

**Sentiment Analysis on Immigration(India-Canada)using Facebook Data**

By

**Atodaria Vaidehi**

A Major Research Project

presented to Toronto Metropolitan University

in partial fulfilment of the requirements for the degree of

Master of Science

in the Program of

Data Science and Analytics

Toronto, Ontario, Canada, 2024

Submitted To: **DR.Anatoliy Gruzd, DR.Ceni Babaoglu**

**AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A MAJOR RESEARCH PROJECT (MRP)**

I, Vaidehi Atodaria, hereby declare that I am the sole author of this Major Research Paper (MRP). This submission is a true representation of my MRP, inclusive of any necessary final amendments.

I grant Toronto Metropolitan University permission to lend this MRP to other institutions or individuals for scholarly research.

Additionally, I allow Toronto Metropolitan University to duplicate this MRP, either partially or in full, through photocopying or other methods, upon the request of other institutions or individuals for scholarly research.

I am aware that my MRP may be accessible to the public in electronic format.

Vaidehi Atodaria

**ACKNOWLEDGEMENTS**

I am deeply thankful to Dr Anatoliy Gruzd for his exceptional guidance and support throughout my research project. As my supervisor, **Dr Anatoliy** has been an invaluable source of knowledge and inspiration. His expertise and insightful feedback have immensely contributed to my academic growth and the success of this project. I am truly grateful for the opportunity to learn from such a distinguished mentor. Thank you, Dr Anatoliy, for your dedication and invaluable contributions to my educational journey

I would also like to express my sincere appreciation to **Dr. Ceni** for their insightful critiques and valuable suggestions, which greatly enhanced the quality of my work.

Thank you, Professor Anatoli and Dr.Ceni Babaoglu.

**Table Of Contents**

AUTHOR'S DECLARATION…………………………………………………………….2

ACKNOWLEDGEMENTS……………………………………………………………….3

List of Figures……………………………………………………………………………..5

1. Introduction……………………………………………………………………………..6

A. Background………………………………………………………………………..7

B. Research Question…………………………………………………………………8

C. Research Objectives……………………………………………………………….9

2. Literature Review……………………………………………………………………...10

3. Descriptive Analytics | Exploratory Data Analysis Acquisition………………………12

A. Data Acquisition / Data Source…………………………………………………...12

4. Data Analysis……………….……………………………………………………........13

5. Text Analysis…………………………………………………………………………..15

6. Methodology and Experiments………………………………………………………..16

A. Aim Study………………………………………………………………………..16

B. Results and Discussion…………………………………………………………...16

Sentiment Analysis…………………………………………………………….16

Topic Modeling……………………………………………………………...…19

7. Conclusion and Future Works….……………………………………………………....30

8. Appendix – A | Sample File & Github Link………………………………….…….….32

Sample File Github Link……………………………………………………………32

9. Appendix – B | List Of Fields In The Dataset…………………………………..……..32

10. References……………………………………………………………………………34

**LIST OF FIGURES**

Figure 1: Overview of Dataset

Figure 2: Top 20 Bia-grams

Figure 3: Bar-Graph Sentiment Distribution

Figure 4: Coherence & perplexity score by number of topics

Figure 5: Distribution of Topics

Figure 6: Frequent Word by Topics(2 and 4)

Figure 7: Frequent Word by Topics(7 and 11)

Figure 8:Sentiment Scores

Figure 9:Topic12 Weights and Sentiments

**1. Introduction**

This study aims to explore the nuanced discussions surrounding India-Canada relations on Facebook and their impact on individuals considering migration to Canada or those who have already settled there. With Facebook's expansive user base representing diverse perspectives and experiences, it serves as an ideal platform for this research endeavour. Drawing inspiration from a similar study titled Twitter-Based Sentiment Analysis and Topic Modeling of Social Media Posts Using Natural Language Processing, to Understand People’s Perspectives Regarding COVID-19 Booster Vaccine Shots in India, which utilized natural language processing (NLP) to analyse Twitter posts on COVID-19 booster shots in India, this project employs comparable Through sentiment analysis and topic modeling techniques, this study seeks to unravel the underlying sentiments and prevalent themes within discussions surrounding India-Canada relations on Facebook. By doing so, it endeavours to offer valuable insights into the social dynamics that shape the perspectives and decisions of both potential migrants and established residents[[1]](https://www.zotero.org/google-docs/?0PKRV5).

The research methodology involves a thorough examination of publicly accessible Facebook posts that discuss India-Canada immigration issues, aiming to analyze and decode the public's reaction to these discussions. By employing computational techniques such as sentiment analysis and topic modelling, the study seeks to dissect and understand the primary themes and sentiments expressed within these online dialogues[[2]](https://www.zotero.org/google-docs/?TcYqhi). The aim is to leverage Facebook's broad reach to gain diverse global perspectives on India-Canada relations, focusing on discerning public perceptions about the discussions' underlying causes and identifying potential constructive outcomes[[3]](https://www.zotero.org/google-docs/?3GoG3K).

Integrating TextBlob with other Natural Language Processing (NLP) techniques such as Latent Dirichlet Allocation (LDA) for topic modelling, this project aims to reveal the intricate dynamics between specific discussion topics and the sentiments they elicit among Facebook users. Adopting a dual-analysis framework similar to that utilized in the research conducted by Molenaar, Lukose, Brennan, Jenkins, and McCaffrey (2024), this endeavour will facilitate a comprehensive understanding of the discourse's emotional and thematic landscapes[[4]](https://www.zotero.org/google-docs/?97SZFM). Such an approach is poised to provide in-depth insights into both the underlying emotional nuances and the substantial content of these discussions, as presented in their study on using NLP to explore social media opinions on food security, which included sentiment analysis and topic modelling.

The aim is to improve discussions on India-Canada relations by adding new insights useful for academic research and shaping immigration policies. This effort seeks to deepen understanding of international relations, capture the experiences of people from various backgrounds settling in new places, and explore how social media affects global issues. Topic modelling helps organise and highlight the main topics within large texts, offering a clear summary of the conversations. Meanwhile, sentiment analysis examines the text's tone to understand the dominant emotions.

Together, these methods provide a thorough look at the sentiments and themes in public Facebook posts about India-Canada immigration, making this complex conversation more understandable for academic and policy work.

**A. Background**

The India-Canada relationship has seen its fair share of complexities, shaped by political, social, and trade factors. Recent events have brought these to the forefront, particularly with the G20 summit in New Delhi on September 9, 2023[[5]](https://www.zotero.org/google-docs/?Wr2Cid). During this summit, discussions between Prime Ministers Modi and Trudeau addressed the delicate topic of anti-India activities, bringing the Khalistan movement's issues into sharp focus. Adding to this, a parade float in Brampton, Ontario on June 4, 2023, depicting the assassination of former Prime Minister Indira Gandhi drew diplomatic ire and intensified bilateral tensions. These events, closely leading up to the data collection phase of this research, provide a critical context for understanding the sentiments and discussions captured in Facebook posts from September 4 to October 4, 2023[[6]](https://www.zotero.org/google-docs/?DAi3Ol).

These incidents have contributed to strained relations, impacting bilateral engagements such as trade negotiations. The postponement of a Canadian trade mission to India in October 2023 and the pause in discussions over a trade deal are indicative of the current frosty relations​​. The two nations have historically cooperated in various sectors, including trade, education, and technology, with bilateral trade reaching approximately $12 billion in 2023. However, issues like trade barriers, human rights concerns, and differing stances on geopolitical matters continue to pose challenges​.

The Indian community in Canada is one of the biggest immigrant groups playing a big part in this situation[[7]](https://www.zotero.org/google-docs/?9HkauJ). The project uses automated methods to collect Facebook posts It will pick out the main topics people are talking about and see how people feel about them, which helps us understand more about what is going on.

