidates at each split must be m << p, at which point, the best node will be chosen based on m number of variables available for splitting [15]
3. Repeat stage 1 and 2 for k times to generate a forest that consists of k trees.
   Breiman and Cutler suggests to observe the error OOB when

$$m = \left(\frac{1}{2}\left|\sqrt{p}\right|, \left|\sqrt{p}\right|, 2\left|\sqrt{p}\right|\right) \quad (1)$$

where p is the total variable and the number of k is small, then m with the smallest error OOB will be chosen.

In order to determine the split used as root node/node, Gini Index is used in Random Forest method. The formula of Gini index can be described as following:

$$\boldsymbol{Gini = 1 - \sum_{i=1}^{k} p_i^2} \quad (2)$$

where:

p_i : probability of an attribute being classified to class i.

k : total number of attributes being classified to a particular class.

The number of k suggested to apply in bagging is k = 50 which will provide satisfied results for classification [16]

RESULT AND DISCUSSION

In this study, the process of labeling public opinion data on stunting in Indonesia, obtained from social media through crawling, was carried out using a manual approach. Initially, a portion of the data was manually labeled by the research team to form the training dataset. Each opinion was classified into two categories: 0 for negative sentiment and 1 for positive sentiment. This manual labeling process is important to ensure that the data used accurately reflects public sentiment towards stunting and to generate a valid classification model.

```
Label                                          komen bersih
    0   dengar juta anak indonesia derita terapi efisi...
    0   bilang sisa rakyat komen lucu lucu gagal dulan...
    0   anak susah makan momok banget butuh nutrisi an...
    0                                 kuwu tatangga desa nu
    0   rakyat indonesia sehat ongkos juta kondisi sak...
  ...                                                   ...
    0   hamil minum puskesmas kasih gratis takut anemi...
    0   iya cupu cupu paham google search baca jurnal ...
    0   banteng jawa barat komitmen entas pdiperjuanga...
    0                                       bahan becandaan
    0   tingkat sabar setip tissue hadap anak emosi mu...
```

Figure 2. Review data that has been labelled

After obtaining a dataset of community reviews on stunting handling, data preprocessing is carried out with various stages such as *case folding, tokenizing, normalization, and stemming*. The first stage, namely preprocessing, aims to homogenize all text into lowercase letters. The results of the case folding process are as follows.

Commented [M.A5]: Please explain this process and why this is matter

Commented [AN6R5]: yes, i will explain this process

2 Data Pre-Processing¶

- Tahap Case Holding => pada tahapan ini bertujuan untuk menyeragamkan seluruh teks kedalam huruf kecil (lowercase)

```
# Case Folding: Mengubah semua teks menjadi huruf kecil
data['komen case folding'] = data['clean_teks'].str.lower()

print("\n Data Setelah Case Folding:")
(data.head(10))
```

clean_teks	komen case folding
dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...
bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...
anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...
kuwu tatangga desa nu	kuwu tatangga desa nu
rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...

Figure 3. Case Holding Stages Result

Next, the tokenizing process is carried out. At this stage, the sentence in the comment will be broken down into words. The results of the tokenizing process are presented as in the following figure.

- Tokenizing proses dilakukannya pemecahan kata pada kalimat. Pada tahapan ini akan dilakukannya pemecahan dari kalimat di komentar menjadi kata per kata.

```
import nltk
from nltk.tokenize import word_tokenize
nltk.download('punkt')
data = data.dropna(subset=['komen case folding'])
data['komen tokenized'] = data['komen case folding'].apply(word_tokenize)
print("\nData Setelah Tokenizing:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]
0	rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...	[rakyat, indonesia, sehat, ongkos, juta, kondi...

Figure 4. Tokenizing Stages Result

The next preprocessing is to perform normalization to change the values of a dataset so that they have a uniform scale. The main purpose is to ensure that variables with different value ranges have equal influence when used in the analysis. The results of data normalization are as follows.

3. Normalization(Stopword Removal) proses Menghapus karakter khusus, angka, dan stopwords (kata-kata umum) pada setiap token.

```
[ ] from nltk.corpus import stopwords
    import re
    nltk.download('stopwords')

    stop_words = set(stopwords.words('indonesian'))

    data['komen normalized'] = data['komen tokenized'].apply(lambda tokens: [word for word in tokens if word.isalpha() and word not in stop_words])

    print("\nData Setelah Normalization:")
    (data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized	komen normalized
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...	[anak, susah, makan, momok, banget, butuh, nut...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]

Figure 5. Normalization Stages Result

The last step in data preprocessing is stemming. This aims to find the base word, by removing all affixes that are fused to the word. The results of stemming performed for sentiment analysis are as follows.

