











## Word Search with Trending Reviews on Twitter

Ritzkal<sup>1\*</sup>, Sutriawan<sup>2</sup>, Bayu Adhi Prakoso<sup>1</sup>, Ahmad Zainul Fanani<sup>3</sup>, Indra Riawan<sup>1</sup>, Hersanto Fajri<sup>1</sup>,  
Ruri Suko Basuki<sup>3</sup>, Farikh Alzami<sup>3</sup>

<sup>1</sup> Faculty of Engineering & Science, Universitas Ibn Khaldun, Bogor 16162, Indonesia

<sup>2</sup> Computer Science, Universitas Muhammadiyah Bima, Bima 8111, Indonesia

<sup>3</sup> Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

Corresponding Author Email: [ritzkal@ft.uika-bogor.ac.id](mailto:ritzkal@ft.uika-bogor.ac.id)

<https://doi.org/10.18280/isi.280210>

### ABSTRACT

**Received:** 19 January 2023

**Accepted:** 10 February 2023

#### **Keywords:**

*normalization, preprocessing, text mining  
Twitter*

Indexing content is the process of text mining. An index is made using the root words that can be located in the text. The section of the text that includes the root can be found using the index. The index can also be used as a database to find trends in text, such as how frequently a word appears. Text mining is essentially the act of turning text into words that are then analyzed. Data collection, preprocessing, Term Weigthing, and categorization are some of the research techniques used. The goal of this study is to identify words that frequently appear in Twitter comments and to choose the best normalization technique based on a dictionary. The dataset for the research approach came from tweets on the rise in petrol prices. According to the research's findings, there are several words used in these comments, including the words "up" and "bbm," which are both frequently used in both positive and negative contexts. Up to 50,000 words were retrieved throughout the preprocessing phase, with 62 documents having a positive class and 180 having a negative class.

## 1. INTRODUCTION

Text data is now being used as a new kind of communication technology rather than only as a source of knowledge and insight due to recent technical advancements. Many text sources today have an impact on how people function. One of the most recent technical advancements is the internet, which has allowed many people to freely express themselves who are unable to do so through traditional media. Most people believe that there are no restrictions on the internet. Massive textual information sources like social media have given people unprecedented chances to express their thoughts in public, but they have also made it extremely difficult to comprehend what they are saying [1].

According to the Director General of Kominfo Budi Setiawan, published in [kominfo.go.id](http://kominfo.go.id), the development of the world of technology is growing very rapidly in every country, including Indonesia, with 19.5 million Indonesians recorded as Twitter users [2]. Twitter is one of the social media that is a source of news for someone's complaints to become a place to do business.

The internet, including newsgroup posts, review sites, and other places, contains a variety of opinions or text messages on governments, products, politicians, and more [3-5]. Because of this, the task of sentiment analysis or opinion mining has gained more attention over the past few years [6], and there are numerous approaches to deal with it. It is necessary to be able to collect pertinent opinions from social media in order to determine information such as opinion polarity, target object (i.e., which section expresses the opinion), holder (who created the opinion), etc. before one may mine sentiment on opinion messages.

Indexing content is the process of text mining. An index is made using the root words that can be located in the text. The section of the text that includes the root can be found using the index. The index can also be used as a database to find trends in text, such as how frequently a word appears. Text mining is essentially the process of turning text into words that are then analyzed [7]. The process of discovering new information and studying vast amounts of data is another definition of text mining. Text mining examines unstructured text that has connections to one another and to other ideas and laws. The anticipated outcome is a new understanding that was before unknowable and uncertain. Information retrieval, categorization, POS tagging, clustering, and other sub-tasks of text mining can all be grouped under the umbrella of knowledge discovery in databases, which is nothing more than a way of describing patterns in a precise, distinctive, and understandable manner [8]. Social media's linguistic idioms make it challenging for computers to automatically analyse messages because of how little they comply to grammatical norms [9]. These variations include typos, impromptu acronyms, emoticons, illegible sentence patterns, and the content of letters that contain other phonetically or even aesthetically identical characters [10]. Text that cannot be interpreted will make the NLP process more difficult. In some other situations, text normalization is crucial for tasks that depend on string matching or word frequency [11].