**B. Research Question**

This study investigates the nature of discourse on Facebook regarding the tensions between India and Canada. The primary objective is to analyze these conversations to identify the underlying sentiments, main topics of discussion, and the nature of interactions among participants. Facebook is chosen as the primary data source for this study due to its significant global user base and its role as a vibrant platform for public discourse, where diverse groups of people engage in discussions on a wide array of topics, including international relations. Its widespread adoption by individuals from different backgrounds, including those within the Indian and Canadian communities, makes it a rich repository of public opinion and sentiment on issues like the tensions between India and Canada. This platform allows for the analysis of real-time, authentic conversations, providing a comprehensive understanding of public perceptions and the dynamics of interactions related to the topics of interest. By focusing on Facebook, this study aims to tap into the grassroots level of public discourse, offering valuable insights into how these tensions are perceived and discussed among the general populace.

Advanced computational techniques, including machine learning and natural language processing, will be employed to scrutinize the content shared on Facebook. Sentiment analysis will be utilized to categorize the emotional tone of the discussions—whether they are predominantly positive, negative, or neutral[[8]](https://www.zotero.org/google-docs/?Dd9Uya). Concurrently, topic modelling will be applied to organize the conversations into distinct themes, facilitating an understanding of the principal subjects of discourse[9[]](https://www.zotero.org/google-docs/?uEKuTc). These tools will help us make sense of the complicated stories and feelings in these discussions and give us a clearer view of how people feel about the tensions between India and Canada.

This research focuses on applying computational methods to analyze online discussions, particularly to identify key themes and sentiments within conversations about tensions between India and Canada. Various algorithms and techniques will be evaluated for their ability to discern significant insights from social media discourse.

By analyzing the content and sentiment of these discussions, the study aims to uncover the influence of these tensions on perceptions of immigration and the experiences of the Indian diaspora in Canada.

**C. Research Objectives**

The primary research question this study aims to address is: "How do the conversations on Facebook reflect the underlying sentiments, main topics of discussion, and nature of interactions among participants regarding the tensions between India and Canada, and how do these elements inform our understanding of public behaviour and perceptions related to immigration and the experiences of Indian expatriates in Canada?"

Addressing this question, the outcomes of this research may inform policy-making, community support initiatives, and further academic inquiries into the relationship between social media, international relations, and the diaspora experience. Through an in-depth analysis of Facebook discourse, this study seeks to uncover nuanced insights into public sentiment and the collective narrative surrounding the India-Canada tensions, contributing valuable perspectives to the broader dialogue on these issues.

**2. Literature Review**

Social media platforms, particularly Facebook, have emerged as vital spaces for public discourse, offering insights into collective opinions and sentiments on a wide range of topics, including international relations and geopolitical issues. The capacity of these platforms to reflect and shape public sentiment makes them invaluable for research, especially in understanding the dynamics of international disagreements as seen through the lens of everyday users[[10]](https://www.zotero.org/google-docs/?lMmdRi).

In this literature review, we specifically focus on the methodologies pertinent to analyzing social media discourse, with an emphasis on sentiment analysis tools that are critical to understanding the emotional underpinnings of Facebook chats in the context of our study. Among these, the TextBlob library stands out as a popular Python tool for processing textual data. It offers a convenient interface for conducting sentiment analysis, making it an excellent choice for researchers looking to analyze the emotional content of large datasets.

The simplicity and versatility of TextBlob have been widely recognized in the field of natural language processing (NLP), as evidenced by its frequent citation in numerous studies. This underscores its utility and effectiveness in extracting meaningful insights from text-based communications.

TextBlob's sentiment analysis function, which applies a pre-trained model to ascertain the polarity and subjectivity of textual data, is a prominent feature that has facilitated research across various domains. Notably, this function has been effectively leveraged to probe public sentiment on a spectrum of issues. In the realm of academic discourse, the paper Sentiment Analysis and Text Analysis of the Public Discourse on Twitter about COVID-19 and MPox available in the MDPI journal Big Data and Cognitive Computing, stands as a testament to the utility of TextBlob[[11]](https://www.zotero.org/google-docs/?qWJL8p). The study delineated public sentiment regarding the health crises of COVID-19 and MPox, illustrating TextBlob's adeptness at categorizing sentiments in extensive text datasets. This illustrates the proficiency of TextBlob in mining and quantifying the positive and negative sentiments reflected in vast corpora of social media discourse, thereby highlighting the potential of TextBlob for comprehensive sentiment analysis in scholarly investigations.

Incorporating TextBlob into our methodology allows for a nuanced analysis of the sentiments expressed in Facebook chats related to the India-Canada issues. By quantifying the polarity scores of the texts, we can discern the overall emotional tone of the discussions, whether they lean towards positive, negative, or neutral sentiments. This quantitative sentiment data, combined with qualitative insights, can offer a comprehensive understanding of public opinions and attitudes towards the topics of interest.

Furthermore, the integration of TextBlob with other NLP techniques, such as topic modelling, can enrich our analysis. For instance, combining sentiment scores with thematic insights from Latent Dirichlet Allocation (LDA) can reveal the relationship between specific topics and the sentiments they evoke among Facebook users[[12]](https://www.zotero.org/google-docs/?EqiGat). This approach mirrors the methodology employed in the study by Patel and Jackson (2024), where the researchers combined sentiment analysis with topic modelling to investigate public discourse on health-related issues on social media platforms (Patel & Jackson, 2024).

The study by Wang and McCallum in 2022 showed that Latent Dirichlet Allocation (LDA) is good at finding themes in large amounts of text, like what people talk about in health forums online. This method is important because it can help understand big conversations on the internet, like those on Facebook about disagreements between countries, such as the issues between India and Canada.

By doing this, the project could give new insights into how social media serves as a place for people to discuss and argue about international issues. This could help in understanding how opinions form and change online, and could be useful for things like diplomacy and making policies that involve different cultures talking to each other.

**3. Descriptive Analytics | Exploratory Data Analysis**

**A. Data Acquisition / Data Source**

Our dataset for this research was gathered from Facebook, focusing specifically on conversations related to the geopolitical tensions between India and Canada. This collection was facilitated by CrowdTangle, a robust analytics tool, during a pivotal one-month period from September 4, 2023, to October 4, 2023. The comprehensive dataset comprises 22,969 posts and is enriched with a detailed array of over 40 metadata elements for each post.

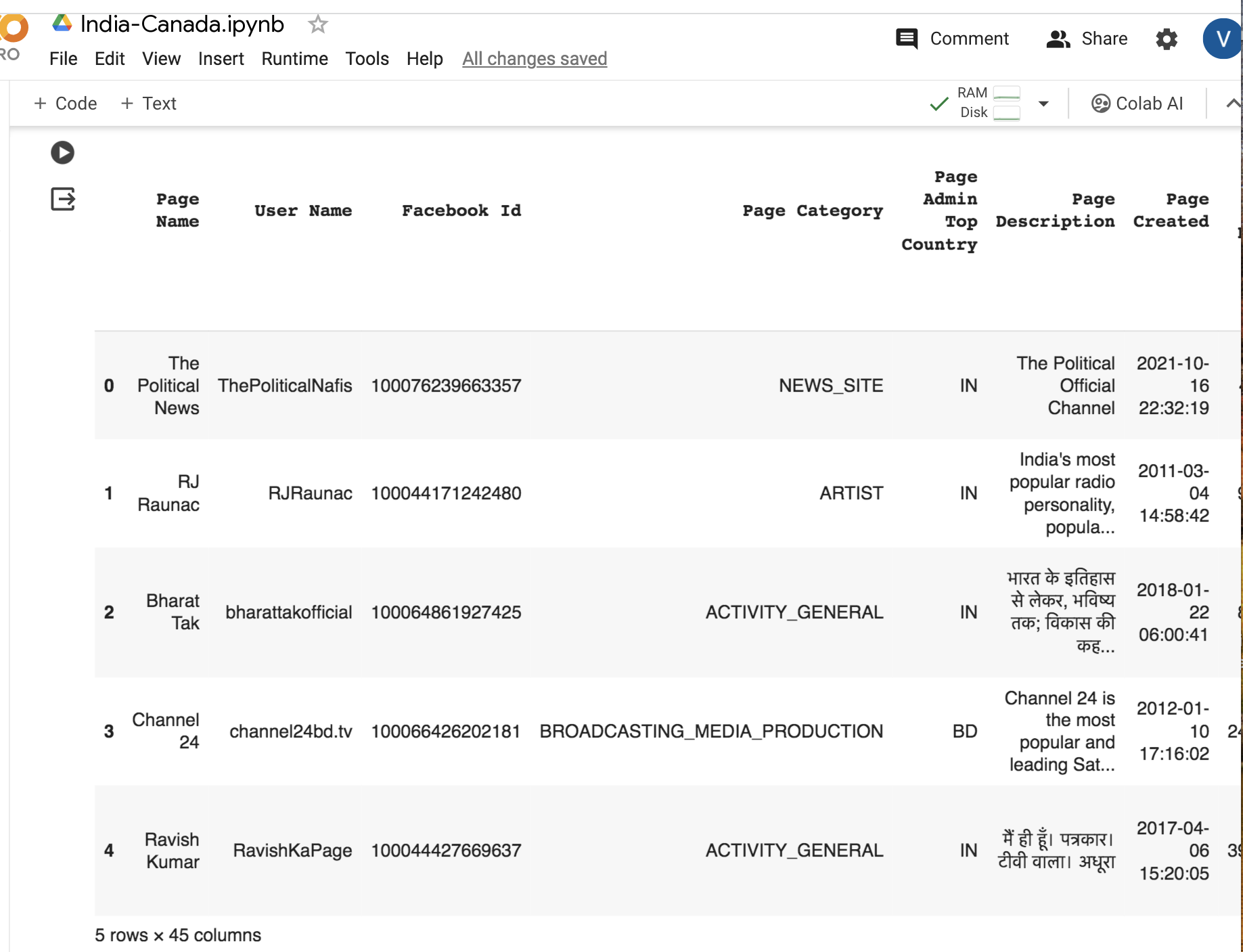
These elements include, but are not limited to, information such as page names, user names, Facebook IDs, and page categories. Additionally, the dataset provides a deep dive into engagement metrics, capturing likes, comments, shares, and a spectrum of emotional reactions from users, such as "The", "love", "wow", "haha", "sad", "angry", and "care" emoticons. This dataset allows for a nuanced exploration of the public discourses surrounding the tensions, offering insights into how these discussions unfold on social media.

