

# Analysis of Indonesian Netizen Sentiment on Platform X Regarding the Arrival of Refugees in Indonesia Using the Multinomial Naive Bayes Method

Muhammad Joefitra Zaqy<sup>1)\*</sup>, Leni Marlina<sup>2)</sup>, Rian Farta Wijaya<sup>3)</sup>

<sup>1,2,3)</sup>Universitas Pembangunan Panca Budi

<sup>1)</sup>[mjoefitraz@gmail.com](mailto:mjoefitraz@gmail.com), <sup>2)</sup>[lheny@pancabudi.ac.id](mailto:lheny@pancabudi.ac.id), <sup>3)</sup>[rianfartawijaya@gmail.com](mailto:rianfartawijaya@gmail.com).

Submitted : July 1, 2024 | Accepted : July 28, 2024 | Published : July 31, 2024

**Abstract:** This research aims to analyze the sentiments of Indonesian netizens regarding the arrival of Rohingya refugees in Indonesia using the Multinomial Naive Bayes method. Sentiment analysis was carried out on comments obtained from platform X. The data collection technique used the crawling method to extract comments from platform X users regarding the issue of the arrival of Rohingya refugees. The tool used for crawling is Google Collab. The data analysis process includes sentiment labeling, data preprocessing (case folding, stopword removal, tokenizing, stemming), and classification using the Multinomial Naive Bayes method. The research results show that the majority of Indonesian netizens' sentiments regarding the arrival of Rohingya refugees in Indonesia are negative, with a percentage of 81%. Positive sentiment reached 8%, while neutral sentiment was 11%. The Multinomial Naive Bayes method produces an accuracy of 82.5% in classifying netizen sentiment. The tools used to process the data are the Orange Data Mining application version 3.36.2. It is hoped that this research can contribute to the development of computer science, especially in the fields of Text Mining, Natural Language Processing, Machine Learning and Artificial Intelligence (AI). It is also hoped that this research will provide benefits to parties related to handling the Rohingya refugee problem in Indonesia, such as the government, humanitarian organizations, mass media, academics, the general public, and other researchers.

**Keywords:** Sentiment Analysis, Multinomial Naive Bayes, Rohingya Refugees, Indonesian Netizens, Platform X, Google Collab, Orange Data Mining, Crawling

## INTRODUCTION

Rohingya are one of the most vulnerable communities in the world. They are vulnerable to lack of citizenship status, discrimination, unfair treatment, ostracism and even persecution (Wulandari, 2022). Approximately 600 people arrived on the east coast of Aceh in four boats. At the same time, 1,018 people landed on three boats north of Langkawi Island, Malaysia. The refugee group found in Malaysia consisted of 865 men and 101 women (Zulkarnain & Kusumawardhana, 2020). On May 15 2015 an increase of 749 Rohingya people were stranded on the coastline of Indonesia, precisely in Aceh Province (Primadi, 2019).

Some Netizens on platform X (Twitter) reject the arrival of Rohingya refugees. So the Rohingya are often the target of slander and hate speech from some netizens who feel disturbed by their arrival. Therefore, efforts are needed to measure how many netizens reject their arrival or who provide negative sentiments. (<https://sekitarkaltim.id/posts/250105/ribuan-netizen-ramaikan-petisi-tolak-rohingya-dan-bubarkan-unhcr-indonesia>).

One effort that can be made is to analyze the sentiments of Indonesian netizens regarding the arrival of Rohingya refugees in Indonesia using the Multinomial Naive Bayes method to group negative, neutral and positive sentiments, so that the final result is a tendency between negative, neutral and positive sentiments by platform netizens. X (Twitter) Indonesia. Sentiment analysis is used to monitor and detect hate speech against Rohingya refugees, enabling quick action against negative content. Sentiment analysis helps in designing effective educational campaigns by highlighting support and positive views towards Rohingya refugees. Sentiment Analysis is a technique for extracting text data to obtain information about positive, neutral or negative sentiment (Sari & Wibowo, 2019). By conducting sentiment analysis, we can find out the attitudes and emotions of netizens towards

\*name of corresponding author



the arrival of Rohingya refugees and can also be used for recommendations for further research development. This sentiment analysis has been carried out several times, one of which is research on "Analisis Sentimen Masyarakat Terhadap Kondisi Perekonomian di Indonesia Pada Masa Pandemi 2020" (Padhana & Sadikin, 2021) and "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi" (Giovani et al., 2020).