Text normalization refers to the process of modifying the format of a text to communicate ideas for a particular purpose. [12, 13]. Text normalization is crucial to employ since a text or reading will often contain words that are challenging to grasp, ranging from vocabulary used differently in different languages or contexts to sentences that have distinct meanings

from their original context [14, 15]. Using contextual information, the first normalization strategy converts non-standard language into standard language, whereas the second substitutes lexically based terms with forms that correspond to

the standard language [11].

Several studies, including the following, discuss normalization based on dictionaries as shown in Table 1:

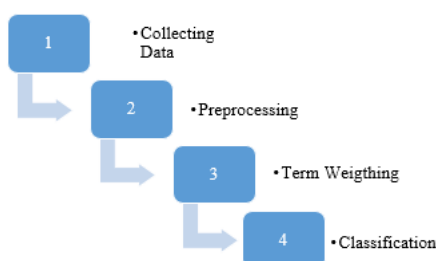
**Table 1.** Data penelitian sebelumnya

No.	Author	Title of article and journal name	Method	Tools	Dataset	Summary
1	Wijaja and Wibowo [16]	Sentiment Analysis with Slang Dictionary in Indonesian Social Media Using Machine Learning Approach. ICIC Express Letters (2022)	Collecting data, preprocessing, feature extraction, feature selection, classification and evaluation. The results compare between the results of using the Slang dictionary with not using. In accordance with the measurement of accuracy, precision, Recall and F1 Score	Python and using the tweepy, sklearn, and nltk libraries	Twitter dengan mengambil 15 kata gaul	The proposed Slang Dictionary has improved accuracy and F1 score by 8% and 9% respectively.
2	Madyatmadja and Yahya [17]	Contextual Text Analytics Framework for Citizen Report Classification: A Case Study Using the Indonesian Language. IEEE Access, 2022	Data collection, data cleaning, text preprocessing and classification	Python, SVM	Aplikasi LAKSA	Context text preprocessing improves the classification performance of the majority classifier by about 3% with the combination of n-grams.
3	Wu et al. [18]	SlangSD: building, expanding and using a sentiment dictionary of slang words for short-text sentiment classification. Lang Resources & Evaluation (2018)	Collect data, configure files and classify	Python	Twitter dan SMS	On the SMS dataset, SentiStrengthSSD outperformed SentiStrength with more than 50% recall on both tasks.
4	Gupta et al. [19]	SLANGZY: A fuzzy logic-based algorithm for English slang meaning selection. Progress in Artificial Intelligence (2019)	Data collection and preprocessing, slang factor generation	Python	Twitter, Youtube, Reddit8	The comprehensive dataset created achieved 88.44% accuracy. while selecting the best definition compared to the top definition of the same word in Urban Dictionary.

In this research, we have decided to conduct a study to get the frequently occurring words and to obtain the right method in normalizing Normalization based on Dictionary to improve the words in Twitter Comments.

## 2. METHOD

This study's dataset incorporates information from Twitter. Slang words from Twitter comments were utilized as the search terms on Twitter. In general, data collection, preparation, and categorization make up the research methodology as shown in Figure 1.



**Figure 1.** Research methods

### 2.1 Collecting data

A method or process for gathering, evaluating, and analyzing different types of data using defined methods. To obtain as much information and data as possible is the main goal of data gathering.

### 2.2 Preprocessing

Preprocessing is a process that includes the following to turn unstructured data into structured data:

#### 2.2.1 Folding a case

The transformation of string form into lowercase is known as case folding. Case folding is helpful in the subsequent stemming step since stemming distinguishes between uppercase and lowercase characters when deleting affixes.

#### 2.2.2 Tokenizing

Tokenizing is a method for breaking up phrases into several parts of speech.

#### 2.2.3 Stopword

Stopwords like from to us and others that are unnecessary

or frequently used are removed using this technique.

#### 2.2.4 Stemming

The technique of stemming allows for the mapping of many morphological word variations to their base or common words. The practice of stemming for Indonesians has changed significantly.

#### 2.2.5 Word cloud

One way to visually show text data is through word clouds, also known as text clouds or tag clouds. Because it is simple to understand, it is often used in text mining. The frequency of words can be presented in an appealing yet educational fashion by utilizing word clouds. The size of a word represented in the word cloud increases with frequency of use.

### 2.3 Term weighing

The word weighting technique lets you choose a phrase's weight based on how significant it is to the document.

### 2.4 Classification

Classification is a process of finding a model or function that describes and distinguishes classes of data or concepts with the aim of predicting classes for data whose classes are unknown.

