

JLK - cek.01

by Firman Noor Hasan, Fajar Sidik

Submission date: 02-Jul-2022 01:10PM (UTC+0700)

Submission ID: 1865681003

File name: Hasan,_Fajar_-_cek.01.doc (242K)

Word count: 2713

Character count: 14394

ANALYSIS SENTIMENT OF COMMUNITY RESPONSE ON COOKING OIL PRICE INCREASE POLICY WITH NAÏVE BAYES CLASSIFIER ALGORITHM

Fajar Sidik^{#1}, Prista Afikah^{#2}, Firman Noor Hasan^{#3}

[#] *Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka
Jl. Tanah Merdeka No.6, Kec. Pasar Rebo, DKI Jakarta 13830, Indonesia*

¹fjrsk@gmail.com

²Pristaafika1112@gmail.com

³firman.noorhasan@uhamka.ac.id

Abstrak— Cooking oil is a basic need for Indonesian people. Indonesia experienced a shortage of oil in March 2022. This has become a hot conversation on Twitter social media last March, many people think positively or negatively. But behind it all there are different assessments of the parties who feel the pros and cons, various parties have different points of view. In this article, we conduct a sentiment analysis on the public's response to the scarcity of cooking oil using a dataset obtained from the Twitter digital platform. This article aims to classify tweets related to the scarcity of cooking oil into positive and negative sentiments using a machine learning strategy using the Naive Bayes method. This algorithm was chosen to make it easier for the public to make choices and to know the level of accuracy of the method, where the level of accuracy obtained from the naive Bayes classifier method 72%.

Kata kunci— sentiment analysis, cooking oil, naive bayes classifier

I. INTRODUCTION

Cooking oil is the main ingredient in food processing [1]. Cooking oil functions as a heat conductor, the need for which is increasing day by day with the increasing demand [2]. Cooking oil can add health benefits and calories as well as certain flavors to dishes [3]. Therefore, food processing using cooking oil is preferred over other processing [4]. Cooking oil that is sold in the community has two types of oil, namely cooking oil that uses packaging and bulk [5]. The difference between the two lies in the filtering which affects the quality of the cooking oil [6]. The difference between the two occurs in the filtering process where packaged cooking oil is filtered twice while the bulk is only filtered once [7].

The increase in the price of cooking oil in Indonesia creates pros and cons in society, in addition to the rising price of cooking oil, it has become scarce in supermarkets and traditional markets [8]. This is due to an increase in demand in

the community but the stock of cooking oil cannot meet the demand of the community because it is stockpiled by several elements [9].

Sentiment analysis is a technique to understand, extract and handle data automatically, this technique is completed to get sentiment data that is composed of an opinion [10]. In some studies, sentiment analysis is included in the big data category, the size of text data is also growing, so it has various meanings in the context [11]. The analysis stages include tweets, review texts, blogs, forums, and preprocessing data which includes, stemming, deletion, sentiment identification, tokenization process, and sentiment classification [12]. Text mining is a technique for grouping documents, clustering, information extraction, sentiment analysis and information retrieval [13]. Text processing should be possible with several stages in preprocessing to be completed to the next grouping stage [14].

Thus, this sentiment analysis aims to determine the public's response, especially Twitter social media users, to the scarcity policy and the ongoing increase in cooking oil prices. This study uses data crawling with the Tweepy library on python software. Research for classification uses the Textblob library in python software.

II. METHODOLOGY

The research method is a guideline for the stages of the flow of research or steps so that research has appropriate results from the initial objectives, the design of this research flow is as shown in Figure 1.

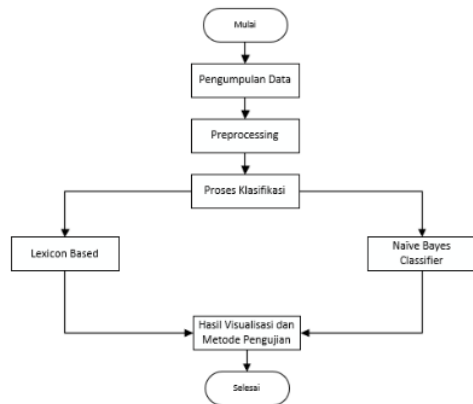


Figure 1. Research Flow

In this study, the Naïve Bayes Classifier method was used in the classification process carried out.