CrowdTangle's utility in analyzing social media conversations, such as the India-Canada geopolitical tensions, is well-documented. For an in-depth look at its impact, Tess's analysis of McClatchy's social media strategy, available in the CrowdTangle Help Center, offers valuable insights[[13]](https://www.zotero.org/google-docs/?bFK23Z). Additionally, Kosnick's study on media's role in migration discussions in Germany further complements the understanding of media's impact on social discourse (Kosnick, K. 2000)[[14]](https://www.zotero.org/google-docs/?aZ7uk9). For this project, CrowdTangle proved invaluable in capturing a diverse array of public posts and pages discussing the India-Canada geopolitical tensions.

The content retrieved spans various types, including status updates, links, photos, and videos, shared by a wide range of entities such as news outlets, community groups, influencers, and public figures. This variety in the dataset shows many different sides of the conversation, including how people with different views and behaviours interact.

The data acquisition process through CrowdTangle was integral to the research, enabling the collection of a dataset that not only quantifies the volume of discussions related to the India-Canada tensions but also qualitatively assesses the nature of these conversations. Using CrowdTangle's metrics helped us understand the Facebook community's reactions to the India-Canada tensions. This method offered a full view of public sentiment, creating a strong base for our study.

**4. Data Analysis**



**Figure 1: Dataset**

Our analysis began with a comprehensive dataset, meticulously collated to include diverse content from Facebook discussions on India-Canada relations. This dataset, depicted in Figure 1, encompasses a wide array of digital interactions, from text posts to multimedia shares, providing a holistic view of public discourse on this subject.

During the initial phase of data preparation, we streamlined the textual content for analytical clarity, standardizing text to lowercase and preserving punctuation to maintain the authenticity of emotional expressions. This process, crucial for sentiment analysis, involved the strategic removal of stopwords to enhance the focus on pertinent themes, as outlined in the methodology section.

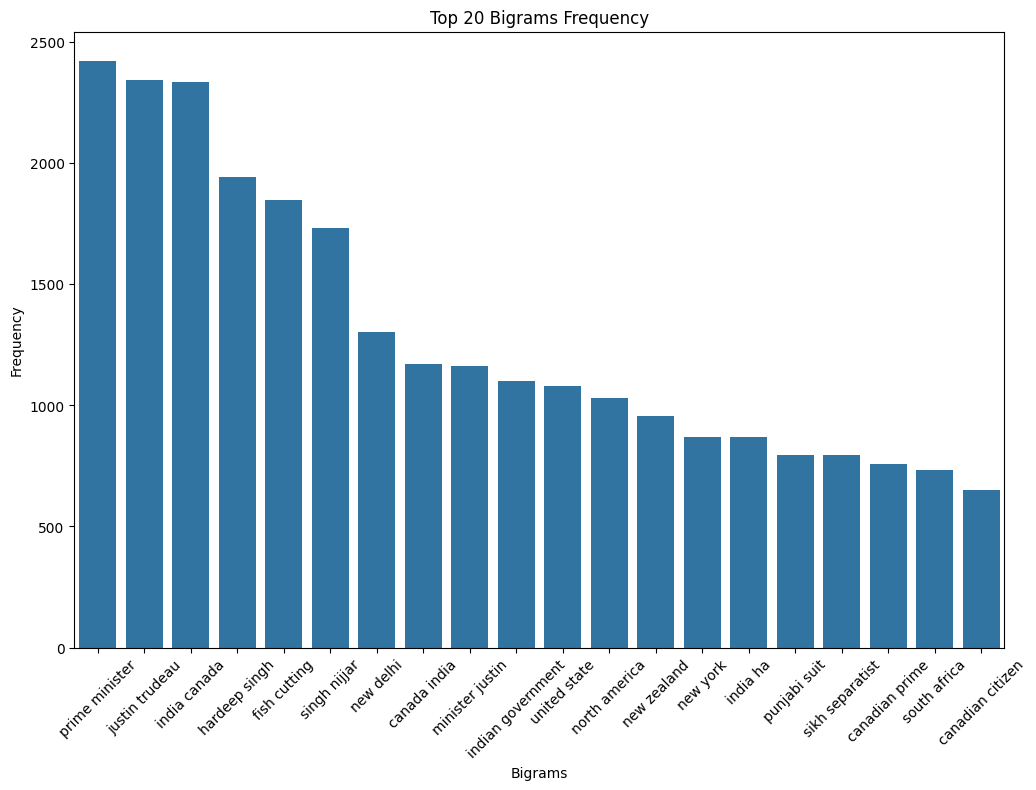
Subsequent steps involved lemmatization, a technique employed to consolidate variant forms of a word into its base form, ensuring uniformity in data analysis. This preparatory work set the stage for the dual analytical approach of sentiment analysis and topic modelling, aimed at dissecting the complex layers of public sentiment and thematic concentration within these digital dialogues.

In addition to sentiment analysis, which will classify posts as positive, negative, or neutral, topic modelling will be utilized to unearth the primary themes within the discussions. This technique will analyze the frequency and co-occurrence of words to identify key subjects being talked about, such as "trade agreements" or "diplomatic relations," providing a comprehensive overview of the conversation topics.

By integrating sentiment analysis with topic modelling and employing careful data-cleaning techniques, including the strategic removal of stopwords and lemmatisation, this study aims to offer a nuanced understanding of the sentiments and thematic focuses in the online discourse surrounding India-Canada relations.

**5. Text Analysis**

Before delving into sentiment analysis, a foundational examination of the text structure was imperative. This entailed analyzing word frequency and the occurrence of bigrams to identify dominant themes and sentiments within the discourse. Such preliminary analysis, visualized in Figure 2, was instrumental in highlighting significant linguistic patterns that underpin the public's response to India-Canada relations on social media.



**Figure 2: Bar-Graph Bigrams-top 20 words**

A bar chart showed the 20 most common pairs of words, helping us see how words connect and pointing out possible feelings or opinions in the text.

**6. Methodology**

1. **Aim of Study**

The primary objective of this study is to analyze and understand the nature of public sentiment and the predominant themes within online discussions related to India-Canada relations on social media platforms. By employing sentiment analysis, the study seeks to categorize the emotional tones of the conversations into positive, negative, and neutral sentiments, offering insights into the general mood and attitudes expressed by the participants.