## 3. RESULT AND ANALYSIS

It is mandatory to have conclusions in your paper. This section should include the main conclusions of the research and a comprehensible explanation of their significance and relevance. The limitations of the work and future research directions may also be mentioned. Please do not make another abstract.

### 3.1 Collecting data

In this data collection process using datasets from Twitter regarding the increase in fuel in Indonesia. This dataset was taken as many as 500 Twitter comments responding to the fuel increase. In the 500 data is processed by filtering the same words or duplicates then in the process obtained 242 data in accordance with the topic of the increase in fuel. For data as much as 258 does not match the topic, then the data is not used or the data will automatically be deleted. After filtering the data, the 242 data are labeled or sentiment classes to determine which words are positive and which words are negative. Obtained positive data as many as 62 documents and obtained negative data as many as 180 documents, the following data has been labeled or positive and negative classes described in Table 2.

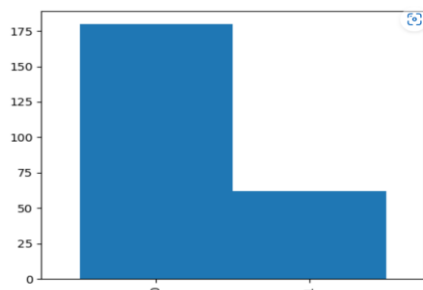


Figure 2. Sentiment analysis chart

Table 2. Positive and negative class labeling

Text	Sentiment
Penyesuaian harga BBM ternyata jadi pengaruh naiknya jumlah orang miskin di Indonesia yang per September 2022 sebanyak 2636 juta jiwa	Negatif
Kompolnas bohong	Negatif
Mobil Esemka bohong	Negatif
Tidak impor bohong	Negatif
BBM tdk naik bohong	Negatif
KOLISTRIKPAJAKPENGANGGURAN NAIK TERUS Presiden Jokowi Perintahkan Seluruh Kepala Daerah Rajin Blusukan ke Pasar	Positif
Emang lu paham gak tentang perdagangan dunia.. ekonomi dunia kayak mana Kalau mau komen harus berilmu bos gak gila kek lu... Lawak lah badut ni... BBM naik salah presiden apa2 naik salahin presiden... Aneh deh ... Pintar sekolah tapi bodoh dalam berpikir	Negatif
BBM naik sama dengan rakyatnya jadi pada miskin kita disuruh ngirit2 DPR minta naik gaji	Negatif
Harga bbm naik gaiz	Positif
Dan kini lpg pkai ktp saat beli dan kadang di aplikasi diterapkan	Positif
Wis rasah aneh <sup>2</sup> masake ngo kayu bakar ae cmiwiy	Positif
Jadi penyesuaian harga BBM berdampak pada kenaikan hargaharga yang harus dibayar kelompok masyarakat miskin dan mempengaruhi daya beli mereka kata Margo	Positif
waktu bbm naik..tol naikin harga juga dgn alasan penyesuaian.dan besok jalan berbayar..kira2 bbm ongkos ojol ongkos taksi bakal naik juga..kita liat ongkos lrt krl dan transjakarta	Positif
Kenaikan harga BBM diiringi dengan meningkatnya kemiskinan terlihat dari kemiskinan September 2022 naik tipis dibandingkan Maret 2022 namun turun dari September 2021 kata Kepala BPS Margo Yuwono dalam konferensi pers daring Senin 16/1	Positif
Selain itu kenaikan GK pada September 2022 juga disebabkan oleh kenaikan harga komoditas nonpangan pasca harga BBM naik	Positif
Misalnya Pertalite naik 3072 persen solar 3204 persen Pertamina non subsidi 16 persen	Positif
Margo mencatat harga beras pada September naik 146 persen dibandingkan Maret 2022. Lalu harga tepung terigu naik 1397 persen telur ayam ras 1901 persen	Positif
Kenaikan harga BBM menyebabkan harga komoditas yg dikonsumsi masyarakat miskin melonjak	Positif
Harga komoditas yang dikonsumsi oleh penduduk miskin ada peningkatan dibandingkan Maret 2022	Positif
Aman itu buat mereka yg berkuasa Rakyat terus di peras pajak naik.BBM naik harga barang semakin mahal.hukum tajam kebawah.lapangan kerja sulit.PHK masal akan terjadi cukup sudah penderitaan Rakyat yg di buat oleh Rezim jkw	Positif
Skrng masyarakat di Eropa susah loh. Inflasinya tinggi banget BBM mereka pake harga ekonomibelum lg hrga gas naik krn perang. Efek ke harga2 pangan yang naik tinggi. Pajaknya apakah turun Ga. Pajak mereka ttp tinggi. Gaji mereka segitu2 aja	Positif
Gk bnr nih om Herman... Catetttt yee om kl utang turun itu bkn krn jkw tp kl BBM naik pengangguran naik hrg <sup>2</sup> pd naik kesenjangan naik itu baru ulah jkw ... Bnr kg kaum onta <sup>2</sup> sekalian	Negatif

The following is a description of the data in diagram form.