## LITERATURE REVIEW

### Analisis Sentimen

Sentiment analysis is a technique in the fields of natural language processing, analytical text, and computational linguistics that identifies and extracts subjective information in text sources. The main goal is to determine the author's attitude towards some topic or the context of the text as a whole. Sentiment analysis, also known as opinion mining, is a Natural Language Processing (NLP) technique used to analyze the emotions of a text. Sentiment analysis is a process that uses Natural Language Processing (NLP) to automatically mine attitudes, opinions, views, and emotions from text, audio, tweets, and other data sources. (Utami & Erfina, 2021). This process involves extracting, understanding and processing unstructured text data to obtain sentiment information contained in an opinion or opinion sentence (Arsi & Waluyo, 2021).

### Rohingya

Rohingya is an ethnic group whose majority are Muslim. They have lived for centuries in Myanmar, a country where the majority of the population is Buddhist. The origin of the term "Rohingya" comes from the words "Rohai" or "Roshangee," which refer to the Muslim population in Rohang or Roshang, as the area was known before it was named Arakan. However, their status as part of Myanmar remains controversial and the subject of historical debate. The Rohingya ethnic conflict has long been simmering in Myanmar (<https://www.idntimes.com/news/world/andir-5-fakta-etnis-rohingya-asal-dan-kenapa-mereka-dibenci-di-myanmar-c1c2>).

As a consequence of this phenomenon, thousands of Rohingya fled their homes. Many of the houses and villages once dominated by this ethnic group are now destroyed (Matthew, 2020)

### X (Twitter)

Twitter is a social media founded by Jack Dorsey in 2006. This social media allows users to send messages or tweets in the form of text, photos, videos or links with a limit of 280 characters. Other features that Twitter also has are tweets, trending topics, mentions, direct messages, retweets and likes which are very useful for communicating and searching for information. (Liani & Rina, 2020).

### Natural Language Processing (NLP)

NLP is a technology used by machines to understand, analyze, manipulate and interpret human language. This is a part of Artificial Intelligence (AI) that is concerned with giving computers the ability to understand natural human language (<https://www.dewaweb.com/blog/nlp-adalah/>). Technological developments in recent decades have enabled significant advances in NLP research and development. Indonesian, as the official language in the country with the fourth largest population in the world, is an important subject in NLP research (Amien, 2023).

### Multinomial Naïve Bayes

The Multinomial Naïve Bayes Method is a probabilistic classification algorithm rooted in Bayes' Theorem. This algorithm assumes that all features in the data are dependent on each other. In the context of sentiment analysis, this method is used to classify text (in this case, netizen comments) into certain sentiment categories (Sabrani et al., 2020).

$$p(c) = \frac{Nc}{Ndoc} \quad (1)$$

Information:

c : Category or class

doc : Document

Nc: The number of categories c in the training document

Ndoc: The total number of training documents used

Further calculation of the probability that word i belongs to a certain category or class can be done using Equation 2 (Lesmana et al., 2020).

$$P(W_i, c) = \frac{\text{count}(w_i, c) + 1}{\sum_w \text{Count}(w, c) + |V|} \quad (2)$$

Information:

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

$w_i$ : The  $i$ th word in all documents in category  $c$   
 $count(w, c)$ : The number of certain words that appear in a category or class  $c$   
 $\sum_w count(w, c)$ : The total number of words in the class  
 $|V|$ : is the number of all unique words in the class.

