Indonesian Journal of Data and Science

Volume 7 Issue 1
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v7i1.313

Research Article

Implementation of Latent Dirichlet Allocation (LDA) to Identify National Pride in YouTube Comments

Rizal Kurniawan [1],*; Amin Akbar [2]; Rinaldi [3]; Aflah Zakinov Irta [4]
[1] Universitas Negeri Padang, Padang, Sumatera Barat 25132, Indonesia, rizal.kurniawan@fip.unp.ac.id
[2] Universitas Negeri Padang, Padang, Sumatera Barat 25132, Indonesia, aminakbar@fip.unp.ac.id
[3] Universitas Negeri Padang, Padang, Sumatera Barat 25132, Indonesia, naldipsi@fip.unp.ac.id
[4] Universitas Negeri Padang, Padang, Sumatera Barat 25132, Indonesia, aflah.zakinov@fpk.unp.ac.id

Correspondence should be addressed to Rizal Kurniawan; rizal.kurniawan@fip.unp.ac.id

Received 12 November 2025; Accepted 09 March 2026; Published 30 March 2026

© Authors 2026. CC BY-NC 4.0 (non-commercial use with attribution, indicate changes).
License: https://creativecommons.org/licenses/by-nc/4.0/ — Published by Indonesian Journal of Data and Science.

Abstract:
Sports can trigger a strong sense of national pride, a psychological factor that contributes positively to collective behavior and societal cohesion. One key source of this pride is the success of athletes in prestigious international competitions. With the growing use of the internet, expressions of national pride are increasingly visible on social media, particularly in the comment sections of content related to national achievements. Previous studies on sports and national pride have mainly relied on survey methods. This study takes a different approach by analyzing expressions of national pride among Indonesian netizens in YouTube comments related to the success of Indonesian athletes in international sporting events. A case study was conducted using videos covering Indonesian achievements at the Paris Olympics 2024 in weightlifting, badminton, and rock climbing. Data were collected from 12 YouTube news videos using Mozdeh, yielding 11,972 comments, of which 10,610 were analyzed after preprocessing. The study applied Latent Dirichlet Allocation (LDA) to identify hidden topics, with the optimal number of topics determined as two based on coherence values. Term frequency–inverse document frequency (TF-IDF) was used for data representation. The results revealed two main themes: gratitude, reflecting thankfulness to God for achievements, and national pride, indicating public pride in Indonesia due to athletes' success. These findings demonstrate that LDA effectively captures expressions of national pride in social media data and highlight the role of sports achievements in strengthening national pride online.

Keywords: National Pride, Athletes Achievements, Topic Modelling, Latent Dirichlet Allocation, YouTube Comments.

1. Introduction

National pride is an important thing that every citizen must possess. National pride is a source of well-being and happiness for people in a country [1], [2]. National pride is also a source of behavior that benefits the country. In the era of the COVID-19 pandemic, national pride plays an important role in increasing a person's willingness to get vaccinated to suppress COVID-19 cases, even though at that time, the vaccine was still in doubt [3]. National pride also plays a role in increasing a person's awareness of paying taxes and can also play a role in increasing a person's desire to maintain the security of their country [4], [5], [6], [7]. Low national pride also has the potential to spread xenophobia [8].

Sports achievement has been an important source for the formation of national pride in a country's society [9]. This emotion is formed from the superiority of a country over other countries, which increases their self-esteem [10], [11], [12]. A country's victory in an international match can increase the national pride of the winning country's citizens and decrease the national pride of the losing country [13].

One way to increase national pride is when a country hosts an international sporting event, such as the Olympics [9]. The Olympics are one of the most significant international sporting events to be held regularly. During the 2021 COVID-19 pandemic, the Olympics were held in Tokyo. Despite receiving many cons and emotions of fear and

anxiety, the event still increased the national pride of its citizens [14]. In addition to the Olympics, the World Cup and continental events can increase the national pride of a host country's citizens [13], [15]. In addition to the host country, participating countries that participate in the tournament can increase the national pride of their citizens [16]. In addition, national symbols displayed during sporting events can also increase a person's pride because they can increase positive feelings toward the country, especially when displayed at prestigious events [17].

In addition to sporting events, athletes' sporting achievements in international competitions can increase national pride. A study conducted by [2] stated that from 2011 to 2016, Germany had a higher level of national pride than other European countries because of its good sporting performance and achievements. In addition, a person's participation in a sport will increase national pride if the sport achieves achievements in sporting events at the international level [18].