A. Tweet Data Collection

The first step in collecting data is creating a Twitter Development account which has a specific function to process data on a Twitter account. After the account creation process is successful, the next step is for researchers to obtain Key Tokens and Key Access which have a function as an Application Programming Interface (API) by accessing <https://developer.twitter.com>. Then do the crawling coding to get the required dataset. Coding is done in python software.

The crawling data comes from Twitter social media tweets regarding "scarcity of cooking oil". Figure 2 is the stage of data crawling.



Figure 2. Data Crawling Stage

B. Preprocessing Data

At this stage, the data that has been collected is then processed, which is usually called preprocessing. At this stage, it is expected to be able to make the text sentences form ideal sentences before the implementation stage is carried out. This stage is related to several stages including Case Folding, Tokenizer, Filtering, Cleansing and also Stemming.

1. Naïve Bayes Classifier

Naïve Bayes Classifier is a classification method that can be used to predict the probability of class membership. Naive Bayes Classifier is also classified as an algorithm that is easy and simple to use and has

high accuracy results in estimating an event. Below is the Naïve Bayes Classifier equation.

$$P(X|Y) = \frac{P(X|Y) \times P(X)}{P(Y)} \quad (1)$$

Where :

X = provisional estimates of data from a specific class

Y = Data with unknown class

P(X|Y) = Estimated probability of X with terms Y (Posterior probability)

P(X) = Estimated probability X (prior probability)

P(Y|X) = Probability of estimating Y with X

P(Y) = Probability Y

Information :

Posterior probability: there may be class X

Prior probability : probability of the initial sample of class Y

III. RESULTS AND DISCUSSION

This sentiment analysis regarding the scarcity of oil uses the Naïve Bayes Classifier method. With the aim of comparing with previous research and increasing accuracy.

A. Tweet Data Collection

The collection of tweet data with the Twitter API Key is 1000 tweets data from January 1, 2022 to June 1, 2022 with the keyword "Scarcity of Cooking Oil". The data obtained is converted into tabular form to make it easier to carry out the next process. contains three attributes, namely:

1. Username: username contains the username of the account that created the tweets obtained
2. Tweetcreatedts : tweetcreatedts contains data when tweets were created
3. Text: contains tweets or opinions commonly called tweets from Twitter user accounts.

B. Preprocessing

Preprocessing is a stage that includes cleaning, case folding, tokenization, filtering and stemming. This stage is carried out to clean crawled data, such as removing punctuation marks, symbols, hashtags, , HTM, mentions and other symbols that are not relevant. In this preprocessing stage, the crawled tweets data are converted as a whole into lowercase letters, cut a sentence into word chunks, remove and change stacked words into basic words.

1. Cleaning

This stage is the stage of removing some irrelevant punctuation marks, including commas, exclamation marks, question marks, , mentions, URLs and hashtags. Table I is the result data from the cleansing stage.

Table 1. Cleaning Result Data

No	Before	After
1	Yg tribun aja bisa di	Yg tribun aja bisa di

No	Before	After
	borong miliaran..gak heran klo minyak goreng langka.. Tetiba ada yg bagikan berton ton https://t.co/diAJrhDobt	borong miliaran gak heran klo minyak goreng langka Tetiba ada yg bagikan berton ton
2	@FOODFESS2 pilih jokowi banyak gilir jokowi bohong lupa janji suara dengar pilih jokowi protes minyak goreng langka	pilih jokowi banyak gilir jokowi bohong lupa janji suara dengar pilih jokowi protes minyak goreng langka.
3	"Pak Jokowi Bilang Dalam 2 Minggu Harga Minyak Goreng Curah Akan Kembali Ke 14RB/L" Nyatanya Tidak! PHP Lagi! lewat Selengkapnya https://t.co/7ch5fMM4bY Mohon ijin tag https://t.co/B6rkIjpkov	Pak Jokowi Bilang Dalam 2 Minggu Harga Minyak Goreng Curah Akan Kembali Ke 14RB L Nyatanya Tidak PHP Lagi lewat Selengkapnya Mohon ijin tag

2. Tokenization

Tokenization is a library of NLTK used in this research. The tokenization stage has a function to break sentences from tweets into words, Table II is the result of the tokenization stage.