Adding a value of one serves to prevent the probability result being 0, this manipulation is called Laplace smoothing. After the learning stage is complete, the next step is to classify new data based on the learning results. For new data classification, calculations are carried out using Equation

$$3.P(c, d) = P(c) \prod_{v \in d} \frac{count(v, c) + 1}{\sum_w count(w, c) + |V|} \quad (3)$$

Information:

$(v, d)$ : Number of unique words in the document

**Confusion Matrix**

Confusion Matrix is a method used to calculate accuracy in data mining concepts (Pratiwi et al., 2021). To produce accuracy, a Confusion Matrix is needed with data generated from the classification model in the form of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) data by comparing the actual data and the classification results data. For accuracy results, calculations are carried out using Equation 4.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

**METHOD**

In Figure 1, this research involves several stages: importing data from Twitter, labeling, preprocessing (including Stopwords), tokenizing, stemming, and splitting the data into 80% training and 20% test. The Multinomial Naive Bayes method is used for sentiment classification. The results are visualized and the model is evaluated.

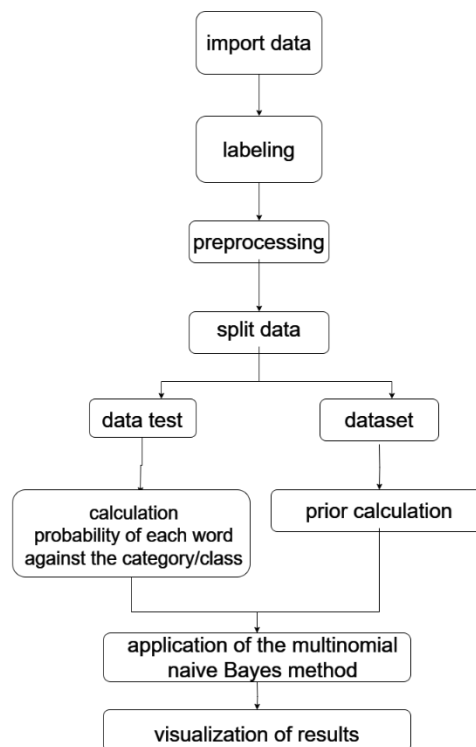


Figure.1 Flow of implementing the method

**Import Data**

At this stage the raw data that has been obtained will be entered for the process of labeling each data.

**Labeling**

At this stage, the raw data will be given labels such as negative, neutral and positive for training data needs as a reference for calculating accuracy for classification.

**Preprocessing**

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

At this preprocessing stage, several processes are carried out to obtain a clean data set, thereby making the process of testing sentiment classification results more accurate. The preprocessing process used in this research includes:

- a. Case folding  
Case Folding is a technique for changing uppercase or capital letters to lowercase.
- b. Stopwords  
Stopword removal is a technique for removing conjunctions or affixes such as 'and', 'yang', 'to', or abbreviations of words to only place meaningful words that have important meaning.
- c. Tokenizing  
Tokenization is a technique for splitting the words that form the input string, which can be interpreted as a technique for breaking sentences into words by splitting the words to parse and determine a set of words, the syntax structure of each word.
- d. Stemming  
The next process in text preprocessing is stemming. At this stage, the tokens obtained from the tokenization process are converted into their basic form. The stemming process is usually carried out on nouns, verbs and adjectives.

## RESULT

### Classification uses the multinomial Naïve Bayes method

At this stage calculations are carried out using sample data which has been divided into training data and test data which can be seen in Table 1 and Table 2 below.

Table 1. Preprocessing Results (dataset)

full_text	Sentiment
orang india islam lolos sasar kebencihan rohingya tega bener	Negative
rohingya representasi yajuj maju	Negative
belain rohingya tampung rumah loe	neutral
harus tolong muslim rohingya	positive
diterima pengungsi rohingya sisihkan anggaran memenuhi kebutuhannya	positive

Table 1. Data test

full_text	Sentiment
Usir orang rohingya jangan sampai kembali	negative

After the data is divided into training data and test data, class probability is then calculated using Eq 1.

$$P(\text{positive}) = \frac{2}{5}$$

$$P(\text{negative}) = \frac{2}{5}$$

$$P(\text{neutral}) = \frac{1}{5}$$

Then proceed to calculate the probability of words sourced from test data using Equation 2.  
Sentiment : Usir orang rohingya jangan sampai kembali