Nowadays, the development of social media is fantastic. According to data from We Are Social, in 2025, internet connections in indonesia will exceed the population of Indonesia itself, which is 125%. The number of Internet users is also relatively high, with data showing that 74.6% of Indonesia's population is connected to the Internet, and 50.2% of them actively use social media [19]. YouTube is one of the most widely accessed social media in indonesia because this social media provides videos for free. In addition, someone can also comment on every video that is watched freely. Comments given by viewers can describe their opinions and emotions about the videos they watch [20]. Several content creators on YouTube provide videos related to sports. This content can meet the need for information about the latest sports developments.

Olympic-level sports are the highest level of competition in sporting events. A person's victory in this event will be seen by worldwide. This situation will bring a positive image of a country and, at the same time, affect the emotions of its citizens. Positive emotions the community feels can be a source of national pride during and after the Olympics [21], [22]. There are also many videos on YouTube showing the victories of Indonesian athletes. The audience who watches the video should have certain emotions when watching the video, such as national pride. Research on national pride with athlete achievements has been widely conducted, but they still use a survey approach [2], [21], [22], [23], [24]. The purpose of this study is to examine the expression of national pride among Indonesian netizens in YouTube comments about Indonesian athletes who won medals at the 2024 Paris Olympics. We use a case study of netizens' comments in indonesia related to a YouTube video about Indonesian athletes' medal achievements at the 2024 Paris Olympics.

This study attempts to answer two questions: Does national pride appear in netizen comments on YouTube videos of Indonesian athletes' victories at the Paris Olympics? Second, what topics emerge in netizen comments on YouTube videos about Indonesian athletes' victory at the Paris Olympics? We hypothesize that national pride is one of the topics of discussion among Indonesian netizens regarding sports achievements in YouTube comments. This study contributes to encouraging future research focusing on the detection of individual behavior on social media, particularly through text analysis and topic modeling approaches aimed at understanding users' social expressions, identity, and psychological dynamics in digital spaces. This research was conducted systematically and reproducibly using the steps we describe in the following sections.

2. Method

This study uses a case study approach to social media [25]. We determined whether national pride appeared in social media comments that contain content about athletes' victories at international sporting events. We use text mining, a study conducted to find meaning from a collection of texts that are usually large in number [26]. Text mining research is also related to content analysis because researchers must determine the meaning of a text. In general, the study process can be described as in **Figure 1**.

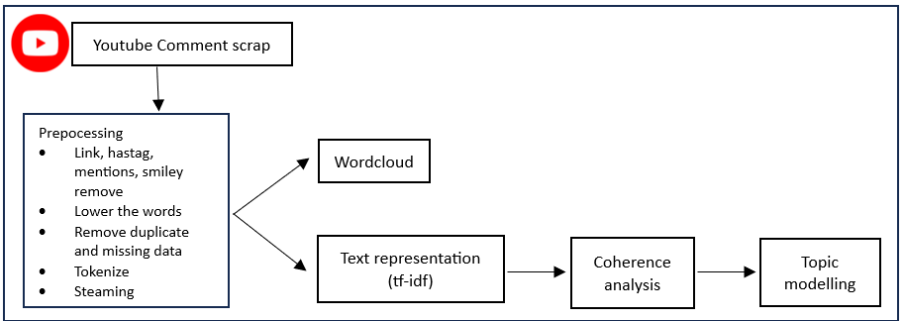

Figure 1. Research Process

Datasets

The researcher used YouTube comments in this study. The researcher did not provide data on the respondents' identities. The researcher only used comments given by netizens as research data for academic purposes. According to [27], when a post given by someone on social media is not protected or stored privately, then the data may be used for scientific purposes. This study also does not interact with humans, so it does not require informed consent.

This study analyzed YouTube comments given by netizens in Indonesia in response to the victory of Indonesian athletes at the 2024 Paris Olympics. We only take videos from official Indonesian-language news sources that show Indonesian athletes' victories at the 2024 Paris Olympics. The researcher used Mozdeh to collect data from YouTube video comments. Mozdeh is a tool developed for text scraping from a YouTube page. There were three athletes consisting of 2 people who won gold medals in rock climbing and weightlifting and one who won a bronze medal from the women's single badminton branch. This study took comments from 12 news stories and clips of each athlete who won a medal that aired on YouTube. We obtained a total of 11,972 comments which we retrieved on December 11, 2024. **Table 1** describes the data sources, access times and the number of comments we obtained.