Concurrently, through topic modelling, the study aims to identify and explore the main subjects and issues discussed within these online interactions. This dual-faceted approach is designed to provide a comprehensive understanding of the public discourse surrounding India-Canada relations, highlighting the aspects that resonate most with the online community and the emotional undertones characterizing these discussions.

1. **Results and Discussion**

**Sentiment Analysis of Facebook Discourse on India-Canada Relations: Unveiling Public Sentiments and Thematic Concerns**

The project aims to explore the intricate landscape of public discourse regarding the geopolitical dynamics between India and Canada as manifested through Facebook conversations. By delving into the sentiments and underlying themes within these dialogues, the study seeks to shed light on the broader societal perceptions and reactions to the unfolding events between these two nations.

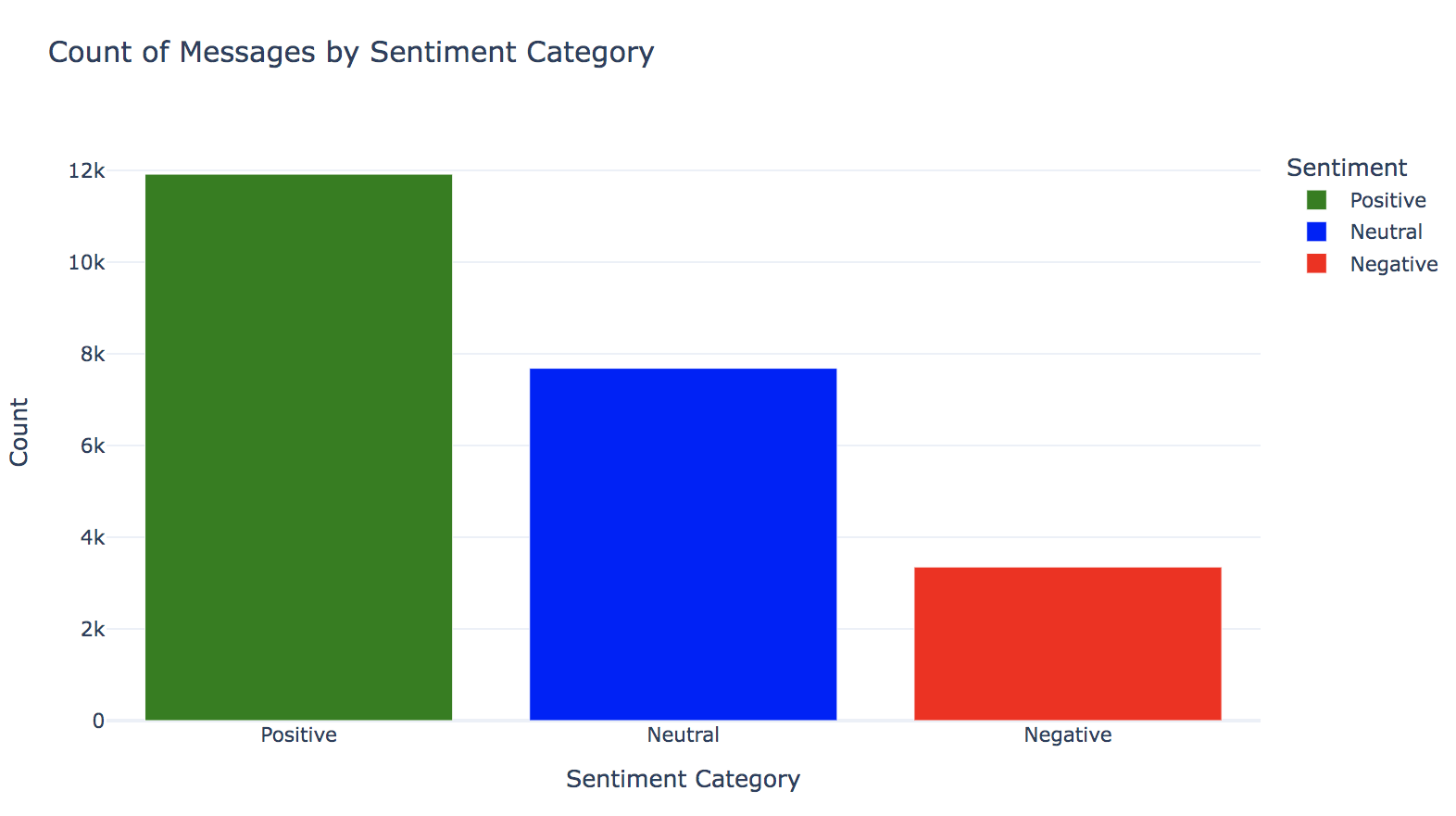
The utilization of Sentiment Analysis in this research is pivotal for discerning the nuanced opinions and emotional undertones prevalent in the bustling realm of social media. This method serves as a conduit to the grassroots-level discourse, offering a real-time snapshot of public sentiment. In the context of India-Canada relations, this analytical approach is instrumental in capturing the public's pulse, unravelling the aspects that resonate positively or negatively with the online community.

The methodology encompasses the aggregation of relevant social media discussions, followed by a meticulous data-cleaning process to eliminate non-essential elements. The core of the analysis employs TextBlob, a versatile tool known for its proficiency in processing natural language. TextBlob excels in its capacity to interpret the sentiment of textual content across multiple languages, a feature of paramount importance given the global nature of social media exchanges. It adeptly categorizes sentiments into positive or negative polarities and distinguishes between subjective opinions and objective statements[[15]](https://www.zotero.org/google-docs/?VCvRho).

For this study, sentiment scores were systematically classified into two categories: polarity and subjectivity. This bifurcation facilitated a more organized and focused analysis, leading to the removal of the original, broader Sentiment column to streamline the dataset. A subsequent review of the cleaned messages alongside their sentiment scores was undertaken to verify the precision of the data preparation phase.

Choosing TextBlob for sentiment analysis in this project is supported by its proven effectiveness in academic studies, like the one analyzing public sentiment about COVID-19 vaccines on Reddit. This reflects TextBlob's strong performance in extracting and interpreting complex sentiments from social media texts, which is crucial for understanding the nuanced public discourse on India-Canada relations[[16]](https://www.zotero.org/google-docs/?8xtQUy).

In conclusion, this research harnesses the power of sentiment analysis to illuminate the public sentiments and thematic concerns surrounding the India-Canada geopolitical discourse on Facebook. Through the strategic application of TextBlob, the study endeavours to provide a comprehensive and nuanced understanding of the prevailing attitudes and viewpoints, contributing valuable insights into the societal implications of international relations as reflected in digital conversations.



**Figure 3: Bar Graph Sentiment Distribution**

This chart shows three different coloured bars, each one representing how many messages fall into one of the sentiment categories The green bar stands for the number of positive or happy messages. The blue bar shows messages that are neutral, which means they're neither happy nor sad. The red bar represents negative or unhappy messages.

Overall, the chart provides a valuable visual summary of the sentiment distribution in the dataset, which can be used for various purposes, including understanding public opinion, identifying trends, and exploring relationships between sentiment and other variables.

**Unveiling the Themes: Topic Modeling of Facebook Conversations on India-Canada Relations**

Building on the sentiment analysis of Facebook discussions regarding India-Canada relations, which provided a broad view of the emotional landscape, a shift towards a more nuanced exploration is now undertaken.

The aim is to uncover specific themes and subjects within these conversations, providing a deeper understanding of the content that drives these sentiments. Topic modelling is employed for this purpose, revealing the predominant themes and offering a detailed view of the topics fueling the India-Canada dialogue on social media.

Selecting the appropriate model for this task was crucial. Latent Dirichlet Allocation (LDA) was chosen, guided by the comprehensive study by Egger R and Yu J (2022), which evaluates various topic modelling techniques, particularly for their applicability to social media texts[[17]](https://www.zotero.org/google-docs/?UGOwTK). This choice was influenced by a detailed analysis of the strengths and limitations of each model, including factors like coherence, perplexity scores, and computational efficiency, to ensure that the selected topics were not only meaningful but also representative of the underlying themes in the discussions.