The ratio of negative to positive terms in the context of Indonesia's fuel price rises is shown in Figure 2. where the positive words arrive after more than 50 and the negative words appear around 175 words.

### 3.2 Preprocessing

#### 3.2.1 Case folding

The initial stage in preprocessing is case folding which functions as changing from uppercase letters to lowercase letters. Case folding technique using python language.

```
df_data['Text'] = df_data['Text'].str.lower()
print('Hasil Case Folding: \n')
print(df_data['Text'].tail())
```

Figure 3. Case folding

Figure 3 shows the coding to use case folding with python programming language, where the data taken is text containing comments on Twitter. In the case folding process can change 242 data with lowercase letters.

```
237 coba nangis lagi kala BBM naik kali aja banyak...
238 hasil simulasi smeru terhadap tingkat kemiskinan...
239 ditunggu tangisanmu saat BBM naik
240 BBM naik berkalkali setelah sbj lengser.
241 airmata mu airmata buaya
Name: Text, dtype: object
```

Figure 4. Case folding result

Five data points that were folded into cases are shown in Figure 4 for your reference.

#### 3.2.2 Tokenizing

The tokenizing process is a process to cut words and remove numbers in a sentence taken from comments on Twitter.

	0	1	2	3	4	5	6	7			
0	penyesuaian	harga	BBM	ternyata	jadi	pengaruh	naiknya	jumlah			
1	kompilnas	bohong	None	None	None	None	None	None			
2	mobil	esemka	bohong	None	None	None	None	None			
3	tidak	impur	bohong	None	None	None	None	None			
4	BBM	tdk	naik	bohong	None	None	None	None			
...	...	...	...	...	...	...	...	...			
237	coba	nangis	lagi	kala	BBM	naik	kali	aja			
238	hasil	simulasi	smeru	terhadap	tingkat	kemiskinan		i			
239	ditunggu	tangisanmu	saat	BBM	naik	None	None	None			
240	BBM	naik	berkalkali	setelah	sbj	lengser.	None	None			
241	airmata	mu	airmata	buaya	None	None	None	None			
7	8	9	...	43	44	45	46	47	48	49	50
jumlah	orang	miskin	...	None	None	None	None	None	None	None	None
None	None	None	...	None	None	None	None	None	None	None	None
None	None	None	...	None	None	None	None	None	None	None	None
None	None	None	...	None	None	None	None	None	None	None	None
None	None	None	...	None	None	None	None	None	None	None	None
...	...	...	...	...	...	...	...	...	...	...	...
aja	banyak	orang	...	None	None	None	None	None	None	None	None
kenaikan	harga	BBM	...	None	None	None	None	None	None	None	None
None	None	None	...	None	None	None	None	None	None	None	None
None	None	None	...	None	None	None	None	None	None	None	None
None	None	None	...	None	None	None	None	None	None	None	None

Figure 5. Result tokenizing

Figure 5 demonstrates how to use the Twitter dataset to remove words and count integers from a text.

#### 3.2.3 Stopwords

Stopword removal is used to clean up irrelevant words or words that are often used such as from to us and others. The stopwords removal process is taken from comments regarding the increase in BBM on Twitter as shown in Figure 6.

	Text	Sentiment	class
0	penyesuaian harga BBM ternyata jadi pengaruh n...	Negatif	0
1	kompilnas bohong	Negatif	0
2	mobil esemka bohong	Negatif	0
3	impur bohong	Negatif	0
4	BBM tdk naik bohong	Negatif	0
5	kolistripajakpengangguran naik terus presiden...	Positif	1
6	emang lu paham gak tentang perdagangan dunia.. ...	Negatif	0
7	BBM naik sama rakyatnya jadi miskin disuruh ng...	Negatif	0
8	dengar pak ganjil genap erp ciptaker pesangonu...	Positif	1
9	harga BBM naik gaiz	Positif	1
10	omnibus law yg rugikan buruh&pekerja disahkan ...	Positif	1

Figure 6. Stopwords removal result

#### 3.2.4 Stemming

The stemming stage is a process of reducing inflected words to their base word, the basic form or root generally in written form. Stem does not need to be identical to the morphology of the root of the word, it is usually sufficient that as shown in Figure 7.