Table 1. Data Sources

	Queries	Access time	Number of comments
1	v=SMepXzeG8HE	11/12/2024 21:55:15	239
2	v=Y7p1RoU-frw	11/12/2024 22:04:06	62
3	v=zXI_YEjw-aw	11/12/2024 22:17:54	504
4	v=7c1lMj34hzo	11/12/2024 20:20:48	354
5	v=oQAkcLwkpR0	11/12/2024 20:32:16	894
6	v=rCgLE6ufsOc	11/12/2024 20:43:18	2632
7	v=f1l7599RT6o	11/12/2024 20:52:25	3454
8	v=LLQL8yZEk1E	11/12/2024 22:10:43	2922
9	v=S62MKRyWHbc	11/12/2024 22:08:41	460
10	v=zzlmBAS6VuI	11/12/2024 21:52:27	196
11	v=A20xzXqYcQw	11/12/2024 21:57:07	241
12	v=aqFkp6abfQ4	11/12/2024 22:14:30	14

Pre-processing

In the first stage of data analysis, researchers conducted preprocessing, which was preparing data to be analyzed in the following process. Furthermore, researchers changed the text to lowercase and continued deleting duplicate data, emojis, links and empty data. In this process, researchers utilized Python 3.10 with the tweet-preprocess module version 0.6.0. We also identify and remove duplicate comments using the pandas module. Furthermore, researchers perform tokenizing, which is separating sentences into words and removing punctuation and meaningless words (stopwords). Our stopwords are a combination of punctuation, Indonesian stopwords from the NLTK (Natural Language Toolkit) Python module version 3.8.1, and several additions from researchers not covered by the two

modules. In the last stage in this section, researchers performed data stemming, which is changing text into basic words; for example, the word "runner" changes to "run". We used the Sastrawi module version 1.0.1 in Python to assist us with the stemming process. Data cleaning revealed several empty comments, such as those containing emoticons and stopwords, that were removed. We also removed these empty comments from the dataset. Researchers managed to get 10610 clean comments ready for further analysis. Researchers also created a word cloud from the data to see the frequency of the most frequently appearing words. We employed orange software version 3.39 to create a word cloud.

Text Representation

Before carrying out the topic modelling process, the text must be transformed into a matrix containing numbers representing the text. The text can be subjected to various analyses, including topic modelling, with this matrix. We use term frequency-inverse document frequency (tf-idf) as a text representation [28]. Tf-idf is a measure to determine how important a word is in a corpus within the entire document. Tf-idf is the result of multiplying text frequency by inverse document frequency. Tf is a measure of how important a word is in a document [29]. The tf value is calculated using the following equation.

$$Tf = \frac{frequency\ term\ occur\ in\ document}{total\ words\ in\ document} \quad (1)$$

Meanwhile, the idf is a measure of a word's rarity within the entire document. Idf can provide information about a word's rarity within the entire document. A rare word can indicate that the word is important in distinguishing a document from the rest of the document (Albrecht, 2020). To calculate the idf value, you can use the following equation.

$$Idf = ln\frac{total\ document}{document\ contain\ term} \quad (2)$$

The tf-idf value is obtained from the vector multiplication of the tf value and the idf value. The equation for the calculation can be written as follows. The tf-idf value describes the importance of a word in a document and its rarity across the entire corpus. The higher the tf-idf value of a word, the more important and rare the word is, so that it can be considered as a topic. To calculate tf-idf values from the data in this research, the researcher used R version 4.5.0 and the tidytext package [30], [31].

$$Tf - idf = tf \times idf \quad (3)$$

Topic Modelling

In addition, researchers must also determine the number of topics before conducting topic modelling analysis. We used the coherence value to determine the number of topics. The coherence value reflects the level of topic homogeneity. The higher the coherence value indicates a good number of topics [32]. We use the C_V coherence approach to determine the coherence value. This approach begins with a segmentation process that creates word pairs from a specified number of topics. Next, we calculate semantic closeness using normalized pointwise mutual information (NPMI). After that, a vector is created for each word in the topic to capture the semantic relationships among words. The final step in the process of measuring the closeness between words using cosine similarity is to calculate the average, which becomes the C_V coherence value. The C_V coherence value moves from the interval 0 to 1. The higher the C_V coherence value, the better the number of topics that represent the entire document [33]. This study uses Python with the gensim module version 4.4.0 to determine the C_V value.

Next, researchers conducted topic modelling analysis, one of the text mining analyses that aimed to identify the topics discussed from an extensive collection of texts [34]. We used latent directlet allocation (LDA). LDA is an analysis that finds themes latently from a collection of texts. LDA assumes that documents have specific topics and certain keywords represent each topic. LDA can bring up these keywords and will be interpreted by researchers to become specific themes [35], [36]. In this study, LDA was used to extract several keywords for each topic from the YouTube comments we get, and the researcher will interpret these words to become themes.