Class Calculation Positive :

$$P(\text{usir, positive}) = \frac{0+1}{12+33} = \frac{1}{45}$$

$$P(\text{orang, positive}) = \frac{0+1}{12+33} = \frac{1}{45}$$

$$P(\text{rohingya, positive}) = \frac{2+1}{12+33} = \frac{3}{45}$$

$$P(\text{jangan, positive}) = \frac{0+1}{12+33} = \frac{1}{45}$$

$$P(\text{sampai, positive}) = \frac{0+1}{12+33} = \frac{1}{45}$$

$$P(\text{kembali, positive}) = \frac{0+1}{12+33} = \frac{1}{45}$$

Class Calculation Neutral :

$$P(\text{usir, neutral}) = \frac{0+1}{8+33} = \frac{1}{41}$$

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

$$P(\text{orang, neutral}) = \frac{0+1}{12+33} = \frac{1}{45}$$

$$P(\text{rohingya, neutral}) = \frac{1+1}{8+33} = \frac{2}{41}$$

$$P(\text{jangan, neutral}) = \frac{0+1}{8+33} = \frac{1}{41}$$

$$P(\text{sampai, neutral}) = \frac{0+1}{8+33} = \frac{1}{41}$$

$$P(\text{kembali, neutral}) = \frac{0+1}{8+33} = \frac{1}{41}$$

Class Calculation Negative :

$$P(\text{usir, negative}) = \frac{0+1}{13+33} = \frac{1}{46}$$

$$P(\text{orang, negative}) = \frac{1+1}{12+33} = \frac{2}{45}$$

$$P(\text{rohingya, negative}) = \frac{2+1}{13+33} = \frac{3}{46}$$

$$P(\text{jangan, negative}) = \frac{0+1}{13+33} = \frac{1}{46}$$

$$P(\text{sampai, negative}) = \frac{0+1}{13+33} = \frac{1}{46}$$

$$P(\text{kembali, negative}) = \frac{0+1}{13+33} = \frac{1}{46}$$

After calculating the probability of the test data words, then the probability of the test data document is calculated for the sentiment class using Equation 3.

$$P(\text{positive,dl}) = \frac{2}{5} x \left(\frac{1}{45}\right)^1 x \left(\frac{1}{45}\right)^1 x \left(\frac{3}{45}\right)^1 x \left(\frac{1}{45}\right)^1 x \left(\frac{1}{45}\right)^1 x \left(\frac{1}{45}\right)^1$$

$$= 0.4 x 0.022 x 0.022 x 0.066 x 0.022 x 0.022 x 0.022$$

$$= 0.0001360558848$$

$$P(\text{neutral,dl}) = \frac{1}{5} x \left(\frac{1}{41}\right)^1 x \left(\frac{1}{41}\right)^1 x \left(\frac{2}{41}\right)^1 x \left(\frac{1}{41}\right)^1 x \left(\frac{1}{41}\right)^1 x \left(\frac{1}{41}\right)^1$$

$$= 0.2 x 0.024 x 0.024 x 0.048 x 0.024 x 0.024 x 0.024$$

$$= 0.0000764411904$$

$$P(\text{negative,dl}) = \frac{2}{5} x \left(\frac{1}{46}\right)^1 x \left(\frac{2}{46}\right)^1 x \left(\frac{3}{46}\right)^1 x \left(\frac{1}{46}\right)^1 x \left(\frac{1}{46}\right)^1 x \left(\frac{1}{46}\right)^1$$

$$= 0.4 x 0.021 x 0.043 x 0.065 x 0.021 x 0.021 x 0.021$$

$$= 0.000217429758$$

Based on calculations, the test data sample has the greatest chance of being classified as negative sentiment. This is because the probability for the negative class (0.000217429758) is greater than the neutral (0.0000764411904) and positive (0.0001360558848) classes.

## DISCUSSIONS

### Testing Comment Data

Testing was carried out with 1327 datasets obtained from Crawling x (twitter) results regarding the arrival of Rohingya refugees in Indonesia which can be seen in table 3 below.