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

data_stemming = []
for index, row in df_data.iterrows():
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    data_stemming.append(stemmer.stem(row["Text"]))
pd.set_option('display.max_rows', 500)
pd.set_option('display.max_columns', 500)
pd.set_option('display.width', 1000)
df_data["Text"] = data_stemming
df_data
```

Figure 7. Stemming coding

In Figure 8 shows words that have not been stemmed where in the picture there are still words that are still irrelevant.

Figure 9 shows that the words have been stemmed.

	Text	Sentiment	class
0	penyesuaian harga BBM ternyata jadi pengaruh n...	Negatif	0
1	kompilnas bohong	Negatif	0
2	mobil esemka bohong	Negatif	0
3	impur bohong	Negatif	0
4	BBM tdk naik bohong	Negatif	0
5	kolistripajakpengangguran naik terus presiden...	Positif	1
6	emang lu paham gak tentang perdagangan dunia.. ...	Negatif	0
7	BBM naik sama rakyatnya jadi miskin disuruh ng...	Negatif	0
8	dengar pak ganjil genap erp ciptaker pesangonu...	Positif	1
9	harga BBM naik gaiz	Positif	1
10	omnibus law yg rugikan buruh&pekerja disahkan ...	Positif	1

Figure 8. Words that have not been stemmed



	Text	Sentiment	class
0	sesuai harga bbm nyata jadi pengaruh naik jum...	Negatif	0
1	kompolnas bohong	Negatif	0
2	mobil esemka bohong	Negatif	0
3	impor bohong	Negatif	0
4	bbm tdk naik bohong	Negatif	0
5	kolistikpajakpengangguran naik terus presiden...	Positif	1
6	emang lu paham gak tntang dagang dunia ekonomi...	Negatif	0
7	bbm naik sama rakyat jadi miskin suruh ngririt2...	Negatif	0
8	dengar pak ganjil genap erp ciptaker pesangonu...	Positif	1
9	harga bbm naik gaiz	Positif	1
10	omnibus law yg rugi buruh kerja sah ribu anak ...	Positif	1

Figure 9. Words that have been stemmed

### 3.2.5 Word cloud

Word cloud is a stage to visualize a collection of words in the term document matrix into an attractive display. In the word cloud process, words that often appear in both positive and negative words will be displayed. The results of the Twitter dataset where positive words often appear can be seen in Figure 10.

Figure 10 illustrates each term that actively influenced the BBM trending topic search on Twitter. These words are closely tied to each review that demonstrates the favorable attitude of people's reviews on the gasoline price hike. Figure 12 shows a word cloud of phrases that recur often in favorable assessments.

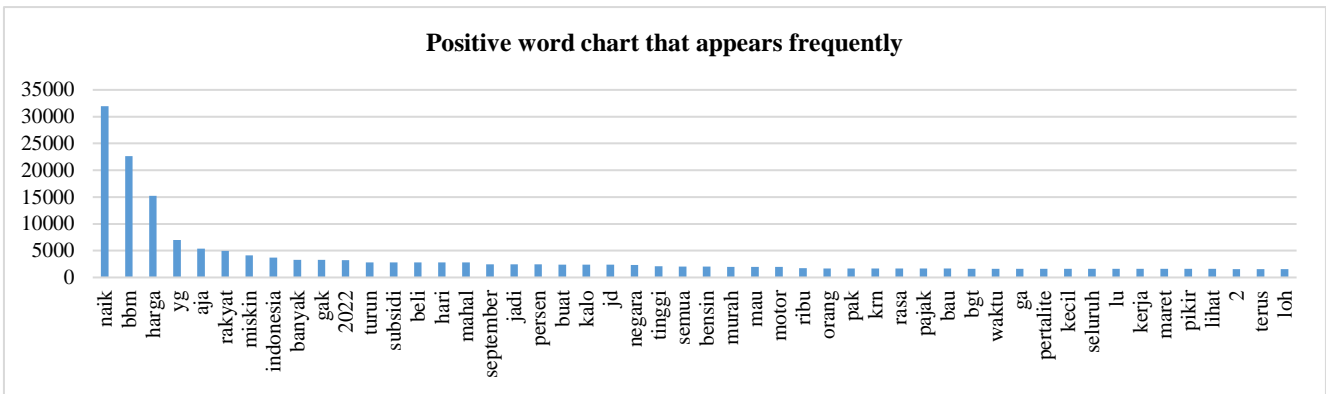


Figure 10. A positive word chart that appears frequently

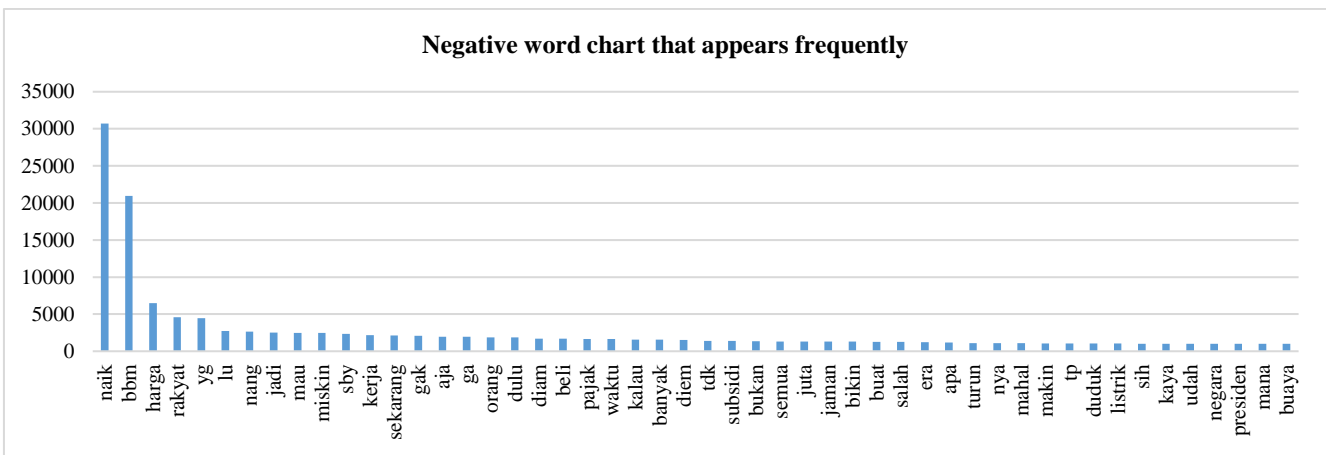


Figure 11. Negative word chart that appears frequently

These terms are directly associated to every review that exhibits a bad sentiment from the public reviews of fuel price rises, according to Figure 11, which shows every word that actively influences the search for trending BBM topics on Twitter. Figure 12 displays a wordcloud of the words that appeared often in bad ratings.