Table 2. Tokenization Result Data

No	Before	After
1	Yg tribun aja bisa di borong miliaran gak heran klo minyak goreng langka Tetiba ada yg bagikan berton ton	['Yg', 'tribun', 'aja', 'bisa', 'di', 'borong', 'miliaran', 'gak', 'heran', 'klo', 'minyak', 'goreng', 'langka', 'Tetiba', 'ada', 'yg', 'agikan', 'berton', 'ton']
2	pilih jokowi banyak gilir jokowi bohong lupa janji suara dengar pilih jokowi protes minyak goreng langka	['pilih', 'jokowi', 'banyak', 'gilir', 'jokowi', 'bohong', 'lupa', 'janji', 'suara', 'dengar', 'pilih', 'jokowi', 'protes', 'minyak', 'goreng', 'langka']
3	Pak Jokowi Bilang Dalam 2 Minggu Harga Minyak Goreng Curah Akan Kembali Ke 14RB L Nyatanya Tidak PHP Lagi lewat Selengkapnya Mohon ijin tag	['Pak', 'Jokowi', 'Bilang', 'Dalam', '2', 'Minggu', 'Harga', 'Minyak', 'Goreng', 'Curah', 'Akan', 'Kembali', 'Ke', '14RB', 'L', 'Nyatanya', 'Tidak', 'PHP', 'Lagi', 'lewat', 'Selengkapnya', 'Mohon', 'ijin', 'tag']

3. Case Folding

At this stage, case folding functions to change the letters contained in the tweets data into lowercase or lowercase letters so that they are easier to read by computers

4. Filtering

At this stage the filtering used comes from the NLTK library in Indonesian to make it easier to remove unnecessary word times to reduce noise which makes the data cleaner.

5. Stemming

In this stemming stage, the library used is the stemmer factory library to facilitate the process. The goal is to make each tweet a root word and remove the affixes from tweets.

In Table III below is a comparison before and after the data passed the case folding, filtering and stemming stages.

Table 3. Case Folding, Filtering, Stemming Result Data

No	Before	After
1	Yg tribun aja bisa di borong miliaran gak heran klo minyak goreng langka Tetiba ada yg bagikan berton ton	tribun aja borong miliar gak heran minyak goreng langka tetiba bagi
2	pilih jokowi banyak gilir jokowi bohong lupa janji suara dengar pilih jokowi protes minyak goreng langka	jokowi banyak jokowi bohong lupa janji suara dengar jokowi protes minyak goreng langka
3	Pak Jokowi Bilang Dalam 2 Minggu Harga Minyak Goreng Curah Akan Kembali Ke 14RB L Nyatanya Tidak PHP Lagi lewat Selengkapnya Mohon ijin tag	jokowi bilang minggu harga minyak goreng curah

C. Translate

After the data has passed the preprocessing stage, the next step is to process the data with the library translator. Table IV below is the results before and after the data is done.

Table 4. Translated Result Data

No	Before	After
1	tribun aja borong miliar gak heran klo minyak goreng langka tetiba bagi ton ton	The tribune just buys billions, it's not surprising that rare cooking oil suddenly gives tons of tons

No	Before	After
2	pilih jokowi banyak gilir jokowi bohong lupa janji suara dengar pilih jokowi protes minyak goreng langka	vote for Jokowi many times, Jokowi lied, forgot to make an appointment to hear, Vote for Jokowi, protested against rare cooking oil
3	jokowi bilang minggu harga minyak goreng curah rb php lengkap ijin tag	Jokowi said sunday the price of bulk cooking oil rb php complete with tag permits

D. Sentiment Analysis and Classification

After performing the steps carried out on the clean tweets data, the next step is to classify, in this study the classification is using the Naive Bayes Classifier method.

1. Naive Bayes Classifier

The grouping method that uses simple probabilities is rooted in Bayes' theorem and has a large independent opinion (not dependent) on each condition [15].

