Jurnal Linguistik Komputasional
JLK
Vol 5, No 2
September
2022

Sentiment Analysis of Community Response on Cooking Oil Price Increase Policy with Naïve Bayes Classifier Algorithm

Firman Noor Hasan[#1], Fajar Sidik[#2], Prista Afikah[#3]

[#] *Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka*

Jl. Tanah Merdeka No.6, Kec. Pasar Rebo, DKI Jakarta 13830, Indonesia

[1]firman.noorhasan@uhamka.ac.id

[2]Fjrsdk@gmail.com

[3]Pristaafika1112@gmail.com

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 nave 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 [15].

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.

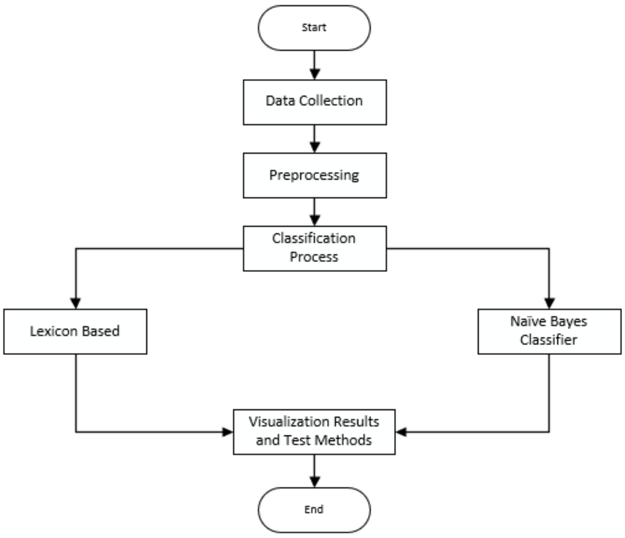

Figure 1. Research Flow

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

A. Tweet Data Collection

The 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 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

the steps taken to get tweet data about the scarcity of cooking oil:

1. The first step in collecting data is creating an account on Twitter Development which has a special function to process data on Twitter accounts.

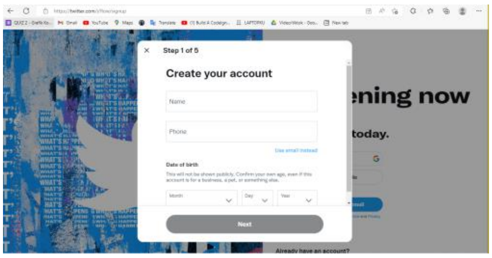

Figure 3. Create Twitter Account

2. After the account creation process is successful, the next step the researcher will get a Key Token and Key Access which has a function as an Application Programming Interface (API) by accessing https://developer.twitter.com.

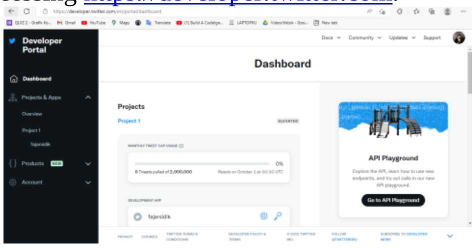

Figure 4. Dashboard Twitter Developer

3. Open google collab and install the required libraries on google collab.

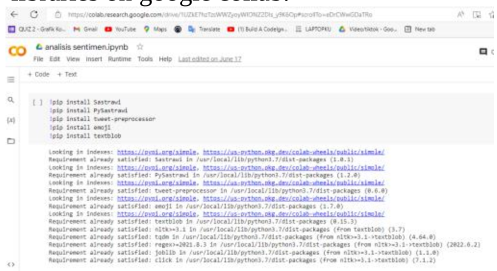

Figure 5. Python Library Installation

4. Call all the libraries that have been installed and also enter the key token and key access that have been obtained previously.

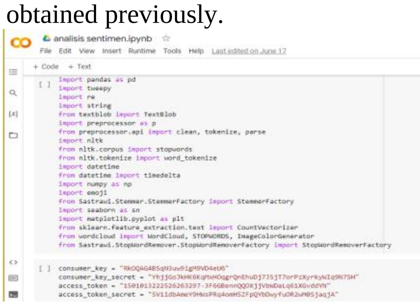

Figure 6. Import Python Library

5. Enter the code for crawling the data. Set words, the start and end dates and the amount of data you want to pull, wait a while and the results of the data pull will appear, then export it into csv form.