Table 2. Total Data

No	Full_text	sentimen
1	Padahal orang India yg islam pun juga ga lolos jadi sasaran kebencihan. Apalagi rohingya... Tega Bener kalian	negative
2	@Arnasamudra Ya kan jauh juga Mas... gak kuat pengungsi Rohingya nyebrang ke sana	neutral
3	@SukeImum4 @GreatAdjie Ulama bela rohingya	positive
----		
1327	Yg menolak pengungsi rohingya dgn berbagai dalil & topeng kompromi mcm2 = fasis Sm sebangun spt israel thd pengungsi eritrea	negative

From the test results using the implementation of the Multinomial Naive Bayes method using the Orange Data Mining Confusion Matrix application which can be seen in table 4 Confusion Matrix below.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Table 3. Confusion Matrix

		prediction		
		negative	neutral	positive
actual	negative	1073	1	2
	neutral	130	9	1
	positive	92	0	19

From the results of the Confusion Matrix calculation in Table 4, accuracy calculations were then carried out using Equation 4.

$$\begin{aligned}
 \text{Accuracy} &= \frac{1073+9+19}{1073+1+2+130+9+1+92+0+19} \times 100\% \\
 &= \frac{1101}{1327} \times 100\% \\
 &= 0.829 \times 100\% \\
 &= 82.9\%
 \end{aligned}$$

The following is the total and percentage chart of the number of negative, neutral and positive sentiments :

Table 4. Total Sentiment

Total Sentiment	
negative	1076
neutral	140
positive	111
total	1327

In table 5 there is a total of 1327 sentiments, including 1076 negative sentiment data, 140 neutral sentiment data and 111 positive sentiment data.

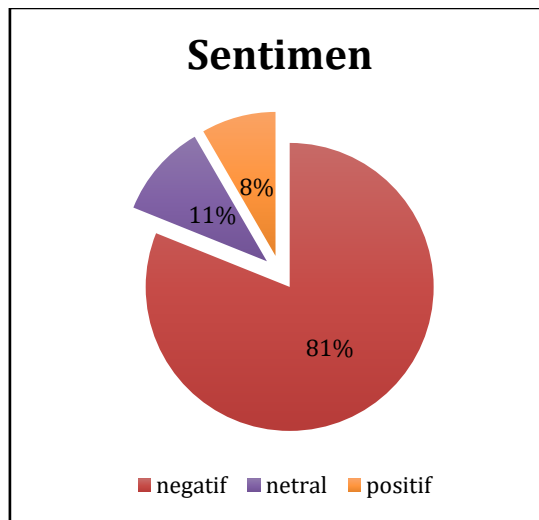


Figura 2. Visualization Sentiment

In Figure 2 there is 81% negative sentiment, 11% neutral sentiment and 8% positive sentiment out of a total sentiment of 100%.

From the test results, there are several supporting factors for obtaining high accuracy, including the following:

- a. Labelling  
The labeling stage is used as a supporting factor because at this stage good data selection is carried out in order to improve the quality of data in certain classes and labeling is done based on sentiment.
- b. Preprocessing

\*name of corresponding author



This preprocessing stage is also a supporting factor because this stage eliminates anomalies in the data so that it can strengthen data by reducing redundancy of words or even useless words. For example, two words were found that had the same meaning, namely "jls 😊" and "jelas", in the preprocessing stage the word "jls 😊" was processed by cleaning, replacing the slang word and the word changed to "jelas" so that the word "jelas" had 2 frequencies. Then, in another example, the stemming process is applied to the words "bantu" and "membantu", so the word "membantu" is changed to "bantu" so that the word "bantu" has 2 frequencies.

c. Data duplication

Don't let there be the same or duplicate data because that will affect accuracy. If there is duplicate data, it must be deleted.

### CONCLUSION

This research focuses on analyzing the sentiment of Indonesian netizens regarding the arrival of Rohingya refugees in Indonesia on platform This research uses the Orange Data Mining application to implement calculations using the Multinomial Naive Bayes method, Confusion Matrix and uses Google Colab to retrieve data on platform x using the Crawling method. From the test results, the accuracy results were 82.9%, which tended to be negative. There is a total of 1327 sentiments, including 1076 negative sentiment data, 140 neutral sentiment data and 111 positive sentiment data and There is 81% negative sentiment, 11% neutral sentiment and 8% positive sentiment out of a total sentiment of 100%.