This methodical approach, informed by prior research, allowed for navigating the complexities of topic modelling, ultimately leading to the choice of LDA. This decision was based not just on its ability to efficiently process the extensive corpus of Facebook discussions but also on its effectiveness in extracting insightful thematic patterns that resonate with the sentiments previously identified.

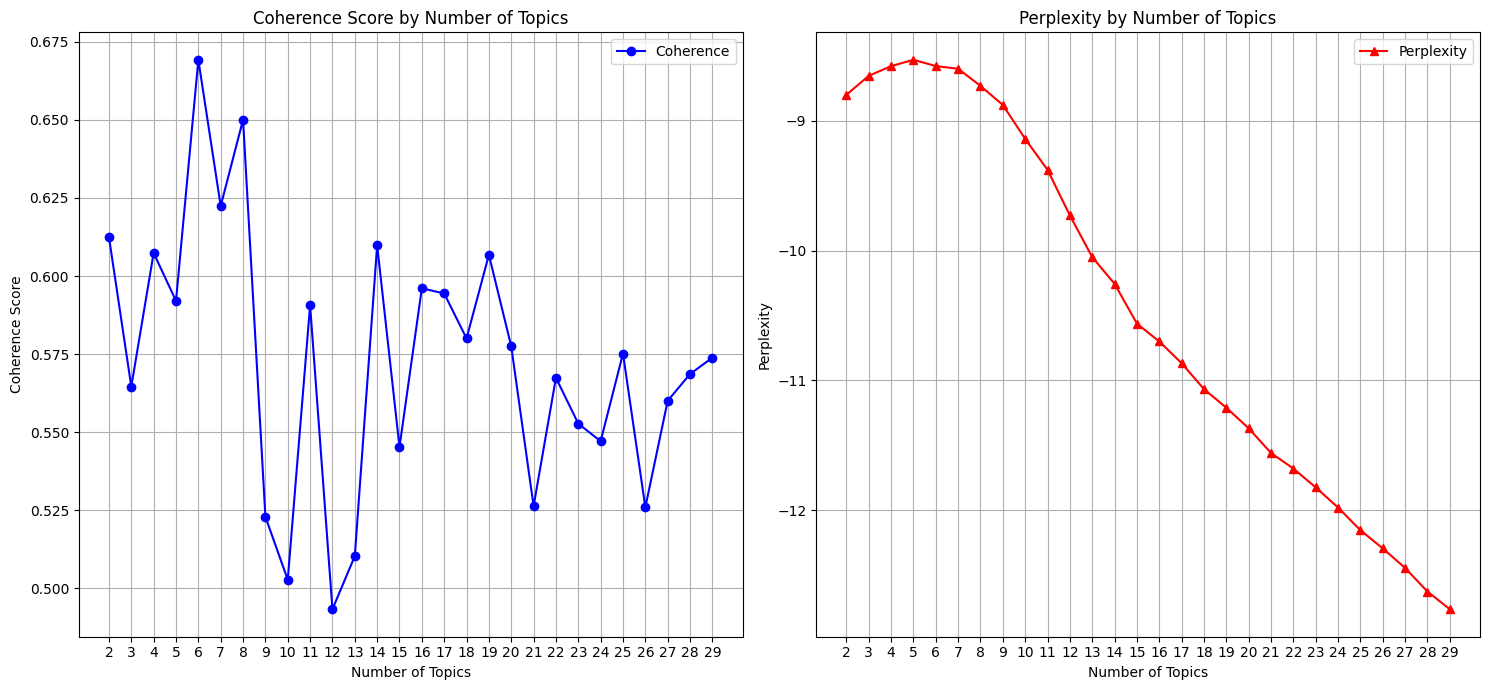
After the text was prepared by breaking it down, removing common but irrelevant words, and simplifying the words to their base forms, the actual analysis of topics within the social media posts commenced.

Various configurations were experimented with to identify the most effective approach. Through trial and improvement, the optimal settings were discovered that highlighted the main topics being discussed in the posts, providing a clearer understanding of the subjects of conversation.

The coherence score was instrumental in this phase. It assessed how meaningfully related the prominent words within each topic were, providing clarity and enhancing relevance. In conjunction with the coherence, the perplexity score measured the model's predictive power, indicating how well the model could anticipate and interpret new data.

Together, these metrics guided the selection of topics. Those with high coherence and low perplexity were chosen for their clear representation of the dataset's thematic structure. Once topics were identified, the most frequent words within each were selected based on their weighting, which denoted both the frequency of occurrence and the uniqueness within a particular topic.

This systematic selection of topics and words facilitated a granular understanding of the themes fueling the discourse between India and Canada on social media, offering insights into the specific content driving the identified sentiments.



**Figure 4: Coherence & perplexity score**

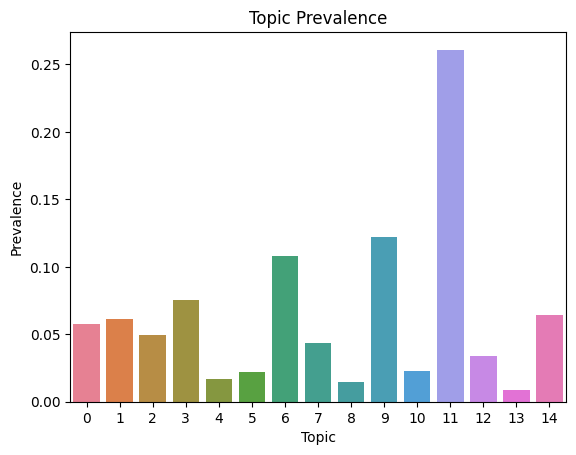
The two charts depict the Coherence Score and Perplexity across different numbers of topics determined by a topic modelling algorithm, likely Latent Dirichlet Allocation (LDA).

The first chart Coherence Score by Number of Topics displays how the coherence score varies with the number of topics. Coherence scores assess the semantic similarity and meaningfulness of the words within a topic.

The scores appear to fluctuate significantly as the number of topics increases. Typically, a higher coherence score is preferred as it indicates that the topics generated are more interpretable and thematically consistent. Peaks in the graph suggest potential numbers of topics that offer the most meaningful representation of the data.

The second chart Perplexity by Number of Topics shows the perplexity score, which is a measure of how well the probability distribution or model predicts a sample. A lower perplexity score generally indicates a better-performing model. The chart shows a declining trend, indicating that as the number of topics increases, the model's perplexity decreases. This trend typically continues to a point after which adding more topics doesn't necessarily lead to better predictive performance, often indicated by the perplexity score levelling off or increasing. In summary, these charts are used in tandem to find a balance between a high coherence score (for interpretability) and a low perplexity score (for predictive performance), which will guide the selection of an optimal number of topics for the model.