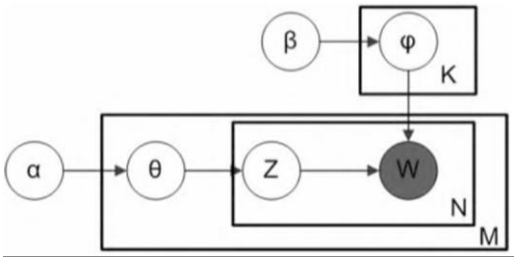

Figure 2. Modelling topic with Laten Dirichlet Allocation

Figure 2 shows the LDA process. The entire document (M) is separated into vocabulary (N), and the number of topics (k) is determined. Then the model estimates the alpha (α) and beta (β) parameters. The α parameter represents the degree of overlap between documents and topics. A low α value indicates a narrow density, making it difficult to develop new topics. The β parameter represents the probability of a word appearing in a topic. A low β value indicates a word that rarely appears in that topic. Next, the model will develop a topic distribution (θ) based on the α value and create a word distribution (φ) based on the β value. From this distribution, the model can determine topics (Z) and form words that appear (W) in each topic [32], [36]. We use the R program version 4.5.0 with the topicmodels package [37], [38] to perform topic modeling analysis. To ensure consistent results, we use seeds to stabilize LDA. We also checked the validity of the research results by comparing model outputs across different seeds.

3. Result and Discussion

Results

Figure 3 is the result of the word cloud that we got. Several words appeared most often from the collection of YouTube comments that we got. The word "Indonesia" is the most frequently appearing word. In addition, there are the words *emas* (gold), *bangga* (pride), *selamat* (congratulations), *medali* (medals), veddriq (athlete names), *terima kasih* (thank you), and olimpiade (olympic). From these words, we can see that netizens in indonesia identify themselves as part of Indonesia and feel proud and grateful to the athletes who have competed and won gold medals at the 2024 Paris Olympics.

Figure 3. Wordcloud Result

Table 2 is the result of calculating the coherence value to determine the number of topics from the YouTube comment data set in this study. We simulated by calculating the coherence value for a range of topic numbers from 2 to 20. We used Python with the gensim module to complete this process. We used the process steps from [34] by using corpus settings using tf-idf, chunk size of 2000, iteration of 400, and random state = 42. We found that the coherence value was high when the number of topics was two topics. Therefore, we used the number of topics as many as 2 to carry out the topic modelling process.

Table 2. Result of Number of Topics and Coherence Value

Numberr of Topics	2	3	4	5	6	7	8	9	10	11
Coherence values	0.61	0.59	0.50	0.49	0.48	0.47	0.47	0.46	0.44	0.42
Numberr of Topics	12	13	14	15	16	17	18	19	20	
Coherence values	0.40	0.45	0.41	0.42	0.41	0.42	0.44	0.43	0.42	

Table 3 shows the top words, word weights, and example comments for each topic across three seed values. The top words for each topic appear consistent. The main difference in the top words for the two topics is that topic 1 includes the word "pride," which, although it appears once in seed 2, has a low weight. Meanwhile, the unique word in topic 2 is "*alhamdulillah*" (praise of God), which does not appear in topic 1. Furthermore, the word "olympic" also appears only in topic 1 and not in topic 2. We interpret this as topic 1 discussing national pride, while topic 2 discusses gratitude.

The six top words for topic 1 are Indonesia (Indonesia), *emas* (gold), *bangga* (pride), *selamat* (congratulations), *medali* (medal), *terimakasih* (thank you), veddriq (athlete name), and rizki (athlete name). Topic 1 explores national pride, showing how Indonesian netizens feel proud of their country following the achievements of Indonesian athletes at the Olympics. They feel proud to be Indonesian because their athletes who competed at the Olympics won gold medals in sport climbing and weightlifting. They also congratulated and thanked these athletes for their contributions to Indonesia in the world's most prestigious sporting competition.

The six keywords for topic 2 were Indonesia, *alhamdulillah* (praise God), *emas* (gold), veddriq (athlete name), rizki (athlete name), leonardo (athlete name), *bangga* (pride), *medali* (medal), *terimakasih* (thank you), and *selamat* (congratulations). Topic 2 explores gratitude, expressing thanks to God for the victories of Indonesian athletes competing at the Paris Olympics. They expressed gratitude for the Indonesian athletes' gold medals. They shared their pride in the victory. They also expressed their gratitude to the athletes who competed.