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()
stemmer = factory.create_stemmer()
data['komen stemming'] = data['komen normalized'].apply(lambda tokens: [stemmer.stem(word) for word in tokens])

print("\nData Setelah Stemming:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized	komen normalized	komen stemming
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...	[anak, susah, makan, momok, banget, butuh, nut...	[anak, susah, makan, momok, banget, butuh, nut...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]
0	rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...	[rakyat, indonesia, sehat, ongkos, juta, kondi...	[rakyat, indonesia, sehat, ongkos, juta, kondi...	[rakyat, indonesia, sehat, ongkos, juta, kondi...

Figure 6 Stemming Performed for sentiment analysis

Furthermore, the results of data preprocessing that have been carried out can be seen visually regarding positive and negative opinions. Visualization aims to display the words that appear most or most often in a sentiment. Wordcloud this time describes each sentiment, the more often a word is used when giving a review, the larger the size of the word displayed on the wordcloud visualization. The following figure shows the visualization results for positive and negative sentiments

Figure 7. Visualization Positive Sentimen

Based on the figure above, it can be seen that in the positive sentiment there are several words that stand out such as, "balanced nutrition,", "helps reduce", "Welfare Growth" and several other words which indicate that the public's response to handling stunting in Indonesia has helped reduce stunting rates, provide balanced nutrition to children and can foster community welfare.

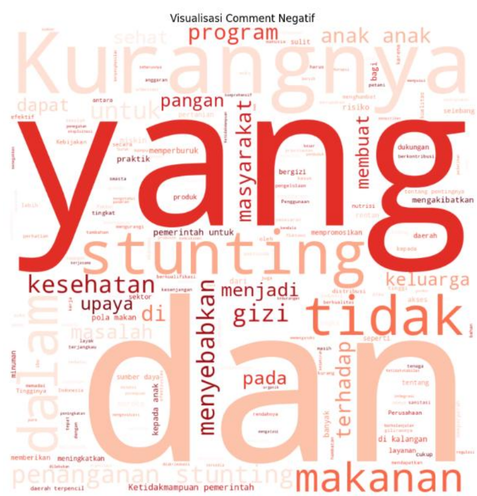

Figure 8. Visualization Negative Sentiment

Based on the figure above, it can be seen that in the negative sentiment regarding the handling of stunting in Indonesia, there are prominent words such as "lack of", "government inability", "gap" and several other words that indicate that the handling of stunting is still lacking, there is a social gap and the lack of government in handling cases of stunting in the society.

Next, Classification on positive and negative comment data that has been done feature extraction using TF-IDF function is to convert text into a numerical vector representation based on the frequency of occurrence of words in it, taking into account the TF-IDF weight of each word. The classification metric evaluation calculations that we use are confution matrix, F1 measure and accuracy. The following are the results of the random forest performance evaluation

Table 1 Confution Matrix

		Actual	
		Negative	Positive
Prediction	Negative	893	0
	Positive	23	5

Based on table 1, which is related to the confusion matrix of prediction results, TP = 5, FP = 23, TN = 893, FN = 0, Total data = 258. The calculation results are obtained from the following calculations:

Table 2. Precision, recall, and f1-score

	Precision	Recall	f1-score
Negative	0.97	1.00	0.99
Positive	1.00	0.18	0.30

The performance generated by the random forest algorithm provides considerable accuracy, which is 97.50%, indicating that this model can classify data has a very good indication, and produces precision on Label 0 (negative comments) of 97% and recall of 100%, the results obtained are very high, and F1 score of 99%, indicating a high balance of precision and Recall. Meanwhile for precision on Label 1 (Positive comments) of 100%, and recall of 18% and the result for f1-score is 30%.

This section explain the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily. On each figure should be given a caption below the figure (Figure 1). The captions on the table are given above the table. Captions are written in lowercase letters except for the first character of each sentence. All figures should be numbered sequentially.

CONCLUSION

Sentiment analysis or opinion mining is a field of study that analyzes people's sentiments, attitudes, or emotions towards certain entities. this research discusses public sentiment towards handling stunting disease in Indonesia. A total of 4601 tweets were selected as the data used for this study. We performed sentiment analysis using random forest algorithm and achieved about

Commented [M.A7]: what the sentiment public goes to? Is it positif or negative? You need to explain

Commented [AN8R7]: Majority are still giving negative reviews on this sentiment

97.5% accuracy. I would recommend you to try using some other machine learning algorithms such as logistic regression, SVM, or KNN and see if you can get better results.

ACKNOWLEDGMENTS

This work was supported by Data Science Program, Faculty of Science and Agricultural Teknologi, Univercity of Muhammadiyah Semarang.

REFERENCE

Commented [M.A9]: add 2 reference from jstatistika

Commented [AN10R9]: oke

[1] H. Rahman, M. Rahmah, and Nur Saribulan, "UPAYA PENANGANAN STUNTING DI INDONESIA," *Jurnal Ilmu Pemerintahan Suara Khatulistiwa (JIPSK)*, vol. VIII, no. 01, pp. 44–59, 2023.

[2] W. H. S. W. H. Organization, " Monitoring Health for the SDGs, World Health Organization," 2022.