**Topic Prevalence Distribution in India-Canada Facebook Discourse Analysis**



**Figure 5: Topics**

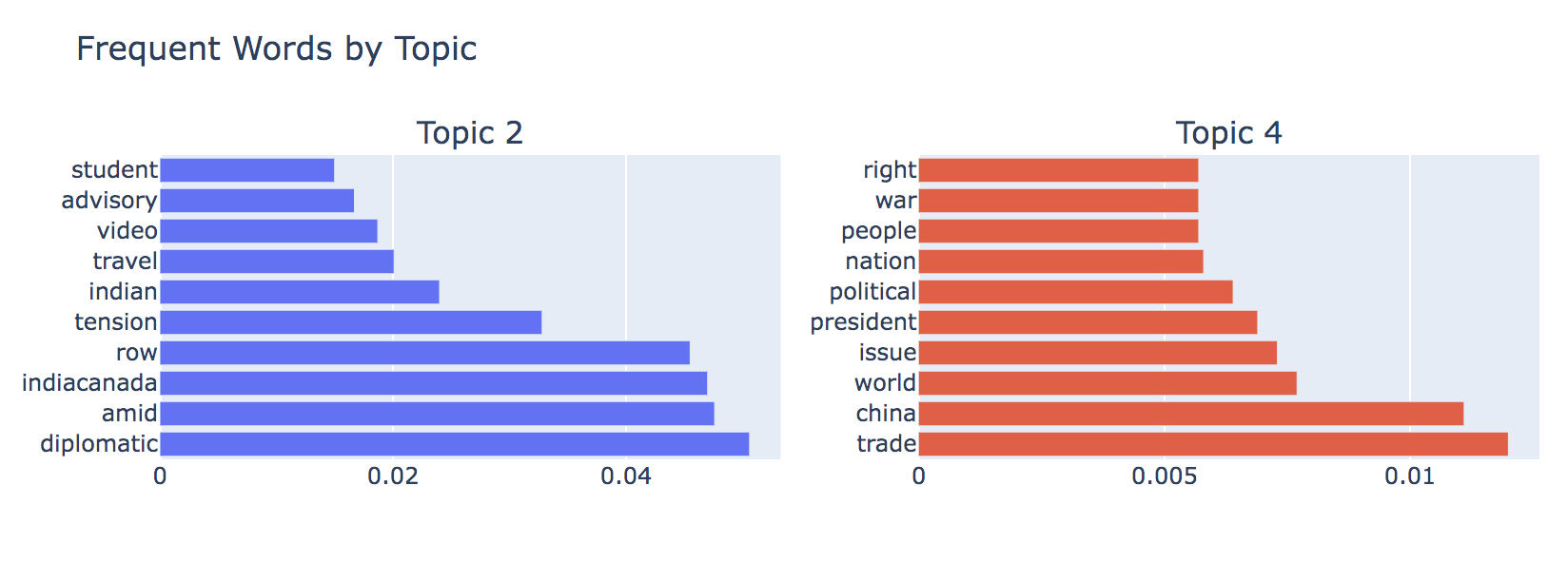
The chart visually represents the results of an LDA topic modelling analysis performed on a dataset of Facebook conversations concerning India-Canada relations. The x-axis categorizes the conversation into 15 distinct topics, numbered from 0 to 14, while the y-axis measures their

prevalence as a share of the total discourse. Notably, some topics have a significantly higher prevalence, indicating they are more frequently discussed or hold more importance among social media users. Topic 11 stands out with the highest prevalence, suggesting that it contains words or is associated with documents that are mentioned more often across the dataset.

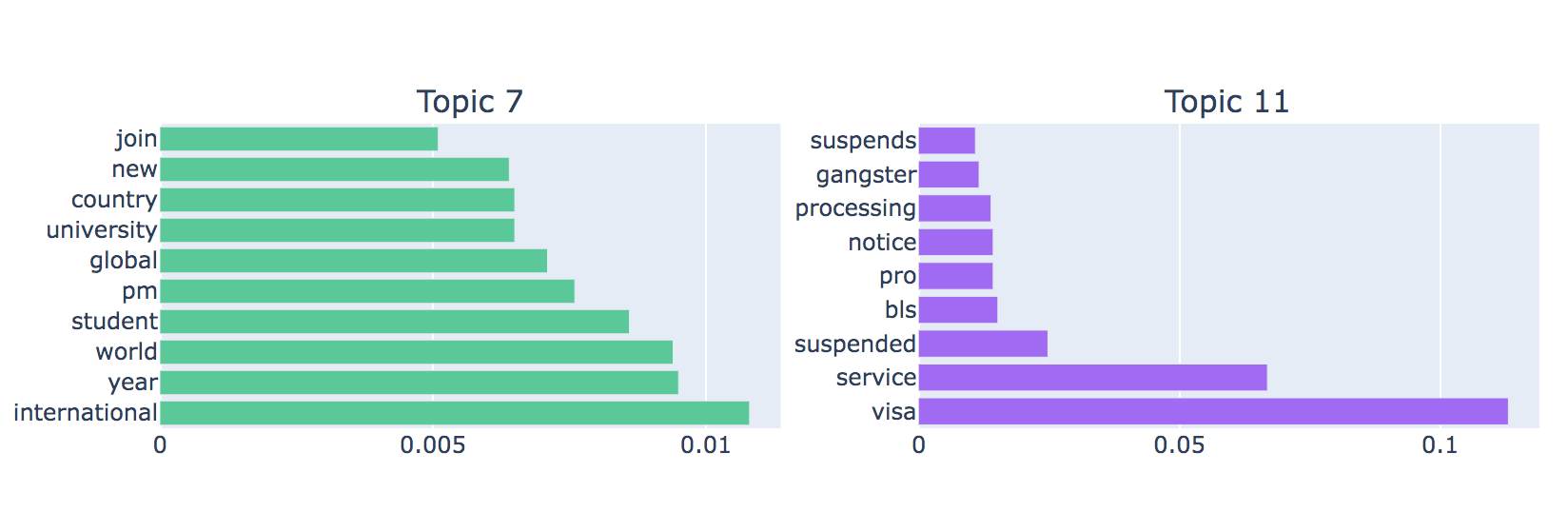
This could point to a topic that encapsulates current events, critical issues, or particularly resonant subjects between the two nations. On the other hand, topics like 0, 1, 4, and 5 are among those with lower prevalence, implying they are of lesser focus in the dialogue.

This visualization is critical as it identifies the relative weight of each topic within the public discourse. By highlighting the topics that dominate the conversation, it provides insight into the subjects that currently engage or concern the audience the most. Such insights are invaluable for decision-makers and leaders who aim to address public interests or respond to prevailing concerns in the realm of India-Canada relations.

**Most Frequent Words by Topic in Facebook Posts from India and Canada.**



**Figure 6: Topics (2 and 4)**



**Figure: 7: Topics (7 and 11)**

The chart displaying frequent words by topic unveils a layered dialogue within the Facebook conversations about India-Canada relations. Topic 2, with keywords like 'student', 'advisory', and 'diplomatic', likely centres on the interplay between education and diplomacy, suggesting a focus on advisories that impact student exchanges amid broader diplomatic relations. In contrast, Topic 4, with terms such as 'right', 'war', and 'trade', appears to delve into the realm of human rights, international conflict, and economic engagements, hinting at a discourse that could involve Canada's and India’s positioning on global trade policies and human rights issues, with a potential undercurrent of China's influence in these areas.

Further into the analysis, Topic 7's key terms like 'join', 'university', and 'international' point to an active conversation about academic collaboration and the experiences of international students navigating cross-border educational opportunities. This suggests a vibrant exchange on the prospects and challenges faced by students within the ambit of India-Canada educational ties. Meanwhile, Topic 11, surfaced by words such as 'allegation', 'prime', and 'government', indicates a politically charged discussion, possibly reflecting on the dynamics of governance and the nuances of political allegations affecting or arising from the policies of both nations.