### REFERENCES

- Amien, M. (2023). *Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia*. 2007, 1–7. <http://arxiv.org/abs/2304.02746>
- Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Idntimes (4 Dec 2023) 5 Fakta Etnis Rohingya : Asal dan Kenapa Mereka Dibenci Myanmar <https://www.idntimes.com/news/world/andi-ir/5-fakta-etnis-rohingya-asal-dan-kenapa-mereka-dibenci-di-myanmar-c1c2s>
- Lesmana, L., Mukrodin, & Nabyla, F. (2020). Analisis Sentimen Pengguna Twitter PPDB Menggunakan Algoritma Multinomial Naive Bayes. *Jurnal Sistem Informasi Dan Teknologi Peradaban (JSITP)*, 1(1). <https://journal.peradaban.ac.id/index.php/jsitp/article/view/604>
- Liani, D. N., & Rina, N. (2020). Motif Penggunaan Media Sosial Twitter (Studi Deskriptif Kuantitatif Pada Pengikut Akun Twitter @EXOind). *Cakrawala: Jurnal Humaniora Bina Sarana Informatika*, 20(1), 63–67. <http://ejournal.bsi.ac.id/ejournal/index.php/cakrawala>
- Matthew, G. (2020). National interests and Indonesian-style diplomacy in resolving the Rohingya ethnic conflict in Myanmar. *Jurnal Hubungan Internasional □ Tahun XIII*, 1, 39–52.
- Mike Napizahni (2 Juni 2022) Natural Language Processing (NLP) : Penjelasan % Contoh Penerapannya <https://www.dewaweb.com/blog/nlp-adalah/>
- Padhana, K. A., & Sadikin, M. (2021). Analisis Sentimen Masyarakat Terhadap Kondisi Perekonomian di Indonesia Pada Masa Pandemi 2020. *Jurnal Ilmu Teknik Dan Komputer*, 5(2), 268–277.
- Pratiwi, B. P., Handayani, A. S., & Sarjana, S. (2021). Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix. *Jurnal Informatika Upgris*, 6(2), 66–75. <https://doi.org/10.26877/jiu.v6i2.6552>
- Primadi, H. (2019). Progres Penanganan Pengungsi Rohingya Oleh Pemerintah Indonesia Di Provinsi Aceh Tahun 2016-2018. *EJournal Ilmu Hubungan Internasional*, 7(1), 299–308.
- Sabrani, A., Putu, I. G., Wedashwara, W., & Bimantoro, F. (2020). *METODE MULTINOMIAL NAÏVE BAYES UNTUK KLASIFIKASI ARTIKEL ONLINE TENTANG GEMPA DI INDONESIA ( Multinomial Naïve Bayes Method for Classification of Online Article About Earthquake in Indonesia )*. 2(1), 89–100.
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, 10(2), 681–686.
- Sekitar Kaltim (5 dec 2023), Ribuan Netizen Ramaikan Petisi Tolak Rohingya dan Bubarkan UNHCR Indonesia <https://sekitarkaltim.id/posts/250105/ribuan-netizen-ramaikan-petisi-tolak-rohingya-dan-bubarkan-unhcr-indonesia>
- Utami, D. S., & Erfina, A. (2021). Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Vector Machine (SVM). *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 1(1), 299–305.
- Wulandari, V. (2022). PERLAKUAN PEMERINTAH MYANMAR TERHADAP MINORITAS MUSLIM ROHINGYA PERSEPTIF SEJARAH DAN HUKUM INTERNASIONAL Veronika Wulandari Program Studi Ilmu Hukum Fakultas Ilmu dan Ilmu Sosial Abstrak Menurut beberapa sejarawan Myanmar , kata Rohingya baru-baru ini m. *Jurnal Ilmu Hukum Sui Generis*, 2(3), 51–68.
- Zulkarnain, Z., & Kusumawardhana, I. (2020). Bersama untuk Kemanusiaan: Penanganan Lintas Sektor terhadap Masalah Pengungsi Rohingya di Aceh 2015. *Jurnal HAM*, 11(1), 67. <https://doi.org/10.30641/ham.2020.11.67-83>