The words "naik" and "harga" have a significant impact on a variety of business aspects of people's life, as seen by Figure 12. The rise in gasoline costs had a significant impact on people's lives; for instance, it increased the cost of basic necessities, reduced people's purchasing power, and reduced traders' profits. As a result of the government's incorrect policy in making decisions about raising fuel prices because it does not take into account the risks that will arise in various business sectors in people's lives, the effect of the increase in

fuel prices has given rise to many unfavorable opinions in society.



Figure 12. WordCloud positive and negative trending words

#### 4. CONCLUSION

The conclusion of this research is in the preprocessing process obtained as many as 50,000 words where documents that have a positive class are 62 and documents that have a negative class are 180 and there are 2 positive and negative words that are often used including the word up as much as 31946 and the word bbm as much as 22655. The words "naik" and "harga" have a significant impact on a variety of business aspects of people's life, as seen by Figure 12. The rise in gasoline costs had a significant impact on people's lives; for instance, it increased the cost of basic necessities, reduced people's purchasing power, and reduced traders' profits.

#### REFERENCES

- [1] Yerasani, S., Tripathi, S., Sarma, M., Tiwari, M.K. (2020). Exploring the effect of dynamic seed activation in social networks. *International Journal of Information Management*, 51: 102039. <https://doi.org/10.1016/j.ijinfomgt.2019.11.007>
- [2] Indonesia Peringkat Lima Pengguna Twitter. (2012). [https://www.kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan\\_media](https://www.kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan_media).
- [3] Henkel, M., Perjons, E., Sneiders, E. (2017). Examining the potential of language technologies in public organizations by means of a business and IT architecture model. *International Journal of Information Management*, 37(1): 1507-1516. <https://doi.org/10.1016/j.ijinfomgt.2016.05.008>
- [4] Martin, N., Rice, J., Arthur, D. (2020). Advancing social media derived information messaging and management: A multi-mode development perspective. *International Journal of Information Management*, 51: 102021. <https://doi.org/10.1016/j.ijinfomgt.2019.10.006>
- [5] Noori, B. (2021). Classification of customer reviews using machine learning algorithms. *Applied Artificial Intelligence*, 35(8): 567-588. <http://dx.doi.org/10.1080/08839514.2021.1922843>
- [6] Georgiadou, E., Angelopoulos, S., Drake, H. (2020). Big data analytics and international negotiations: Sentiment analysis of Brexit negotiating outcomes. *International Journal of Information Management*, 51(2): 102048. <http://dx.doi.org/10.1016/j.ijinfomgt.2019.102048>