Table 3. Top Word, word weight, and example comment on Topic Based on Seed Difference

Topic 1 (National pride)		
Seed	Top words	Example comments
Seed 1	Indonesia (Indonesia) (0.040), *Emas* (Gold) (0.024), *Selamat* (Congratulation) (0.023), *Bangga* (Proud) (0.018), Rizki (Athlete name) (0.015), *Terimakasih* (Thank you) (0.015).	"*Terharu sekaligus merindinggg terimakasih sudah membanggakan nama indonesia di kancah internasional para atlet indonesia*". (touched and goosebumps, thank you for making Indonesia proud on the international stage, Indonesian athletes). "*Betapa bangga dan bahagianya bisa berprestasi dan menyumbang emas olimpiade dan menyanyikan lagu indonesia raya di negara orang selamat rizki*". (How proud and happy to be able to achieve and contribute gold at the Olympics and sing the Indonesian national anthem in a foreign country. Good luck).
Seed 2	Indonesia (Indonesia) (0.059), *Emas* (Gold) (0.021), *Medali* (Medal) (0.019), Rizki (Athlete name) (0.018), *Bangga* (Proud) (0.017), *Terimakasih* (Thank you) (0.013).	
Seed 3	*Emas* (Gold) (0.044), *Bangga* (Proud) (0.025), Indonesia (Indonesia) (0.022), *Terimakasih* (Thankyou) (0.015), Veddriq (Athlete name) (0.015), *Olimpiade* (Olympic) (0.012).	
Topic 2 (Gratitude)		
Seed	Top words	Example comments
Seed 1	Indonesia (Indonesia) (0.055), *Emas* (Gold) (0.028), *Alhamdulillah* (Praise the God) (0.027), Veddriq (Athlete name) (0.018), Lenardo (Athlete name) (0.015), *Medali* (Medal) (0.013)	"*Alhamdulillah selamat kepada rizky juniansyah dan team angkat besi yang sudah meraih pencapaian tertinggi, medali emas.* (Thank God, congratulations to Rizky Juniansyah and the weightlifting team who have achieved the highest achievement, a gold medal.)" "*Alhamdulillah emas lagi keren rizky dr juara dunia kmrin sdh yakin bangat anak ini bakal naik podium di olimpiade paris moga atlet terakhir kita nurul akmal juga akan membawa pulang medali nanti yg laga di hari*
Seed 2	Indonesia (Indonesia) (0.035), *Emas* (Gold) (0.031), *Selamat* (Congratulation) (0.021), *Alhamdulillah* (Praise the God) (0.020), *Terimakasih* (Thank you) (0.017), *Bangga* (Proud) (0.012)	

Seed 3	Indonesia (Indonesia) (0.072), *Alhamdulillah* (Praise the God) (0.030), *Selamat* (Congratulation) (0.023), *Medali* (Medal) (0.013), Rizki (Athlete name) (0.013), Leonardo (Athlete name) (0.012)	*terakhir olimpiade".* (Thank God, Rizky's gold medal is cool from yesterday's world championship. I'm very sure that this kid will be on the podium at the Paris Olympics. Hopefully our last athlete, Nurul Akmal, will also bring home a medal in the match on the last day of the Olympics.)

Note: seed 1 = 1234; seed 2 = 2222; seed3 = 1543

Discussion

The main objective of this study is to find out what topics are most often raised by Indonesian netizens when watching athletes from Indonesia achieve prestigious achievements in the most prestigious international sports tournaments on social media. The results of the topic modeling analysis showed that two important topics were most often raised by Indonesian netizens: gratitude and national pride.

The results of this study indicate that gratitude is an important part of Indonesian society. This result is because Indonesia is a country that prioritizes religious values in all aspects of life. Pride and gratitude are positive emotions, but pride is greater for egoistic behaviour, while gratitude is closer to altruistic behaviour, such as donations [39]. In fact, pride does not always give rise to egoistic behavior. Two types of pride can lead to different social behaviors. The first is authentic pride, a feeling that arises from effort and hard work in achieving one's goals. Authentic pride will give rise to prosocial behavior. Conversely, hubristic pride, a feeling of superiority that leads to looking down on others, can give rise to selfish behavior [40]. Meanwhile, gratitude is an emotion that arises when someone receives a benefit from an external source and prompts the intention to repay the kindness [41]. The results of this research show that netizens in Indonesia feel proud when athletes receive medals in international sports competitions. They express gratitude to God for this achievement as a reward for the athletes.