Table 5. Data Classification Naive Bayes Classifier

No	Tweet	Klasifikasi	Klasifikasi Naive Bayes
1	photo of market subsidized cooking oil price socialization	Netral	Netral
2	vote for Jokowi many times, Jokowi lied, forgot to make an appointment to hear, Vote for Jokowi, protested against rare cooking oil	Positif	Positif
3	shadow of rare cooking oil at the eastern end of Indonesia, the taste is suddenly fixed, the price will take time tomorrow will be abundant, there is a cooking oil mafia with regional structures	Positif	Positif

E. Visualization

After all stages are completed, the next stage is visualization, the output of this research visualization is in the form of a histogram which displays the percentage of each

class. Figure 6 is a histogram display as the output of this analysis visualization.

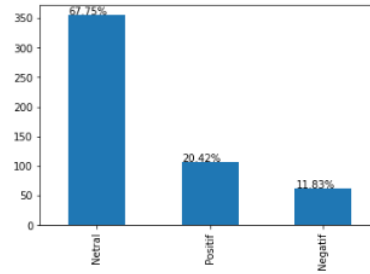


Figure 3. Histogram Visualization Results Display

F. Algorithm Implementation

The next stage is the implementation of the algorithm. At this stage, training and testing of the data is carried out, where the output produced is a prediction algorithm. The results of the algorithm are very important because they are an indication that the method used is correct or not while simultaneously testing the accuracy of the algorithm. The following are the results of the training and testing that have been carried out.

	Positif	Netral	Negatif
True Positif (TP)	98	232	53
False Positif (FP)	10	143	3
False Negative (FN)	75	5	76

Figure 4. Confusion Metric Results

G. Performance Metric

At this stage, testing is carried out using the sklearn library to assist the process in displaying the classification report. In building a machine learning model, measures or metrics are commonly used to find out how well the model is performing. Precision, recall, f1-score, macro avg and weighted avg were used to evaluate this model. Figure 4 is the result of the report from the Naive Bayes Classification method.

	precision	recall	f1-score	support
Negatif	0.95	0.46	0.62	114
Netral	0.64	0.98	0.78	237
Positif	0.91	0.57	0.70	173
accuracy			0.73	524
macro avg	0.83	0.67	0.70	524
weighted avg	0.80	0.73	0.72	524

Figure 5. Classification Report Results Naive Bayes Classifier Method

IV. CONCLUSION

This research is an analysis of sentiment regarding public opinion about the scarcity of cooking oil on the digital twitter platform using the naive Bayes classifier method. The crawling data resulted from deviations amounting to 1000 tweets data, after doing several processes such as preprocessing and the next stage testing the naive Bayes classifier method with the total data obtained is 523 tweet data. The accuracy result of the naive Bayes classifier is 72%. With this accuracy, it shows that sentiment analysis using the naive Bayes classifier method has a high accuracy value.