[3] A. Muhaimin *et al.*, "Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia Analisis Sosial Media dan Pemodelan Topik: Kasus Studi Stunting di Indonesia," *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.

[4] R. Rizqi Robbi Arisandi, B. Warsito, and A. Rachman Hakim, "Aplikasi Naive Bayes Classifier (NBC) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *Jurnal Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[5] U. R. Gurning, S. F. Octavia, D. R. Andriyani, N. Nurainun, and I. Permana, "Prediksi Risiko Stunting pada Keluarga Menggunakan Naïve Bayes Classifier dan Chi-Square," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 172–180, Jan. 2024, doi: 10.57152/malcom.v4i1.1074.

[6] A. A. Reza1 and M. S. Rohman2, "JITE (Journal of Informatics and Telecommunication Engineering) Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization," *JITE*, vol. 7, no. 2, 2024, doi: 10.31289/jite.v7i2.10628.

[7] A. Muhaimin *et al.*, "Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia Analisis Sosial Media dan Pemodelan Topik: Kasus Studi Stunting di Indonesia," *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.

[8] M. Banurea, D. Betaria Hutagaol, and O. Sihombing, "KLASIFIKASI PENYAKIT STUNTING DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST," *Jurnal TEKINKOM*, vol. 6, no. 2, 2023, doi: 10.37600/tekinkom.v6i2.927.

[9] Stephenie, B. Warsito, and A. Prahutama, "Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method," in *E3S Web of Conferences*, EDP Sciences, Nov. 2020. doi: 10.1051/e3sconf/202020216006.

[10] M. Ali Fauzi, A. Z. Arifin, S. C. Gosaria, and I. S. Prabowo, "Indonesian news classification using naïve bayes and two-phase feature selection model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 8, no. 3, pp. 610–615, Dec. 2017, doi: 10.11591/ijeecs.v8.i3.pp610-615.

[11] Bahrawi, "Sentiment Analysis Using Random Forest Algorithm Online Social Media Based," Makassar, Dec. 2019.

[12] G. Khanvilkar and D. Vora, "Sentiment analysis for product recommendation using random forest," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 3, pp. 87–89, 2018, doi: 10.14419/ijet.v7i3.3.14492.

[13] L. Breiman, "Random Forests," 2001.
[14] Leo Breiman and A. Cutler, " Manual on Setting Up, Using, and Understanding Random Forest 4.0," 2003.
[15] M. A. Fauzi, "Random forest approach fo sentiment analysis in Indonesian language," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 1, pp. 46–50, Oct. 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
[16] I. Afdhal *et al.*, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.

J Statistika Vol. 17, No. 1, (2024), Hal. 645 - 654 jurnal.unipasby.ac.id

Sentiment Analysis Of Public Opinion On Handling Stunting In Indonesia Using Random Forest

Ariska Fitriyana Ningrum (1), Ihsan Fathoni Amri (2)

[1,2]Program Studi Sains Data, Universitas Muhammadiyah Semarang

Jln. Kedungmundu No 18, Kedungmundu, Kec. Tembalang, Kota Semarang
e-mail: ariskafitriyana@unimus.ac.id(1), ihsanfathoni@unimus.ac.id(2)

ABSTRAK

Masalah stunting penting untuk diselesaikan, karena berpotensi mengganggu potensi sumber daya manusia dan berkaitan dengan tingkat kesehatan, bahkan kematian anak. Pemerintah Indonesia menargetkan angka stunting turun menjadi 14 persen pada tahun 2024 melalui program percepatan penurunan stunting sebagai upaya meningkatkan status gizi masyarakat dan juga menurunkan prevalensi stunting atau balita pendek. Memahami sentimen publik terhadap inisiatif stunting sangat penting bagi para pembuat kebijakan dan pemangku kepentingan untuk merancang intervensi yang efektif dan mengalokasikan sumber daya secara efisien. Pada penelitian ini dilakukan klasifikasi pada sentiment positif dan negatif menggunakan algoritma random forest. Data yang digunakan adalah data komentar pada salah satu laman media sosial yaitu twitter mengenai sentiment masyarakat terhadap penanganan kasus stunting di Indonesia. Tahapan pertama pada penelitian ini setelah didapatkan sebuah data yaitu dilakukan prepocessing data. Tahapan preprocessing data dalam analisis sentimen berguna untuk membersihkan dan menormalkan teks, menghilangkan kata-kata tidak relevan, serta mempersiapkan data agar algoritma dapat menganalisis sentimen dengan lebih akurat dan efisien. Selanjutnya hasil data yang sudah di prepocessing diberikan label 0 untuk positif dan 1 untuk label negatif. Klasifikasi terhadap sentiment positif dan negatif ini dilakukan menggunakan random forest dan menghasilkan nilai akurasi sebesar 97,5%. Model ini sudah baik, namun kami menyarankan untuk mencoba algoritma lain dalam penelitian selanjutnya.

