Charging Up Opinions: Unveiling Canadian Sentiments on Electric Vehicles (EVs) through Sentiment Analysis and Topic Modeling of Reddit Data

A paper submitted for
ITE Canada/WATT Consulting Group "Transportation in a Sustainable World" Student Award

Word Count: 3,661 Words

Ian Christopher Roman, EIT
Master of Science (MSc), Transportation Engineering Specialization
University of Calgary
Email: ian.roman@ucalgary.ca | **Phone:** 587-585-7414

Contents

List of Figures

Abstract

Social media has been everyone's way of communicating privately and publicly, where data is collected quicker than a nanosecond. Information from social media platforms can be used in various ways and leveraged to gain insights. One way to leverage social media data is text mining, where unstructured text data is analyzed and modeled to get general sentiment or determine common discussion themes. In Canada, electric vehicles (EVs) have been talked about quite a lot, either in public settings or online. The purpose of this study is to utilize Reddit thread and comment data to uncover Canadian sentiment regarding EVs and to model common discussion themes in this platform using topic modeling. It was found that there has been an increase in discussion about EVs since 2015, which has the same response as the number of EV registrations. About 65% of the threads and comments dataset are leaning toward a positive sentiment, and topic modeling reveals that the common themes Canadians are talking about are EV infrastructure, vehicle efficiency and affordability, charging facilities, and incentives. This study hopes that vehicle manufacturers and policymakers will leverage social media data and text-mining techniques to improve current practices and legislation that will help Canadians change their perception towards EVs and potentially increase adoption of these vehicles to achieve net-zero goals.

Chapter 1

Introduction

In the age of rapid technological advancement, adopting electric vehicles (EVs) stands out as a milestone in pursuing sustainable transport modes, safer travel alternatives, and improved overall driving experience (Suresha and Tiwari, 2021). According to Karami and Mackenzie (2020), social media plays a vital role in shaping public opinion and sentiment, particularly regarding technological advancements and societal shifts. Understanding public sentiments and opinions from various social media platforms becomes crucial as the transportation industry shifts toward safer, more efficient, and sustainable alternatives in a society where social media interactions are continuously growing.

In Canada, 94% of Canadians who spend time online have at least one (1) social media account, with top contents or topics that Canadians often interact with being political expression and public and mental health (Dubois et al. 2023). Social media sites have become the medium in which citizens voice political concerns and are exposed to political news, increasing the opportunities for political participation (Zuniga et al., 2014). Discussions of EVs on various social media platforms are often political and according to Zhou et al. (2020), the development of EVs is affected by economic, technological, policy, and social media network factors. Given the velocity and abundance of social media data, we can gauge citizens' awareness and opinions with specific policies and legislation using text data analytics (text mining) of online discussions about EVs.

This paper aims to perform basic text-mining techniques on *Reddit* threads (posts) and comment data to measure the *polarity* (*how negative or positive*) of each thread and comment and to perform topic modeling to determine common themes discussed within this online forum. This study will give insights to Canadian policymakers and help shape future EV legislation(s), considering what citizens' are voicing in online communities.

Chapter 2

Review of Related Literature

2.1 Canadian Perspectives on EVs

Attitudes and viewpoints of Canadians towards EVs are multifaceted based on some literature analyzing this topic on various social media platforms. A study conducted in 2017 revealed that while Canadian respondents have heard of key EV models, their familiarity and experience with EVs has remained low (Long et al., 2019), which suggests that while EV awareness has increased, there is still a lack of experience and familiarity on this innovation. Additionally, Barkenbus (2020) accentuated the growing usage of EVs implying that EVs are becoming a reality of the future which signifies a shit in public opinion towards the acceptance and transition to EVs in Canada. Moreover, another study has highlighted the potential for reusing EV batteries for *stationary energy storage*[1] in Canada, which implies a growing interest in sustainable energy solutions and integrating EV technology into public infrastructure (Catton et al., 2019). Mierlo and Maggetto (2007) also emphasized the benefits of EVs which indicated a favorable opinion towards EVs from an environmental perspective. This further suggests that Canadians support electric vehicles to reduce emissions and promote sustainable travel modes. Lastly, electric vehicle adoption has increased, indicating a positive trend toward integrating EVs into the Canadian automotive market (Nayak & Bohre, 2022).

However, along with these positive perspectives, the Canadian viewpoints on EVs also include some concerns. The high upfront cost of EVs compared to their diesel/gasoline counterparts remains

[1] **Stationary energy storage** refers to the use of rechargeable batteries to store energy in a fixed location, such as a charging station or grid infrastructure, to support various applications including grid stabilization, renewable energy integration, and smart grid management (Sick et al., 2015; Hesse et al., 2017; Thanh et al., 2018)

a significant barrier for the average consumer (Basso et al., 2022). Moreover, another concern was regarding charging infrastructure, particularly in rural communities and harsh winter conditions, which can hinder potential EV buyers (Transport Canada, 2022). Furthermore, limited EV model availability and perceived range anxiety compared to traditional diesel vehicles may raise questions about the practicality of EVs for all consumer's needs (Barkenbus, 2020). Acknowledging these concerns and the positive perspectives to comprehensively understand the multifaceted Canadian narrative surrounding EVs is important. This approach will inform effective policies that address barriers and encourage the adoption of EVs in Canada.