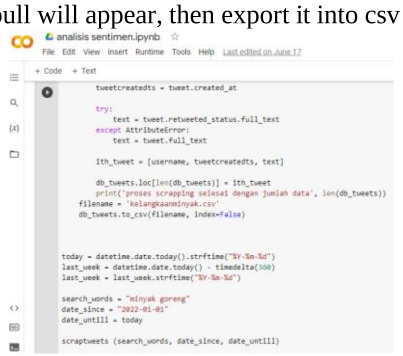

Figure 7. Python Coding to Pull Twitter Data

6. the result of exporting tweet data taken from twitter.

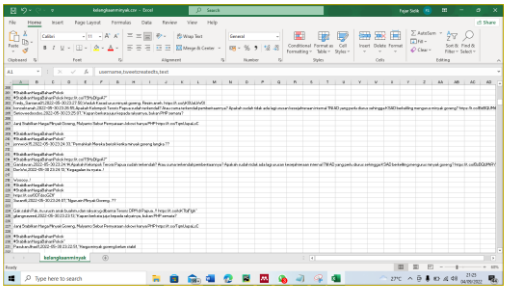

Figure 8. Data Pull Results in Python

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, Tokenizen, Filtering, Cleansing and also Stemming.

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.

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

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 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 [16].

Table 1. Cleaning 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 https://t.co/diAJrhDobt	Yg tribun aja bisa di 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 [17].

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’, ‘bagikan’, ‘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 [18].

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 [19].

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 [20].

   In Table.3 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
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. Naïve 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 [21].

Table 5. Data Classification Naive Bayes Classifier

No	Tweet	Classification	Naïve Bayes Classification
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	Positif	Positif

No	Tweet	Classification	Naïve Bayes Classification
	be abundant, there is a cooking oil mafia with regional structures		

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.

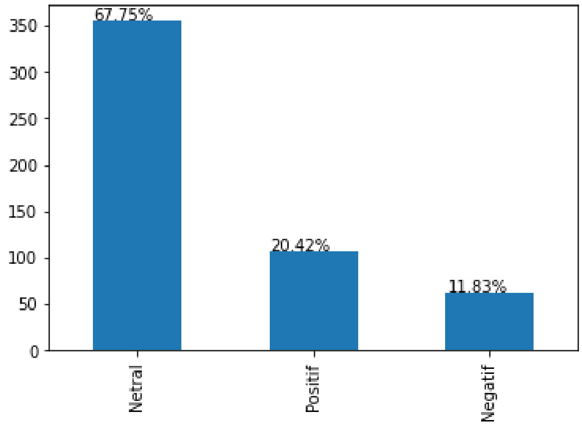

Figure 9. Histogram Visualization Results Display

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.

Table 6. Confusion Metric Results

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

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, maccri avg and weighted avg were used to evaluate this model. Figure 4 is the result of the report from the Naïve Bayes Classification method.

Table 7. Classification Report Results Naive Bayes Classifier 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

IV. CONCLUSION

This research can be developed further by incorporating Deep Learning to achieve more accurate results. In addition, there is still a lot of room to conduct sentiment analysis research using various pre-processing, modeling, and evaluation of data. In addition, the scope of research is also a research area that is not limited.

Based on the research that has been done on sentiment analysis about public opinion on the scarcity of cooking oil on the digital twitter platform using the Naive Bayes Classifier method, the results are in the form of crawling data from deviations that open 1000 tweet data. After doing several processes such as preprocessing and the next stage is testing the Naive Bayes Classifier method with the total data obtained is 523 tweet data. The accuracy of the Naive Bayes Classifier is 72%. With this accuracy, it shows that sentiment analysis using the Naïve Bayes Classifier method has a high accuracy value.

REFERENSI

[1] F. A. Nova and G. Erdiana, "Analisa Kadar Asam Lemak Bebas Pada Minyak Jelanta Yang Diperjualbelikan Pada Penjual Goreng Di Kecamatan Medan Helvetia Secara Alkalimetri," *J. Anal. Lab. Med.*, vol. 4, no. 1, 2019.