Kata kunci: Analisis Sentiment, Random Forest, Stunting

ABSTRACT

The problem of stunting is important to solve, as it has the potential to disrupt human resource potential and is linked to health outcomes and even child mortality. The Indonesian government targets the stunting rate to drop to 14 percent by 2024 through an accelerated stunting reduction program as an effort to improve the nutritional status of the community and also reduce the prevalence of stunting or short toddlers. Understanding public sentiment towards stunting initiatives is essential for policy makers and stakeholders to design effective interventions and allocate resources efficiently. In this research, classification of positive and negative sentiment is carried out using the random forest algorithm. The data used is comment data on one of the social media pages, namely Twitter, regarding public sentiment towards handling stunting cases in Indonesia. The first stage in this research after obtaining a data is data preprocessing. The data preprocessing stage in sentiment analysis is useful for cleaning and normalizing text, removing irrelevant words, and preparing data so that algorithms can analyze sentiment more accurately and efficiently. Furthermore, the results of the preprocessed data are labeled 0 for positive and 1 for negative labels. The classification of positive and negative sentiment was done using random forest and resulted in an accuracy value of 97.5%. This model is good, but we suggest trying other algorithms in future research.

Keywords: *Sentiment Analyst, Random Forest, Stunting*

INTRODUCTION

Stunting is a condition in children who experience growth disorders, so that the height and weight of children are not normal due to problems of nutritional deficiencies for a long time [1]. The problem of stunting in Indonesia is still quite large in the health sector today. According to the World Health Organization (WHO), as many as 22% or around 149.2 million children in the world under the age of five were recorded as stunted in 2020. Indonesia's position on the prevalence of stunting in the world is ranked 115 out of 151 countries. Meanwhile, in Southeast Asia, Indonesia is ranked second at 31.8% after Timor Leste at 48.8%. The third is Laos at 30.2%, the fourth is Cambodia at 29.9%, and the fifth is the Philippines at 29.9% [2] Stunting is caused by health problems, environmental factors and health services received by children. Genetic factors do not significantly affect stunting. Lack of nutrition in the fetus is the biggest cause of stunting in children. The first 1000 days of a child's life (1000 HPK) is the starting point for making important conclusions on long-term growth Thus, ineffective parenting and diet can increase the chance of stunting. Mental disorders and hypertension in mothers also affect the behavior and practices of nutrition in children. Limited access to health and sanitation services exacerbates the stunting conditions that occur in Indonesia such as lack of clean water, unclean latrines, and so on [3] Stunting in Indonesia is a deep-rooted problem. The problem of stunting is important to solve, because it has the potential to disrupt human resource potential and is related to health levels, even child mortality. In early 2021, the Indonesian government targeted the stunting rate to drop to 14 percent by 2024 through the accelerated stunting reduction program as an effort to improve the nutritional status of the community and also to reduce the prevalence of stunting or short toddlers [4].

Stunting in Indonesia is a deep-rooted problem. The problem of stunting is important to solve, because it has the potential to disrupt human resource potential and is related to health levels, even child mortality. In early 2021, the Indonesian government targeted the stunting rate to drop to 14 percent by 2024 through the accelerated stunting reduction program as an effort to improve the nutritional status of the community and also to reduce the prevalence of stunting or short toddlers[5] Understanding public sentiment towards stunting initiatives is crucial for policymakers and stakeholders to design effective interventions and allocate resources efficiently. Several previous studies have analyzed stunting predictions using the random forest algorithm which resulted in a classification accuracy value of 90.7%. [6]. Another study on social media analysis with the topic of stunting in Indonesia was conducted where the results showed that negative sentiment dominated by 60.6%, positive sentiment by 31.5%, and neutral by 7.9% [7]. In addition, this study shows that 'children', 'decline', 'numbers', 'prevention', and 'nutrition' are words that often appear in stunting [7]. Another study comparing SVM and random forest algorithms for the classification of stunting disease. the results show that the random forest algorithm provides higher accuracy of 88.2% compared to SVM of 65.6% [8].

The background of this study is based on the need to analyze public sentiment regarding the handling of stunting in Indonesia, which is a significant public health problem. The data used are the results of positive and negative reviews from the public on social media such as twitter regarding the handling of stunting cases in Indonesia. Some previous studies have also analyzed using comment data on twitter such as research on Sentiment Analysis of Twitter Netizens on the News of VAT on Basic Food and Education Services with Social Network Analysis and Naive

Bayes Classifier Approaches with data obtained as many as 4090 tweets [9]. While other studies have also conducted sentiment analysis of twitter users regarding online transportation service users [10].

The method used in this research is Random Forest. Random Forest was chosen because of its superior ability to handle complex and varied data, and provide accurate results in classification and prediction. This method is suitable for sentiment analysis because it is able to overcome overfitting, works well on large and irregular datasets, and provides a good interpretation of the features that affect the results. Random Forest is one of the state-of-the-art methods in machine learning that consists of a number of independently trained decision trees and the results are combined to produce more accurate and stable predictions. In the context of sentiment analysis, Random Forest can handle variations in language expression and identify relevant patterns from unstructured text data. In addition, Random Forest's ability to handle data with many features is very useful in sentiment analysis involving various emotional aspects and public opinions related to stunting treatment [11].