2.2 Social Media Text Mining

Social media text mining involves applying text mining techniques to extract valuable insights and information from the abundant user-generated content on social media platforms (Yu et al., 2021). This process involves various approaches, such as sentiment analysis, topic modeling, and text categorization, where the goal is to unveil meaningful patterns and sentiments on unstructured text data (SS, 2023). This form of analysis has been utilized not just for social media analysis but also in applications such as competitive analysis, crisis detection, and public opinion mining, which demonstrates the versatility of this tool in uncovering insights from diverse social media contexts (He et al., 2013; Rahman et al., 2023; D & Babu, 2017). The application of text-mining techniques in social media has also been a helpful aid in analyzing public responses to significant events such as the recent COVID-19 pandemic and having a better understanding of customer preferences and opinions towards products and services (Xie et al., 2021; Arifin & Purnama, 2023; Budi et al., 2017).

Text-mining techniques, such as sentiment analysis and topic modeling, have enabled researchers to gain insights into consumer sentiments, experiences, and service quality, contributing to a deeper understanding of customer preferences and behaviors (He et al., 2017; Ordenes & Zhang, 2019). In addition, integrating text mining with social media data has facilitated the identification of public sentiments toward specific topics, enabling key organizations to gauge public opinion and sentiment toward various topics, including government policies and crisis management (Yuyun et al., 2021).

2.3 About Reddit

Reddit is a social media platform hub for discussions, news, and content sharing. Academic scholars have increasingly turned to this platform for text mining and sentiment analysis, leveraging the platform's broad user-generated content to gain insights into many topics. For example, a study by Hodges et al. (2022) utilized text mining and sentiment analysis to examine discussions on gifted education, showcasing the platform's potential for sentiment analysis and social media research. Furthermore, Proferes et al. (2021) provided an overview of the diverse disciplines and methods used by researchers in accessing Reddit data, highlighting the platform's growing significance as a data source for research across various fields. Lastly, Gozzi et al. (2020) identified the platform's relevance in studying public responses to specific topics, such as health-related discussions during the COVID-19 pandemic, indicating its utility for understanding public interests and concerns.

In Canada, a wide array of topics are discussed on this social media platform regarding the multifaceted nature of EVs. These discussions investigate various aspects, such as the commercialization of lithium battery technologies and the consideration of performance, production, and cost (Zeng et al., 2019). Additionally, another focus delves into the environmental implications, emphasizing air quality improvement and climate change mitigation through policies (Slovic et al., 2015). Another prominent theme is the impact of road transportation as one of the major greenhouse gas contributors highlighting the significance of EVs in addressing environmental concerns in Canada (Panchal et al., 2016). Lastly, another theme of the discussion was Canada's commitment to selling zero-emission vehicles and trucks after 2035, which aligns with the global efforts towards carbon neutrality policies (He et al., 2023).

Chapter 3

Methodology

The Reddit platform has many subreddits[2] where thread (post) and comment data are accessible. This analysis will focus only on specific subreddits known to most Canadians and where discussions of "electric vehicles" or "EVs" are likely. This section will describe the process of extracting data from Reddit and briefly discuss the analysis and modeling techniques used to derive post sentiments and topic modeling.

3.1 Scraping Data from Reddit

Web scraping refers to the process of extracting data from websites. This study extracted Reddit data using `R` and `R Studio` and the package `RedditExtractoR`. Some functions within this package were utilized for this study. The first function, **`find_thread_urls()`**, takes input such as *subreddit name, time,* and *search keywords* to extract the data as a *data frame* and contains attributes such as *post date, author,* and *post URL*. The second function used for this study was **`get_thread_content()`**, where the **URL** column from the result of **`find_thread_urls()`** function call is the input. The output of the second function call will be a list of 2 elements: thread and comment data frames and exported to comma-separated values (CSV) files for more effortless data loading and analysis.

The subreddits of interest for this analysis were **CanadaPolitics**, **canada**, **electricvehicles**, and **EVCanada**. Note that these are selected only based on which subreddits have top searches for the words "EV"/"electric vehicle"/"EVs" and checked if there is sufficient text data from

[2] A **subreddit** is a specific community or forum within the social media platform Reddit dedicated to a particular topic of interest

these subreddits for analysis. **Figure 3.1** below illustrates the number of threads for each chosen subreddit.