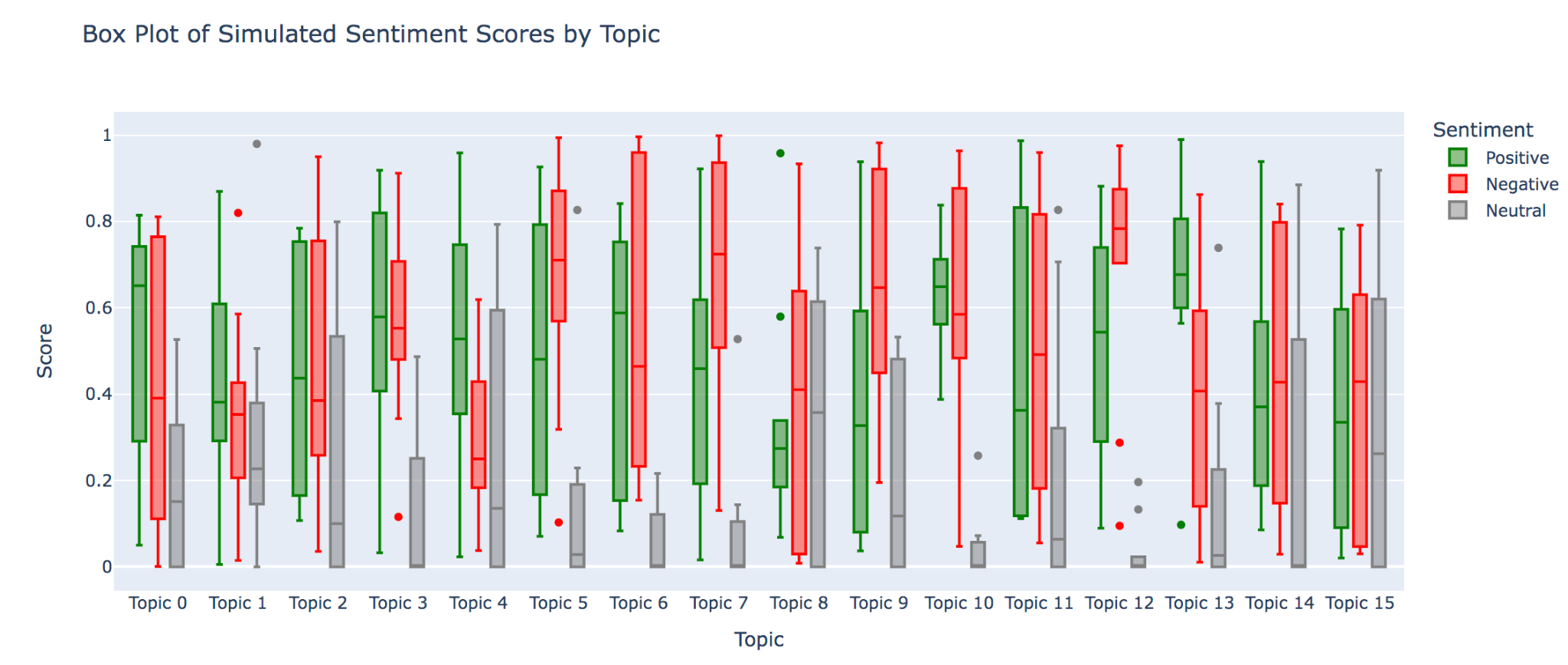
Lastly, Topic 15 brings a cultural and commercial dimension into the fold with terms like 'shop', 'online', and 'Punjabi', painting a picture of the Indian diaspora's commercial activities in Canada and the cultural narratives that unfold within this space. The focus here might be on the burgeoning online marketplace and the representation of Punjabi culture within the Canadian mosaic. Each topic, distinguished by its unique set of terms, contributes to a multifaceted understanding of the issues at the heart of the public discourse on India-Canada relations on social media platforms.

**Comparative Sentiment Analysis Across Topics**

A Box Plot of Simulated Sentiment Scores by Topic, provides a visual representation of sentiment analysis conducted across various topics extracted from a dataset..

The box plots display the distribution of sentiment scores for each topic, which have been grouped into three categories: positive, negative, and neutral sentiments.

These scores are likely derived from a sentiment analysis algorithm and have been normalized to fall between 0 and 1, where scores closer to 1 indicate a stronger presence of the respective sentiment.

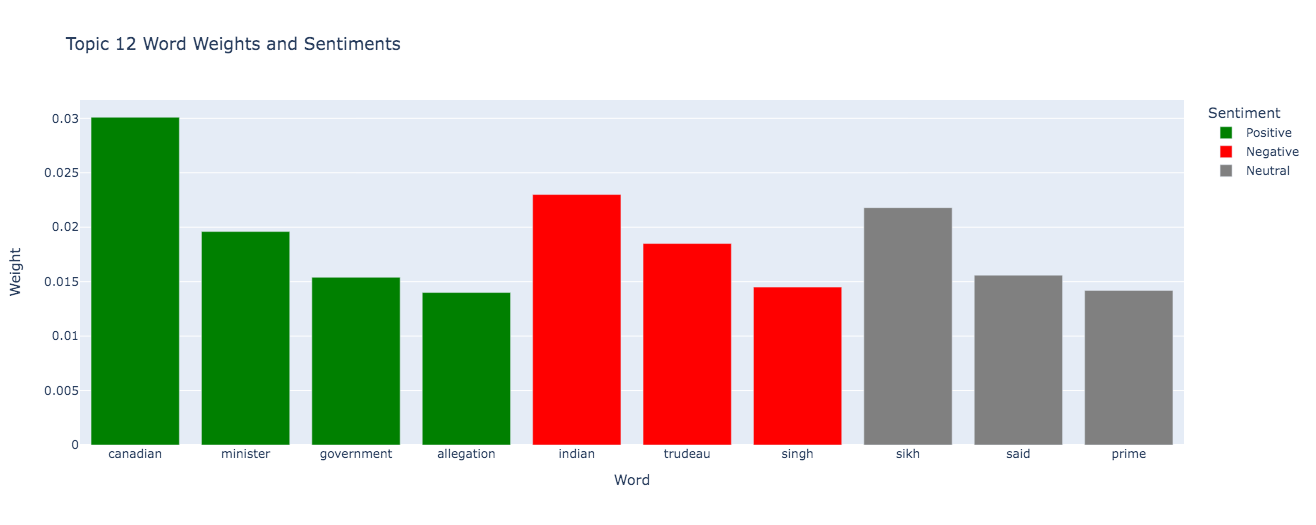


**Figure 8: Sentiment Scores**

By analyzing the chart, we can see that: Topic 10 has a relatively high median positive score, indicating that discussions within this topic tend to be more positive in sentiment.

Topic 5 has a wider box and a lower median score, suggesting a more varied distribution of sentiment with a mix of positive and negative discussions Topic 13 has a relatively low median negative score, implying that discussions within this topic tend to be more negative in sentiment.

**Examining India-Canada Relations: An Analysis of Keywords and Sentiments in Social Media Discourse**

Our primary focus in this project is to analyze the India-Canada tensions as reflected in Facebook posts. We identified Topic 12 as the most relevant to this topic based on the Chart.****

**Figure 9: Topic12 Weights and Sentiments**

Focusingon **Topic 12**  is crucial for our project as it focuses on the ongoing tensions between India and Canada, a key area of interest due to its implications on diplomatic relations, trade, and community ties. This topic, identified through the analysis of frequently occurring words and their significance, serves as a rich source of public sentiment and opinion reflected in Facebook posts.

By examining the discussions encapsulated within this topic, we gain insights into the prevailing perceptions and attitudes of individuals towards the actions and policies of both nations. The specific issues and themes emerging from this analysis not only shed light on the main factors fueling the tensions but also provide a nuanced understanding of the underlying causes. Such insights are invaluable for policymakers, diplomats, and analysts, offering a basis for informed decision-making aimed at addressing the concerns and improving bilateral relations. Moreover, tracking the evolution of the discourse around India-Canada tensions through Topic 12 enables us to identify trends and anticipate future developments, making it a pivotal aspect of our project that contributes to a comprehensive understanding of the dynamics at play.

Topic 12 gives us a clear picture of what people are talking about when it comes to the relationship between India and Canada online. Words like **Canadian, minister, and government,** along with names like **Trudeau** and **Singh** show us that a lot of the conversation is about politics and specific people. The word 'allegation' being used near 'Sikh' and 'prime' might point to certain events or issues that are causing a lot of talk[[18]](https://www.zotero.org/google-docs/?EV4jWt).

This situation shows how social media can strongly influence public discussions. Posts that involve strong emotions or talk about well-known political leaders tend to attract more attention, leading to a wider spread of information. This can quickly change how people understand and talk about certain topics.

The tension between India and Canada has grown recently due to accusations related to the death of Hardeep Singh Nijjar, a Sikh leader, in Canada. The issue escalated when Canada's Prime Minister, Justin Trudeau, spoke of "credible allegations" that Indian agents were involved in Nijjar's killing. This has led to a serious diplomatic issue, with both countries taking steps like expelling diplomats and India asking Canada to reduce its diplomatic staff. This conflict has strained relations but hasn't yet affected business and trade between the two countries. However, it has slowed down talks about a possible trade agreement[[19]](https://www.zotero.org/google-docs/?zd7ZGE).

The heart of this issue is the long-standing matter of Sikh separatism. This has been a point of conflict between India and Canada, mainly because many Sikhs live in Canada[[20]](https://www.zotero.org/google-docs/?PBLSeQ). India has often said Canada is too lenient towards Sikh separatists, which Canada denies.

The recent events have brought this issue to the forefront, marking a low point in the relationship between the two countries, with no quick solution in sight.

On the other hand, spotting negative sentiments can help address people's worries more directly. It might lead to clearing up misunderstandings, starting new public diplomacy efforts, or rethinking policies that might be causing issues.