- [7] Sonalitha, E., Asriningtias, S.R., Zubair, A. (2021). Text Mining. Graha Ilmu., Yogyakarta. ISBN: 978-623-228-820-1.
- [8] Sharda, R., Delen, D., Turban, E. (2021). Business intelligence, analytics, and data science: A managerial perspective. ISBN-13: 978-013-730-571-1.
- [9] Ritter, A., Cherry, C., Dolan, W.B. (2010). Unsupervised modelling of Twitter conversations. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. [http://ai.cs.washington.edu/www/media/papers/tmpS\\_Kd9R.pdf](http://ai.cs.washington.edu/www/media/papers/tmpS_Kd9R.pdf).
- [10] Han, B., Cook, P., Baldwin, T. (2013). Lexical normalization for social media text. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(1): 1-27. <http://dx.doi.org/10.1145/2414425.2414430>
- [11] Bertaglia, T.F.C., Nunes, M.D.G.V. (2017). Exploring word embeddings for unsupervised user-generated content normalization. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, Osaka. <https://doi.org/10.48550/arXiv.1704.02963>
- [12] Duran, M.S., Avançaço, L.V., Nunes, M.D.G.V. (2015). A normalizer for UGC in Brazilian Portuguese. *Proceedings of the ACL 2015 Workshop on Noisy User-generated Text*, pp. 38-47. <https://doi.org/10.18653/v1%2FW15-4305>
- [13] Purbo, O.W. (2019). Text mining: Analisis medsos, kekuatan brand dan intelijen di internet. Yogyakarta: Penerbit Andi, ISBN: 978-979-29-9518-3.
- [14] Levanti, D., Monastero, R.N., Zamani, M., Eichstaedt, J.C., Giorgi, S., Schwartz, H.A., Meliker, J.R. (2023). Depression and anxiety on Twitter during the COVID-19 stay-at-home period in seven major U.S. cities. *AJPM Focus*, 2(1): 100062. <https://doi.org/10.1016/j.focus.2022.100062>
- [15] Aldinata, Soesanto, A.M., Chandra, V.C., Suhartono, D. (2023). Sentiments comparison on Twitter about LGBT. *Procedia Computer Science*, 216: 765-773. <https://doi.org/10.1016/j.procs.2022.12.194>
- [16] Wijaja, J.A., Wibowo, A. (2022). Sentiment analysis with slang dictionary in Indonesian social media using machine learning approach. *ICIC Express Letters*, ISSN 1881-803X, pp. 1169-1177. <https://doi.org/10.24507/icicel.16.11.1169>
- [17] Madyatmadja, E.D., Yahya, B.N. (2022). Contextual text analytics framework for citizen report classification: A case study using the Indonesian language. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2022.3158940>
- [18] Wu, L., Morstatter, F., Liu, H. (2018). SlangSD: Building, expanding and using a sentiment dictionary of slang words for short-text sentiment classification. *Lang Resources & Evaluation*. <https://doi.org/10.1007/s10579-018-9416-0>
- [19] Gupta, A., Taneja, S.B., Malik, G., Vij, S., Tayal, D.K., Jain, A. (2019). SLANGZY: A fuzzy logic-based algorithm for English slang meaning selection. *Progress in Artificial Intelligence*. <https://doi.org/10.1007/s13748-018-0159-3>