The state of the art of the Random Forest method shows that this technique has been successfully applied in various domains, including text and sentiment analysis, with satisfactory results. Previous studies have shown that Random Forest often outperforms other methods such as logistic regression and Support Vector Machines (SVM) in terms of accuracy and robustness to noise in the data. This makes Random Forest an appropriate choice for this research in an effort to understand and measure public sentiment towards stunting response efforts in Indonesia.

METHOD

In this study, a sentiment analysis about stunting in Indonesia was conducted. The remaining data will be analyzed using the random forest algorithm to check the accuracy of the sentiment results. The following are the stages of sentiment analysis research on stunting using the random forest algorithm

1. Data Collection

The data used in this study is the result of crawling Twitter data related to positive and negative responses to stunting conditions in Indonesia. The data collection process was carried out over a one-month period, starting from January 1, 2024 to January 31, 2024, to get a current picture of public sentiment on the issue of stunting. In the crawling process, certain filters and keywords related to infant stunting were used. The keywords used included "stunting", "child growth", and "child nutrition", as well as other keyword variations related to stunting and child health in Indonesia. In applying a series of filters, including language settings (Bahasa Indonesia) and geographical location (Indonesia) were used to ensure that the data captured was specifically related to responses to stunting conditions in Indonesia. This resulted in a total of 4601 comments showing both positive and negative opinions. Labeling the data into positive and negative categories was done through a manual process by the researcher, where each opinion was classified based on the sentiment expressed towards stunting conditions. This assessment is based on the context within each opinion, with the aim of gaining an accurate understanding of public sentiment.

2. Prepocessing Data

The first stage of the system is preprocessing. This stage involves several processes including Case Folding, Tokenization, Normalization, and Stemming. Case Folding is a task of splitting review text into smaller units called tokens or terms [12]. For infant stunting cases, what is done before and after case folding is, for example, "Breastfeeding mothers must have good nutrition" becomes "breastfeeding mothers must have good nutrition". Next is Tokenizing, in this process the separation is carried out on each word that makes up a document. In general, each word is identified or separated from other words by space characters, so the tokenizing process relies on space characters in the document to perform word separation [13] In sentiment analysis of stunting cases, what is done is to present the number of tokens generated from a review or comment. For example, from the sentence "breastfeeding mothers should have good nutrition", the tokens generated are "mother", "breastfeeding", "should", "nutrition", "which", "good". Normalization (Stopword Removal) process Removes special characters, numbers, and stopwords (common words) from each token. In the case of sentiment analysis, it shows a list of stopwords used and examples of text before and after stopword removal. For example, from "mother", "breastfeeding", "should", "nutrition", "which", "good", after removing the stopwords "should", "which", then "mother", "breastfeeding", "nutrition", "good" remains. This research also uses Stemming techniques which aim to find the base word, by removing all affixes that are fused to the word.[14] In Indonesian, this usually involves the removal of prefixes, suffixes or infixes. As an example of words before and after the stemming process, for example, "menyusui" can be reduced to "susu".

3. Sentiment Analyst Using Random Forest

The last stage is sentiment classification. Each review will be classified into positive or negative category. In this study, we employ random forest for the classification task. Random forest algorithm is a supervised classification algorithm. It is an ensemble learning technique based on decision tree algorithm [15]. Random Forest Algorithm is the advancement of Classification and Regression Tree (CART) method with the implementation of bootstrap aggregating (bagging) and random feature selection. Procedure of random forest algorithm on the data of n observations and p predictor [16]

a. Random samples of size n are drawn with the possibility of obtaining the same data (with replacement). This phase is called bootstrap.
b. Using the bootstrap samples, the tree is grown until the maximum size is reached, which is done without pruning. At each node, the random feature selection is used to determine the split, which m number of variables randomly sampled as candidates at each split must be m << p, at which point, the best node will be chosen based on m number of variables available for splitting [17]
c. Repeat stage 1 and 2 for k times to generate a forest that consists of k trees.

Breiman and Cutler suggests to observe the error OOB when

$$m = \left(\frac{1}{2}\left|\sqrt{p}\right|, \left|\sqrt{p}\right|, 2\left|\sqrt{p}\right|\right) \quad (1)$$

where *p* is the total variable and the number of *k* is small, then m with the smallest error OOB will be chosen.

In order to determine the split used as root node/node, Gini Index is used in Random Forest method. The formula of Gini index can be described as following:

$$\boldsymbol{Gini = 1 - \sum_{i=1}^{k} p_i^2} \quad (2)$$

where:

p_i : probability of an attribute being classified to class i.

k : total number of attributes being classified to a particular class.