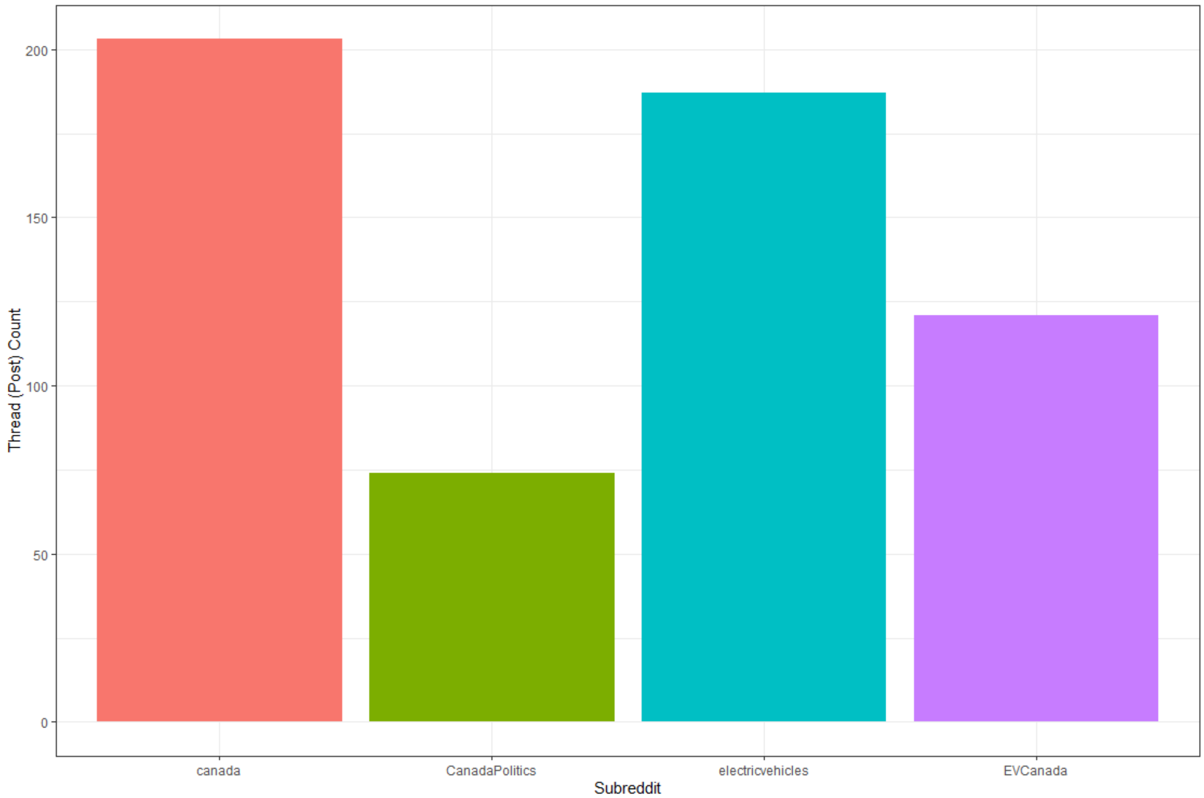

Figure 3.1: Distribution of Threads per Selected Subreddits

3.2 Text Mining Procedure

Since the `R` language was utilized in data preparation and modeling, **Figure 3.2** below will show the steps of the text mining procedure, from the input data to the visualizations that resulted from various analyses performed (Silge & Robinson, 2017). There are some processing techniques needed before starting any analysis involving text data. One of the essential steps is converting the text data into **tokens** (i.e., for each row of text data, the text is split into single words and inherits the unique ID assigned for that particular sentence or paragraph text data), which is the building block for subsequent analysis. A common challenge is that once the text data is *tokenized*, there will be an overwhelming amount of stop words (i.e., a set of commonly used words in a language; for instance, in English, the words "the", "a", "is", "are" are some examples) which can skew the effectiveness of any text mining method. A built-in list of stop words can be accessed in `R` to compare with the tokenized data frame and perform an *anti-join*[3] to filter these words from the data frame. The next step is to visualize the number of occurrences of each word by visualizing the transformed data to a word frequency bar plot, which gives insight into the words commonly mentioned on each thread and comment.