Several netizens commented on YouTube videos for the government to give prizes to several athletes. This comment was driven by the emotion of gratitude. Gratitude also plays an important role in the context of sports. Gratitude can increase well-being, such as sports satisfaction and social support, and reduce ill-being, such as psychological distress and burnout [42]. Many positive comments appeared on YouTube about the victory of Indonesian athletes at the Paris Olympics. However, amid these comments, some negative comments were also made by netizens. This comment contains one of the athletes who received a medal because his opponent was injured. One of the media in Indonesia called it a "giveaway", and in the end, it also influenced public opinion. Comments containing gratitude can cover up these negative comments.

Gratitude also plays a role in the psychological formation of young athletes [43]. Therefore, cultivating gratitude is important because it can be a crucial factor in the development of athletes, ultimately benefiting the future of sports in a country [44]. One of the important psychological factors that must be developed from a young age is mental toughness, as it plays a crucial role in shaping the spirit of athletes to achieve their goals [45]. Additionally, gratitude is a factor that can also enhance mental toughness [46].

In addition, this study also found that the topic that emerged was national pride. This result is in line with previous studies which stated that sports achievements can be a source of national pride for people in a country [2], [21], [23]. The results of this study add information that national pride in sports achievements can also appear in comments given by netizens on social media. The results of this study show that the expression of pride in YouTube comments is national pride. This result is in line with the results of research conducted by [13], which states that achievements in the field of sports can increase national pride and prosocial society in a country. Netizens also showed that the sense of national pride was also influenced by playing the Indonesian national anthem on the Podium. These results strengthen the results of previous research conducted by [17] which stated that national attributes and symbols found at a sporting event can also increase the national pride of a country's people.

Several previous studies also support these findings. National pride can be influenced by events occurring during sporting competitions. Situations during sporting events, such as flags being raised, clothing, and face paint worn by supporters, can strengthen national pride [17]. This research shows that raising flags and playing the national anthem

after medals are awarded to athletes on the podium heightens netizens' national pride when they watch videos of athletes' victories broadcast on social media.

4. Conclusion

This study was conducted to explore the national pride of Indonesian netizens in responding to the victory of athletes winning medals at the 2024 Paris Olympics. This study analyzed comments on YouTube videos showing news about the victory of Indonesian athletes at the 2024 Paris Olympics. The results of this study indicate that Latent LDA was capable of detecting expressions of national pride in YouTube users' comments, particularly in response to the achievements of Indonesian athletes in international events. Other results show that Indonesian netizens also widely expressed gratitude for the gold medals they won. Indonesia is one of the countries that still upholds religious values. Therefore, gratitude appears as the most common comment in the comments. The results of this study can have implications for the government, including using athletes' achievements to increase netizen nationalism. Limitations of this study include the short data-collection period and the limited use of YouTube comments in news content. Furthermore, the data in this study include short comments and lack external validity. Furthermore, further research can be conducted on national pride across other social media platforms such as Instagram, Facebook, TikTok, and X. Future research can conduct multilingual checks, use a supervised classifier to detect national pride, and conduct user-level analysis to study the persistence of national pride over time.

References:

[1] T. Pawlowski, P. Downward, and S. Rasciute, "Does national pride from international sporting success contribute to well-being? An international investigation," *Sport Management Review*, vol. 17, no. 2, pp. 121–132, Apr. 2014, doi: 10.1016/j.smr.2013.06.007.

[2] S. Shibli, G. Ramchandani, and L. Davies, "The impact of British sporting achievements on national pride among adults in England," *European Sport Management Quarterly*, vol. 21, no. 5, pp. 658–676, Oct. 2021, doi: 10.1080/16184742.2021.1916553.

[3] P. K. H. Mo *et al.*, "Dualistic determinants of COVID-19 vaccination intention among university students in China: From perceived personal benefits to external reasons of perceived social benefits, collectivism, and national pride," *Vaccines (Basel)*, vol. 9, no. 11, p. 1323, 2021.

[4] Șerban Olah, "The National Identity, the National Pride and the Willingness to Fight for the Country. A Study Case of the Ukraine-Russia Conflict," *Eurolimes*, vol. 31, no. 31–32, pp. 165–173, 2022.

[5] V. Rutkauskas, "Factors affecting willingness to fight for one's own country: The case of baltic states," *Special Operations Journal*, vol. 4, no. 1, pp. 48–62, 2018.

[6] W.-C. Wu, "Distributive Unfairness, National Pride, and Willingness to Fight: Evidence From Taiwan," *J Asian Afr Stud*, p. 00219096231168047, 2023.

[7] A. Macintyre, H. F. Chan, M. Schaffner, and B. Torgler, "National pride and tax compliance: A laboratory experiment using a physiological marker," *PLoS One*, vol. 18, no. 1, p. e0280473, 2023.