[2] A. Nurrahmah and S. R. F. R. Putri, "Analisis Perbandingan Penggunaan Minyak Curah Dan Minyak Kemasan Menggunakan Uji Hipotesis Dua Proporsi," *Bull. Appl. Ind. Eng. Theory*, vol. 1, no. 2, pp. 65–70, 2020.

[3] D. M. Sinaga, R. Rusdina, A. P. Windarto, and A. Wanto, "Analisis Metode ELECTRE Pada Pemilihan Produk Minyak Goreng Kemasan Terbaik Berdasarkan Konsumen," *J. Sains dan Inform. - Res. Sci. Inform.*, vol. 5, no. 2, pp. 129–135, 2019.

[4] Badryah and A. Rasmito, "Pemanfaatan Kunyit Untuk Meningkatkan Kualitas Minyak Goreng Curah," *J. Tek. Ind. dan Kim.*, vol. 1, no. 1, pp. 7–15, 2018.

[5] G. F. J. Tarigan and L. J. Pinem, "Analisis Kepuasan Dan Loyalitas Konsumen Minyak Goreng Kemasan Bimoli (Studi Kasus pada Ibu-Ibu Rumah Tangga Kecamatan Medan Baru)," *Agriprimatech*, vol. 4, no. 2, pp. 57–64, 2021.

[6] S. M. 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 Factor Evaluation Process)," *Semin. Nas. Sains dan Teknol. Inf.*, no. 2021, pp. 577–580, 2021.

[7] N. A. Putra 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. 1, pp. 9–14, 2021.

[8] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinominal Naïve Bayes Classifier," *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021.

[9] A. Ftiri, F. Margasaty, Kusmaria, R. Desfaryani, and V. U. Dewi, "Peramalan Harga Minyak Goreng Di Tengah Pandemi Covid-19 Kota Bandar Lampung," *J. dwijenAGRO*, vol. 10, no. 1, pp. 21–26, 2020.

[10] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier," *J. Linguist. Komputasional*, 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," *Media Inform. Budidarma*, vol. 6, no. 1, pp. 535–541, 2022.

[12] F. N. Hasan and A. Febriandirza, "Perancangan Data Warehouse Untuk Data Penelitian di Perguruan Tinggi Menggunakan Pendekatan Nine Steps Methodologhy," *Pseudocode*, vol. VIII, no. 1, pp. 49–57, 2021.

[13] A. Wibowo, F. N. Hasan, R. Nurhayati, and A. Wibowo, "Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier," *J. Asiimetrik J. Ilm. Rekayasa dan Inov.*, vol. 4, no. 2, pp. 239–248, 2022.

[14] P. Afikah, A. Avorizano, I. R. Afandi, and F. N. Hasan, "Implementasi Business Intelligence Untuk Menganalisis Data Kasus Virus Corona di Indonesia Menggunakan Platform Tableau," *Pseudocode*, vol. 9, no. 1, pp. 25–32, 2022.

[15] D. D. Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *JITET - J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022.

[16] H. Sulastri and A. I. Gufroni, "Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia," *J. Nas. Teknol. Sist. Inf.*, vol. 3, no. 2, pp. 299–305, 2017.

[17] A. Salam, J. Zeniarja, and R. S. U. Khasanah, "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspress Indonesia)," in *SINTAKS 2018 - Seminar Nasional Teknologi Informasi & Aplikasi Komputer dan Call For Papers*, 2018, pp. 480–486.

[18] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *J. RESTI - Rekayasa Sist. dan Teknol. Inf.*, vol. 1, no. 1, pp. 19–25, 2017.

[19] F. N. Hasan, "Implementasi Sistem Business Intelligence Untuk Data Penelitian di Perguruan Tinggi," in *Prosiding Seminar Nasional TEKNOKA 4*, 2019, vol. 4, no. 2502, pp. I1–I10.

[20] H. K. Wardana, I. Swanita, and B. W. Yohanes, "Sistem Pemeriksa Pola Kalimat Bahasa Indonesia berbasis Algoritme Left-Corner Parsing dengan Stemming," *Int. J. Inf. Technol. Electr. Eng. - JNTETI*, vol. 8, no. 3, pp. 211–217, 2019.

[21] 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, pp. 1–8, 2020.