REFERENCE

- [1] N. F.A and E. G, "Analisa kadar asam lemak bebas pada minyak jelanta yang diperjualbelikan pada penjual goreng di kecamatan medan helvetia secara alkalimetri 1," *Anal. Lab. Med.*, vol. 4, no. 1, pp. 1–3, 2019.
- [2] S. . . Firly and A. Nurrahmah, "Analisis perbandingan penggunaan minyak curah dan minyak kemasan menggunakan uji hipotesis dua proporsi," *Bull. Appl. Ind. Eng. Theory*, vol. 2, no. 1, pp. 65–66, 2019.
- [3] D. M. Siriza, Rusdina R, R. Alfah, A. Perdana Windarto, and A. Wanto, "Analisis Metode ELECTRE Pada Pemilihan Produk Minyak Goreng Kemasan Terbaik Berdasarkan Konsumen." *Sci. Inf. J.*, vol. 5, no. 2, pp. 129–135, 2019, [Online]. Available: <http://doi.org/10.22216/jsi.v5i2.4095>.
- [4] . B. and A. Rasmito, "B. Pemanfaatan Kunyit Untuk Meningkatkan Kualitas Minyak Goreng Curah," *J. Tek. Ind. dan Kim.*, vol. 1, no. 1, pp. 7–15, 2018, doi: 10.54980/jtik.v1i1.59.
- [5] G. Franko, J. Tarigan, and L. J. Pinem, "Agriprimatech Vol. 4 No. 2, April 2021," vol. 4, no. 2, pp. 57–64, 2021.
- [6] S. Dewi, A. N. Syahputri, K. Deni, A. P. Sari, and I. S. Damanik, "Sistem Pendukung Keputusan Penentuan Produk Minyak Goreng Kemasan dikalangan Masyarakat dengan Metode MFEP (Multi Faktor Evaluation Process)," pp. 577–580, 2021.
- [7] N. A. P. NICO and R. Azara, "Comparative of the Quality of Cooking Oil With Four Times Frying on Packaged and Bulk Cooking Oil," *J. Trop. Food Agroindustrial Technol.*, vol. 2, no. 01, pp. 9–14, 2021, doi: 10.21070/jtft.v2i01.1576.
- [8] G. Widjaja, "HILANGNYA MINYAK GORENG DARI PASAR," file:///C:/Users/Fajar_Sidik/Downloads/675-729-1-PB.pdf, vol. 2, pp. 1–11, 2022.
- [9] Annisa Fitri, Fadila Margasaty, Kusmaria, Rini Desfaryani, and Vivi Utami Dewi, "Peramalan Harga Minyak Goreng Di Tengah Pandemi Covid-19 Kota Bandar Lampung," *dwijenAGRO*, vol. 10, no. 1, pp. 21–26, 2020, doi: 10.46650/dwijenagro.10.1.859.26.
- [10] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, "Analisis Sentimen Terhadap Pembelajaran Daring dengan Algoritma Naive Bayes Classifier," vol. 5, no. 1, pp. 34–43, 2022.
- [11] M. A. Amrustian, W. Widayat, and A. M. Wirawan, "Analisis Sentimen Evaluasi Terhadap Pengajaran Dosen di Perguruan Tinggi Menggunakan Metode LSTM," vol. 6, pp. 535–541, 2022, doi: 10.30865/mib.v6i1.3527.
- [12] W. A. Prabowo and C. Wiguna, "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM," *J. Media Inf. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [13] A. V. Sudiantoro *et al.*, "Analisis Sentimen Twitter Menggunakan Algoritma Mining Dengan," vol. 10, no. 2, pp. 398–401, 2018.
- [14] A. N. Indraini, I. Ernawati, and A. Zaidah, "Analisis Sentimen Terhadap Pembelajaran Daring di Indonesia Menggunakan Support Vector Machine (SVM)," pp. 1–11, 2022.
- [15] A. R. Prananda and I. Thalib, "Sentiment Analysis for Customer Review: Case Study of GO-JEK Expansion," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 1, p. 1, 2020, doi: 10.20473/jisebi.6.1.1-8.

ORIGINALITY REPORT

12%

SIMILARITY INDEX

11%

INTERNET SOURCES

8%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1	inacl.id Internet Source	4%
2	e-journal.unair.ac.id Internet Source	1%
3	jtfat.umsida.ac.id Internet Source	1%
4	www.sinta.ristekbrin.go.id Internet Source	1%
5	ejournal.undwi.ac.id Internet Source	1%
6	prosiding.seminar-id.com Internet Source	1%
7	ur.aeu.edu.my Internet Source	1%
8	Submitted to Loughborough University Student Paper	1%
9	www.jim.unindra.ac.id Internet Source	1%

- | | | |
|----|---|------|
| 10 | N Nasution, P Prajitno, D S Soejoko. "Effectiveness of using computer aided detection based on convolutional neural network for screening microcalcification On USG Mammae", Journal of Physics: Conference Series, 2021
Publication | <1 % |
| 11 | garuda.kemdikbud.go.id
Internet Source | <1 % |
| 12 | Submitted to Universitas 17 Agustus 1945 Surabaya
Student Paper | <1 % |
| 13 | ejurnal.univalabuhanbatu.ac.id
Internet Source | <1 % |
| 14 | core.ac.uk
Internet Source | <1 % |
| 15 | D R Ramdania, B Subaeki, E Nugraha, L Muliawaty, M A Ramdhani. "Decision support system algorithm for the beneficiary of uninhabitable housing funds", IOP Conference Series: Materials Science and Engineering, 2021
Publication | <1 % |
| 16 | Fajar Sidik, Ibnu Suhada, Azhar Haikal Anwar, Firman Noor Hasan. "Analisis Sentimen Terhadap Pembelajaran Daring Dengan | <1 % |

Algoritma Naive Bayes Classifier", Jurnal Linguistik Komputasional (JLK), 2022

Publication

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off