[8] M. Hjerm, "National identities, national pride and xenophobia: A comparison of four Western countries," *Acta sociologica*, vol. 41, no. 4, pp. 335–347, 1998.

[9] R. Kurniawan *et al.*, "Sport and national pride: A literature review," *Retos: nuevas tendencias en educación física, deporte y recreación*, no. 65, pp. 44–54, 2025.

[10] A. Luksyte and D. R. Avery, "The Effects of citizenship dissimilarity and national pride on attitudes toward immigrants: Investigating mediators and moderators of intergroup contact," *International Journal of Intercultural Relations*, vol. 34, no. 6, pp. 629–641, Nov. 2010, doi: 10.1016/j.ijintrel.2010.04.001.

[11] H. Tajfel and J. C. Turner, "The social identity theory of intergroup behavior," in *Political psychology (Key Readings in Social Psychology)*, 1st ed., John. T. Jost and J. Sidanius, Eds., Psychology Press, 2004, pp. 276–293. doi: https://psycnet.apa.org/doi/10.4324/9780203505984-16.

[12] T. W. Smith and S. Kim, "National pride in comparative perspective: 1995/96 and 2003/04," *Int J Public Opin Res*, vol. 18, no. 1, pp. 127–136, 2006.

[13] L. R. Rosenzweig and Y.-Y. Zhou, "Team and Nation: Sports, Nationalism, and Attitudes Toward Refugees," *Comp Polit Stud*, vol. 54, no. 12, pp. 2123–2154, Oct. 2021, doi: 10.1177/0010414021997498.

[14] T. Shimokubo, "Hosting Olympic Games under a state of emergency: are people still proud of their country?," *International Journal of Sport Policy and Politics*, vol. 15, no. 1, pp. 147–161, Jan. 2023, doi: 10.1080/19406940.2022.2162102.

[15] F. Gassmann, J. Haut, and E. Emrich, "The effect of the 2014 and 2018 FIFA World Cup tournaments on German national pride. A natural experiment," *Appl Econ Lett*, vol. 27, no. 19, pp. 1541–1545, Nov. 2020, doi: 10.1080/13504851.2019.1696448.

[16] V. Halldorsson, "National sport success and the emergent social atmosphere: The case of Iceland," *Int Rev Sociol Sport*, vol. 56, no. 4, pp. 471–492, Jun. 2021, doi: 10.1177/1012690220912415.

[17] L. Brown, S. Richards, and I. Jones, "Sojourner perceptions of the St George Cross flag during the FIFA 2010 World Cup: A symbol of carnival or menace?," *Int Rev Sociol Sport*, vol. 49, no. 1, pp. 102–120, Feb. 2014, doi: 10.1177/1012690212450647.

[18] K. Hallmann, C. Breuer, and B. Kühnreich, "Happiness, pride and elite sporting success: What population segments gain most from national athletic achievements?," *Sport Management Review*, vol. 16, no. 2, pp. 226–235, Apr. 2013, doi: 10.1016/j.smr.2012.07.001.

[19] We are social, "Digital 2025: Indonesia," https://datareportal.com/reports/digital-2025-indonesia.

[20] M. Deori, M. K. Verma, and V. Kumar, "Sentiment analysis of users' comments on Indian Hindi news channels using Mozdeh: An evaluation based on YouTube Videos," *Journal of Creative Communications*, p. 09732586211049232, 2021.

[21] A. Elling, I. Van Hilvoorde, and R. Van Den Dool, "Creating or awakening national pride through sporting success: A longitudinal study on macro effects in the Netherlands," *Int Rev Sociol Sport*, vol. 49, no. 2, pp. 129–151, Apr. 2014, doi: 10.1177/1012690212455961.

[22] I. van Hilvoorde, A. Elling, and R. Stokvis, "How to influence national pride? The Olympic medal index as a unifying narrative," *Int Rev Sociol Sport*, vol. 45, no. 1, pp. 87–102, Mar. 2010, doi: 10.1177/1012690209356989.

[23] H. E. Meier and M. Mutz, "Political regimes and sport-related national pride: a cross-national analysis," *International Journal of Sport Policy and Politics*, vol. 10, no. 3, pp. 525–548, Jul. 2018, doi: 10.1080/19406940.2018.1447498.

[24] R. K. Storm and T. G. Jakobsen, "National pride, sporting success and event hosting: An analysis of intangible effects related to major athletic tournaments," *International Journal of Sport Policy and Politics*, vol. 12, no. 1, pp. 163–178, Jan. 2020, doi: 10.1080/19406940.2019.1646303.