Also, neutral sentiments might show where people are unsure or don't have enough information. This can lead to efforts to educate or communicate better to help people understand the more complex parts of the India-Canada relationship.

For our project, focusing on Topic 12 is key because it zeroes in on the ongoing issues between India and Canada, which is important for understanding their diplomatic relationship, trade, and how communities interact. This topic, which we picked out by looking at common words and their meanings, shows us how people feel and think about what both countries are doing. Understanding these issues and themes not only highlights the main reasons for the tensions but also helps us grasp what is causing them.

**7** . **Conclusion and Future Work**

The research offers insights into the dynamics of conversations surrounding the India-Canada conflict as observed in Facebook posts. Through topic modelling and sentiment analysis, key themes and the emotional undertones associated with this dispute have been identified, highlighting issues that resonate deeply with those affected.

The study plans to enhance its scope by incorporating data from "X (formerly Twitter)", various news platforms, and online forums. This expanded approach aims to provide a more comprehensive understanding of the broader implications of the conflict. A particular interest lies in examining the variances in discourse across different platforms and locales, potentially revealing shifts in public sentiment and focal points of discussions.

A closer examination of critical issues such as immigration policies and trade disputes is intended, to delve deeper into these dialogues. The insights gained aim to propose viable solutions that address the identified concerns, advocating for policies that promote fairness and open communication. The objective is to widely disseminate the findings, thereby enhancing awareness and fostering more informed debates on the India-Canada tension.

The research targets two primary audiences: policymakers and the general public. For policymakers, the study aims to present concise, actionable recommendations that tackle the main issues uncovered through analysis. This effort seeks to assist in the development of policies that are equitable and encourage open dialogue, ensuring that policy formulations are responsive to the sentiments and concerns of those affected by the dynamics of the India-Canada relationship.

By sharing the findings with the general public, the study strives to improve collective understanding and stimulate more informed discussions on this subject. This effort is crucial for elevating the level of public discourse and empowering individuals to participate more effectively in conversations about the India-Canada relationship.

The overarching aim of this strategy is to foster a more informed and equitable dialogue between policymakers and the public. By aligning policy with public sentiment and enhancing the quality of public discussions, the study hopes to contribute to a more sophisticated and constructive resolution of the issues at stake, promoting a culture of collaboration and inclusivity in addressing challenges in international relations.

Exploring specific contentious issues such as immigration policies and trade disagreements in greater depth is anticipated to uncover detailed nuances within these discussions. From the findings, the study aims to propose practical solutions that address the voiced concerns, advocating for policies that encourage equity and dialogue. Furthermore, by disseminating the insights through various channels, the aim is not only to raise awareness but also to provoke more informed debates concerning the India-Canada conflict.

**8 . Appendix – A | Sample File & Github Link**

Sample File

[**https://github.com/vaidehi1994/Vaidehi-Atodaria**](https://github.com/vaidehi1994/Vaidehi-Atodaria)

**9 . APPENDIX – B| LIST OF FIELDS IN THE DATASET**

1. Page Name
2. User Name
3. Facebook Id
4. Page Category
5. Page Admin Top Country
6. Page Description
7. Page Created
8. Likes at Posting
9. Followers at Posting
10. Post Created
11. Post Created Date
12. Post Created Time
13. Type
14. Total Interactions
15. Likes
16. Comments
17. Shares
18. Love
19. Wow
20. Haha
21. Sad
22. Angry
23. Care
24. Video Share Status
25. Is Video Owner?
26. Post Views
27. Total Views
28. Total Views For All Crossposts
29. Video Length
30. URL
31. Message
32. Link
33. Final Link
34. Image Text
35. Link Text
36. Description
37. Sponsor Id
38. Sponsor Name
39. Sponsor Category
40. Total Interactions (weighted — Likes 1x Shares 1x Comments 1x Love 1x Wow 1x Haha 1x Sad 1x Angry 1x Care 1x)
41. Overperforming Score

**10. References**

[1]<https://www.researchgate.net/publication/378344041_Analysis_of_Public_Sentiment_regarding_COVID-19_Vaccines_on_the_Social_Media_Platform_Reddit>

[2]<https://monkeylearn.com/sentiment-analysis/#:~:text=By%20combining%20machine%20learning%2C%20computational,and%20emotions%20from%20written%20language>.

[3]<https://www.researchgate.net/publication/273945598_The_Indian_Diaspora_Hindus_and_Sikhs_in_Australia>.

# 

# [4]Using Natural Language Processing to Explore Social Media Opinions on Food Security: Sentiment Analysis and Topic Modeling Study https://www.jmir.org/2024/1/e47826

[4]<https://www.thehindu.com/business/Economy/the-potential-economic-and-trade-fallout-of-strained-indo-candian-diplomatic-relations-explained/article67349441.ece>

[5]<https://www.ndtv.com/india-news/justin-trudeau-s-oldest-son-accompanies-him-to-india-for-g20-summit-4373903>

[6]<https://www.aljazeera.com/news/2023/9/15/canada-hits-pause-on-trade-mission-to-india-after-tensions-at-g20-summit>

[7]The Potential Economic and Trade Fallout of Strained Indo-Canadian Diplomatic Relations Explained.<https://www.thehindu.com/business/Economy/the-potential-economic-and-trade-fallout-of-strained-indo-candian-diplomatic-relations-explained/article67349441.ece>

[8]<https://www.researchgate.net/publication/371811054_Sentiment_analysis_Textblob_for_decision_making>

[9] Selecting the Number and Labels of Topics in Topic Modeling: A Tutorial.

[10]<https://www.researchgate.net/publication/324863961_Social_Media_and_the_Public_Sphere>

[11]Sentiment Analysis and Text Analysis of the Public Discourse on Twitter about COVID-19 and MPox https://www.mdpi.com/2504-2289/7/2/116.

[12]LDA-Based Topic Modeling Sentiment Analysis Using Topic/Document/Sentence (TDS) Model.

[13]<https://help.crowdtangle.com/en/articles/1214650-how-mcclatchy-used-crowdtangle-to-define-their-social-media-strategy>.

[14]Building bridges: Media for migrants and the public-service mission in Germanyhttps://journals.sagepub.com/doi/10.1177/136754940000300304

[15]Sentiment Analysis of Twitter Data

[16]<https://www.researchgate.net/publication/378344041_Analysis_of_Public_Sentiment_regarding_COVID-19_Vaccines_on_the_Social_Media_Platform_Reddit>.

[17]A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Postshttps://www.frontiersin.org/articles/10.3389/fsoc.2022.886498/full

[18]<https://www.aljazeera.com/news/2023/10/3/timeline-the-fraying-india-canada-relations-over-sikh-separatists-killing>.

[19]<https://www.indiatoday.in/india/story/india-canada-relationship-in-deep-crisis-perhaps-its-worst-ever-experts-2458616-2023-11-06>.

[20]<https://thediplomat.com/2023/09/india-canada-ties-are-caught-in-downward-spiral/>

21. Wang, Y., Huang, C., Wang, Y., & Chen, X. (2023). Leveraging BERT and VADER for Aspect-Based Sentiment Analysis in E-Commerce Reviews.

22. Gupta, S., Mittal, A., & Kumar, P. (2021). TF-IDF and Naive Bayes for Sentiment Analysis on Movie Reviews.

23. Gupta, R., Kumar, S., & Singh, M. (2020). Combining TextBlob and Support Vector Machines for Twitter Sentiment Analysis.

24. Sharma, A., Gupta, V., & Kumar, P. (2019). A Comparative Study of Machine Learning Algorithms for Sentiment Analysis.

25.Grootendorst, M. (2020). BERTopic: Topic Modeling Using BERT Embeddings.

26..Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation for Topic Modeling.

27..Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep

Bidirectional Transformers for Language Understanding.

28. Blei, D. M., & McAuliffe, J. D. (2008). Topic Modeling with Large Datasets.

29.Le, Q., & Mikolov, T. (2014). LDA2Vec: Topic Modeling with Document Embeddings.

30.Zhang, L., Wang, S., & Liu, B. (2020). Sentiment analysis of Twitter data: Insights into global events. *ResearchGate*.

<https://www.researchgate.net/publication/365618365_Sentiment_Analysis_of_Twitter_Data>