The number of k suggested to apply in bagging is k = 50 which will provide satisfied results for classification [18]

RESULT AND DISCUSSION

In this study, the process of labeling public opinion data on stunting in Indonesia, obtained from social media through crawling, was carried out using a manual approach. Initially, a portion of the data was manually labeled by the research team to form the training dataset. Each opinion was classified into two categories: 0 for negative sentiment and 1 for positive sentiment. This manual labeling process is important to ensure that the data used accurately reflects public sentiment towards stunting and to generate a valid classification model.

Label	komen bersih
0	dengar juta anak indonesia derita terapi efisi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...
0	anak susah makan momok banget butuh nutrisi an...
0	kuwu tatangga desa nu
0	rakyat indonesia sehat ongkos juta kondisi sak...
...	...
0	hamil minum puskesmas kasih gratis takut anemi...
0	iya cupu cupu paham google search baca jurnal ...
0	banteng jawa barat komitmen entas pdiperjuanga...
0	bahan becandaan
0	tingkat sabar setip tissue hadap anak emosi mu...

Figure 1. Review data that has been labelled

After obtaining a dataset of community reviews on stunting handling, data preprocessing is carried out with various stages such as *case folding, tokenizing, normalization, and stemming*. Data preprocessing is important because it helps to optimally clean, prepare, and organize data before analysis, thereby improving the accuracy, reliability, and interpretation of results from the model or analysis technique used.

The first stage, namely preprocessing, aims to homogenize all text into lowercase letters. The results of the case folding process are as follows.

2 Data Pre-Processing¶

- Tahap Case Holding => pada tahapan ini bertujuan untuk menyeragamkan seluruh teks kedalam huruf kecil (lowercase)

```
# Case Folding: Mengubah semua teks menjadi huruf kecil
data['komen case folding'] = data['clean_teks'].str.lower()

print("\n Data Setelah Case Folding:")
(data.head(10))
```

clean_teks	komen case folding
dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...
bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...
anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...
kuwu tatangga desa nu	kuwu tatangga desa nu
rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...

Figure 2. Case Holding Stages Result

Next, the tokenizing process is carried out. At this stage, the sentence in the comment will be broken down into words. The results of the tokenizing process are presented as in the following figure.

- Tokenizing proses dilakukannya pemecahan kata pada kalimat. Pada tahapan ini akan dilakukannya pemecahan dari kalimat di komentar menjadi kata per kata.

```
] import nltk
  from nltk.tokenize import word_tokenize
  nltk.download('punkt')
  data = data.dropna(subset=['komen case folding'])
  data['komen tokenized'] = data['komen case folding'].apply(word_tokenize)
  print("\nData Setelah Tokenizing:")
  (data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]
0	rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...	[rakyat, indonesia, sehat, ongkos, juta, kondi...

Figure 3. Tokenizing Stages Result

The next preprocessing is to perform normalization to change the values of a dataset so that they have a uniform scale. The main purpose is to ensure that variables with different value ranges have equal influence when used in the analysis. The results of data normalization are as follows.

3. Normalization(Stopword Removal) proses Menghapus karakter khusus, angka, dan stopwords (kata-kata umum) pada setiap token.

```
[ ] from nltk.corpus import stopwords
    import re
    nltk.download('stopwords')

    stop_words = set(stopwords.words('indonesian'))

    data['komen normalized'] = data['komen tokenized'].apply(lambda tokens: [word for word in tokens if word.isalpha() and word not in stop_words]

    print("\nData Setelah Normalization:")
    (data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized	komen normalized
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...	[anak, susah, makan, momok, banget, butuh, nut...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]

Figure 4. Normalization Stages Result

The last step in data preprocessing is stemming. This aims to find the base word, by removing all affixes that are fused to the word. The results of stemming performed for sentiment analysis are as follows.

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()
stemmer = factory.create_stemmer()
data['komen stemming'] = data['komen normalized'].apply(lambda tokens: [stemmer.stem(word) for word in tokens])