[25] M. Kosinski, "Using Big Data," in *The Handbook of Social Psychology*, 6th ed., D. T. Gilbert, S. T. Fiske, E. J. Finkel, and W. B. Mendes, Eds., Situational Press, 2025, ch. 6.

[26] J. Žižka, F. Dařena, and A. Svoboda, *Text mining with machine learning: principles and techniques*. Crc Press, 2019.

[27] D. Wilkinson and M. Thelwall, "Researching personal information on the public web: Methods and ethics," *Soc Sci Comput Rev*, vol. 29, no. 4, pp. 387–401, 2011.

[28] Z. Antić, *Python Natural Language Processing Cookbook: Over 50 recipes to understand, analyze, and generate text for implementing language processing tasks*. Packt Publishing Ltd, 2021.

[29] Hobson. Lane, Cole. Howard, and H. Max. Hapke, *Natural language processing in action: Understanding, analyzing, and generating text with Python*. Manning, 2019.

[30] J. Silge and D. Robinson, "tidytext: Text mining and analysis using tidy data principles in R," *J Open Source Softw*, vol. 1, no. 3, p. 37, 2016.

[31] J. Silge and D. Robinson, *Text mining with R: A tidy approach*. " O'Reilly Media, Inc.," 2017.

[32] J. Atkinson-Abutridy, *Text Analytics: An Introduction to the Science and Applications of Unstructured Information Analysis*. Chapman and Hall/CRC, 2022.

[33] M. Röder, A. Both, and A. Hinneburg, "Exploring the Space of Topic Coherence Measures," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, New York, NY, USA: ACM, Feb. 2015, pp. 399–408. doi: 10.1145/2684822.2685324.

[34] J. Albrecht, S. Ramachandran, and C. Winkler, *Blueprints for text analytics using Python*. " O'Reilly Media, Inc.," 2020.

[35] D. Maier *et al.*, "Applying LDA topic modeling in communication research: Toward a valid and reliable methodology," in *Computational methods for communication science*, Routledge, 2021, pp. 13–38.

[36] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[37] R Core Team, *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, 2022.

[38] B. Grün and K. Hornik, "topicmodels: An R package for fitting topic models," *J Stat Softw*, vol. 40, pp. 1–30, 2011.

[39] T. H. Baek and S. Yoon, "Pride and gratitude: Egoistic versus altruistic appeals in social media advertising," *J Bus Res*, vol. 142, pp. 499–511, Mar. 2022, doi: 10.1016/j.jbusres.2021.12.066.

[40] M. J. J. Wubben, D. De Cremer, and E. van Dijk, "Is pride a prosocial emotion? Interpersonal effects of authentic and hubristic pride," *Cogn Emot*, vol. 26, no. 6, pp. 1084–1097, Sep. 2012, doi: 10.1080/02699931.2011.646956.

[41] M. E. McCullough, S. D. Kilpatrick, R. A. Emmons, and D. B. Larson, "Is gratitude a moral affect?," *Psychol Bull*, vol. 127, no. 2, pp. 249–266, 2001.

[42] N. T. Gabana, J. Steinfeldt, Y. J. Wong, Y. B. Chung, and D. Svetina, "Attitude of Gratitude: Exploring the Implementation of a Gratitude Intervention with College Athletes," *J Appl Sport Psychol*, vol. 31, no. 3, pp. 273–284, Jul. 2019, doi: 10.1080/10413200.2018.1498956.

[43] L. H. Chen, "Gratitude and Adolescent Athletes' Well-Being: The Multiple Mediating Roles of Perceived Social Support from Coaches and Teammates," *Soc Indic Res*, vol. 114, no. 2, pp. 273–285, Nov. 2013, doi: 10.1007/s11205-012-0145-2.

[44] A. Akbar *et al.*, "Individual psychological characteristics of under-17 football players in Indonesia and Malaysia: sport psychology approach," *Retos*, vol. 58, pp. 368–376, Jul. 2024, doi: 10.47197/retos.v58.107674.

[45] A. Akbar *et al.*, "The Role of Mental Toughness in Improving Achievement: The Perspective of Sport Psychology on Student-Athletes," *International Journal of Human Movement and Sports Sciences*, vol. 11, no. 6, pp. 1376–1382, Nov. 2023, doi: 10.13189/saj.2023.110623.

[46] A. Kiarostami, M. Jeddi, S. Roohinezhad, and R. Hakimi, "The Effectiveness of Positive Thinking Training on Mental Toughness and Social Well-being in Derelict Adolescents," *Iranian Evolutionary and Educational Psychology*, vol. 4, no. 2, pp. 398–409, May 2022, doi: 10.52547/ieepj.4.2.398.