print("\nData Setelah Stemming:")
(data.head(10))
```

Label	clean_teks	komen case folding	komen tokenized	komen normalized	komen stemming
0	dengar juta anak indonesia derita terapi efisi...	dengar juta anak indonesia derita terapi efisi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...	[dengar, juta, anak, indonesia, derita, terapi...
0	bilang sisa rakyat komen lucu lucu gagal dulan...	bilang sisa rakyat komen lucu lucu gagal dulan...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...	[bilang, sisa, rakyat, komen, lucu, lucu, gaga...
0	anak susah makan momok banget jawab butuh nutr...	anak susah makan momok banget jawab butuh nutr...	[anak, susah, makan, momok, banget, jawab, but...	[anak, susah, makan, momok, banget, butuh, nut...	[anak, susah, makan, momok, banget, butuh, nut...
0	kuwu tatangga desa nu	kuwu tatangga desa nu	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]	[kuwu, tatangga, desa, nu]
0	rakyat indonesia sehat ongkos juta kondisi sak...	rakyat indonesia sehat ongkos juta kondisi sak...	[rakyat, indonesia, sehat, ongkos, juta, kondi...	[rakyat, indonesia, sehat, ongkos, juta, kondi...	[rakyat, indonesia, sehat, ongkos, juta, kondi...

Figure 5 Stemming Performed for sentiment analysis

Furthermore, the results of data preprocessing that have been carried out can be seen visually regarding positive and negative opinions. Visualization aims to display the words that appear most or most often in a sentiment. Wordcloud this time describes each sentiment, the more often a word is used when giving a review, the larger the size of the word displayed on the wordcloud visualization. The following figure shows the visualization results for positive and negative sentiments

Figure 6. Visualization Positive Sentiment

Based on the figure above, it can be seen that in the positive sentiment there are several words that stand out such as, "balanced nutrition,", "helps reduce", "Welfare Growth" and several other words which indicate that the public's response to handling stunting in Indonesia has helped reduce stunting rates, provide balanced nutrition to children and can foster community welfare.

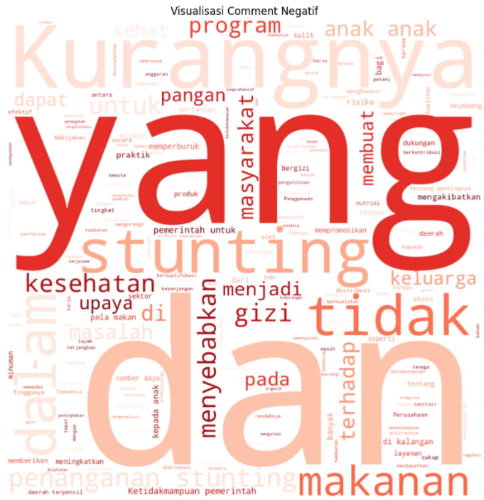

Figure 7. Visualization Negative Sentiment

Based on the figure above, it can be seen that in the negative sentiment regarding the handling of stunting in Indonesia, there are prominent words such as "lack of", "government inability", "gap" and several other words that indicate that the handling of stunting is still lacking, there is a social gap and the lack of government in handling cases of stunting in the society. Next, Classification on positive and negative comment data that has been done feature extraction using TF-IDF function is to convert text into a numerical vector representation based on the frequency of occurrence of words in it, taking into account the TF-IDF weight of each word. The classification metric evaluation calculations that we use are confution matrix, F1 measure and accuracy. The following are the results of the random forest performance evaluation

Table 1 Confution Matrix

		Actual	
		Negative	Positive
Prediction	Negative	893	0
	Positive	23	5

Based on table 1, which is related to the confusion matrix of prediction results, TP = 5, FP = 23, TN = 893, FN = 0, Total data = 258. The calculation results are obtained from the following calculations:

Table 2. Precision, recall, and f1-score

	Precision	Recall	f1-score
Negative	0.97	1.00	0.99
Positive	1.00	0.18	0.30

The performance generated by the random forest algorithm provides considerable accuracy, which is 97.50%, indicating that this model can classify data has a very good indication, and produces precision on Label 0 (negative comments) of 97% and recall of 100%, the results obtained are very high, and F1 score of 99%, indicating a high balance of precision and Recall. Meanwhile for precision on Label 1 (Positive comments) of 100%, and recall of 18% and the result for f1-score is 30%.

CONCLUSION

This study shows that public sentiment towards the handling of stunting cases in Indonesia can be divided into positive and negative based on the analysis of 4601 comments from Twitter social media. The results show that positive responses include the view that the handling of stunting has succeeded in reducing stunting rates, providing balanced nutrition to children, and potentially improving the general welfare of society. On the other hand, negative responses include dissatisfaction with the effectiveness of stunting handling, the existence of unresolved social inequalities, and the lack of effort from the government in handling stunting cases in the community.

This research continued with the classification of comment data based on sentiment using TF-IDF feature extraction. This method is important because it converts text into a numerical vector representation, where the TF-IDF weight of each word gives an idea of the importance of the word in determining positive or negative sentiment. Through this classification, it is possible to identify and categorize the sentiments present in the text data, enabling a deeper understanding of the public's views on stunting in Indonesia.

We performed sentiment analysis using random forest algorithm and achieved about 97.5% accuracy. I would recommend you to try using some other machine learning algorithms such as LSTM or KNN and see if you can get better results.

ACKNOWLEDGMENTS

This work was supported by Data Science Program, Faculty of Science and Agricultural Teknologi, Univercity of Muhammadiyah Semarang.

REFERENCE

[1] H. Rahman, M. Rahmah, and Nur Saribulan, "UPAYA PENANGANAN STUNTING DI INDONESIA," *Jurnal Ilmu Pemerintahan Suara Khatulistiwa (JIPSK)*, vol. VIII, no. 01, pp. 44–59, 2023.

[2] W. H. S. W. H. Organization, " Monitoring Health for the SDGs, World Health Organization," 2022.

[3] A. Muhaimin *et al.*, "Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia Analisis Sosial Media dan Pemodelan Topik: Kasus Studi Stunting di Indonesia," *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.

[4] R. Rizqi Robbi Arisandi, B. Warsito, and A. Rachman Hakim, "Aplikasi Naive Bayes Classifier (NBC) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *Jurnal Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[5] U. R. Gurning, S. F. Octavia, D. R. Andriyani, N. Nurainun, and I. Permana, “Prediksi Risiko Stunting pada Keluarga Menggunakan Naïve Bayes Classifier dan Chi-Square,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 172–180, Jan. 2024, doi: 10.57152/malcom.v4i1.1074.

[6] A. A. Reza1 and M. S. Rohman2, “JITE (Journal of Informatics and Telecommunication Engineering) Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization,” *JITE*, vol. 7, no. 2, 2024, doi: 10.31289/jite.v7i2.10628.

[7] A. Muhaimin *et al.*, “Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia Analisis Sosial Media dan Pemodelan Topik: Kasus Studi Stunting di Indonesia,” *Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.

[8] M. Banurea, D. Betaria Hutagaol, and O. Sihombing, “KLASIFIKASI PENYAKIT STUNTING DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST,” *Jurnal TEKINKOM*, vol. 6, no. 2, 2023, doi: 10.37600/tekinkom.v6i2.927.

[9] J. A. Nursiyono and C. Chotimah, “Analisis Sentimen Netizen Twitter terhadap Pemberitaan PPN Sembako dan Jasa Pendidikan dengan Pendekatan Social Network Analysis dan Naive Bayes Classifier,” *Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 14, pp. 52–58, 2021.

[10] “Penerapan Text Mining pada Analisis Sentimen Pengguna Twitter Layanan Transportasi Online Menggunakan Metode Density Based Spatial Clustering of Applications With Noise (DBSCAN) dan K-Means,” *Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 15, pp. 184–194, 2022.

[11] Stephenie, B. Warsito, and A. Prahutama, “Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method,” in *E3S Web of Conferences*, EDP Sciences, Nov. 2020. doi: 10.1051/e3sconf/202020216006.

[12] M. Ali Fauzi, A. Z. Arifin, S. C. Gosaria, and I. S. Prabowo, “Indonesian news classification using naïve bayes and two-phase feature selection model,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 8, no. 3, pp. 610–615, Dec. 2017, doi: 10.11591/ijeecs.v8.i3.pp610-615.

[13] Bahrawi, “Sentiment Analysis Using Random Forest Algorithm Online Social Media Based,” Makassar, Dec. 2019.

[14] G. Khanvilkar and D. Vora, “Sentiment analysis for product recommendation using random forest,” *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 3, pp. 87–89, 2018, doi: 10.14419/ijet.v7i3.3.14492.

[15] L. Breiman, “Random Forests,” 2001.

[16] Leo Breiman and A. Cutler, “ Manual on Setting Up, Using, and Understanding Random Forest 4.0,” 2003.

[17] M. A. Fauzi, “Random forest approach fo sentiment analysis in Indonesian language,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 1, pp. 46–50, Oct. 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.

[18] I. Afdhal *et al.*, “Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.

BUKTI PENERIMAAN

JURNAL ILMIAH TEORI DAN APLIKASI STATISTIKA
Program Studi Statistika, Universitas PGRI Adi Buana Surabaya
ISSN 2089-0028 e-ISSN 2654-7511,
Email : jstat@unipasby.ac.id
http://jurnal.unipasby.ac.id/index.php/jstatistika

SURAT PENERIMAAN

Nomor : 41/J-Statistika/VII/2024
Hal : Informasi Artikel Diterima

Yth,
Ariska Fitriyana Ningrum

Yang bertanda tangan di bawah ini, Editor Jurnal J-Statistika (Jurnal Ilmiah Teori dan Aplikasi Statistika) Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas PGRI Adi Buana Surabaya menerangkan bahwa artikel yang ditulis oleh:

Penulis : Ariska Fitriyana Ningrum (1), Ihsan Fathoni Amri (2)
Lembaga : Program Studi Sains Data, Universitas Muhammadiyah Semarang
Judul Artikel : Sentiment Analysis Of Public Opinion On Handling Stunting In Indonesia Using Random Forest

Diterima dan sedang dalam proses layout editing dan akan diterbitkan dalam Jurnal **J-Statistika** Periode Juli 2024 Volume 17 Nomor 1. Selanjutnya, penulis dimohon untuk segera melakukan pembayaran sebesar Rp. 500.000,- untuk biaya publikasi jurnal ke:

No Rekening : 0691974088
Bank : BNI
Atas Nama : Muhammad Athoillah

Serta mengirim bukti pembayaran ke kontak editor (WA) +62 856 453 583 46. Atas perhatian dan kerjasamanya, kami ucapkan terimakasih.

Surabaya, 5 Juli 2024
Chief Editor

Jurnal Ilmiah Teori Dan Aplikasi Statistika
ADI BUANA SURABAYA

Muhammad Athoillah, S.Si, M.Si