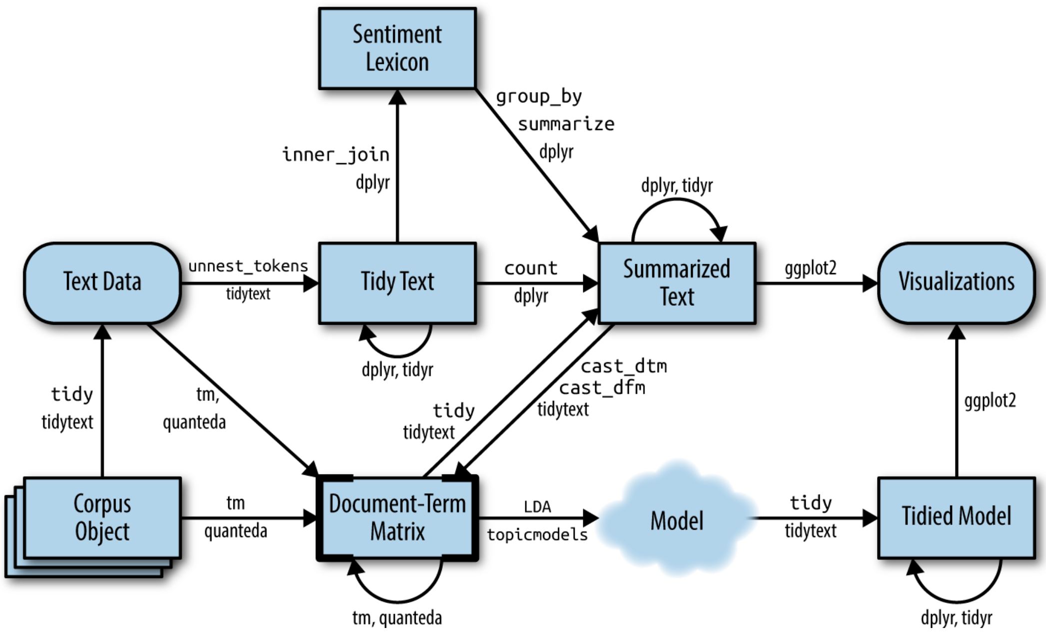

Figure 3.2: The Text Mining Process from Silge & Robinson (2017)

3.2.1 Sentiment Analysis

Sentiment analysis, also known as *opinion mining*, leverages the capabilities of natural language processing (NLP) to decode emotional undercurrents present in text data (Pang & Lee, 2008). Given the amount of text data on social media platforms, manual sentiment analysis can be intimidating for an average human being. Sentiment analysis can classify words into positive, negative, or neutral, offering valuable insights into the general underlying emotions of a given text data (Liu, 2012). The sentiment analysis process involves comparing a **lexicon dictionary**[4] where common

[3] An **anti-join**, also known as a left anti-join or right anti-join depending on the emphasis, is a type of join operation in databases that focuses on excluding matching rows between two tables based on a specific join condition.

[4] A **lexicon dictionary** is a structured and organized database of words, and in the context of sentiment analysis, a collection of words with their corresponding polarity (either positive, negative, or neutral, in numerical or categorical format)

words are scored numerically or categorically to get a sense of the *polarity*[5] of the word data. Using the `R` language, three (3) available lexicon dictionaries are **afinn**, **bing**, and **nrc**. Of the three, **afinn** is the only lexicon dictionary that scored numerically and was utilized for this analysis. **AFINN** is a lexicon of English words, with each word having an integer value ranging from negative five (-5) to positive five (+5). From 2009 to 2011, Finn Arup Nielsen labeled these words in this dictionary (Silge & Robinson, 2017).

The **AFINN** lexicon dictionary is then *inner-joined*[6] with the word token dataset to capture the respective score from the lexicon dictionary to the transformed dataset. This process will populate a new column containing the positive and negative word scores to determine the polarity of each word. The polarity of each thread/comment record is determined by grouping words according to their index. Subsequently, the new column is averaged to ascertain the overall polarity, reflecting the general sentiment of a thread or comment.

3.2.2 Topic Modeling

For collections of documents such as blog posts or social media posts, it is natural to group these data to understand their nature better. Topic modeling is an unsupervised classification method where documents are clustered on similar characteristics, and this process is synonymous with clustering algorithms (Silge & Robinson, 2017). Latent Dirichlet Allocation (LDA) is a well-known topic modeling algorithm. It is a probabilistic tool that determines latent topics within a collection of documents (e.g., published texts, social media posts, comments). LDA follows the logic that *every document is a mixture of topics,* and *every topic is a mixture of words* (Blei et al., 2003). LDA is better represented by the equation and **Figure 3.3** below, where:

$$p(\theta, \zeta, \omega | \alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(\zeta_n|\theta) \cdot p(\omega_n|\zeta_n, \beta)$$

- M is the total number of documents;
- N is the total number of words *in* a document;

[5] Negative, positive, or neutral

[6] An **inner join** is an operation that combines rows from two tables based on a matching value in a standard column. It creates a new table containing only the rows where the specified columns have the same value in both tables.

- ω or w is the observed words *in* a document;
- ζ or z indicates the specific topic from which *each* word in a document is drawn;
- θ is the topic proportions for *each* document;
- β is the word distribution for *each* topic and;
- α is the topic proportions *across* documents

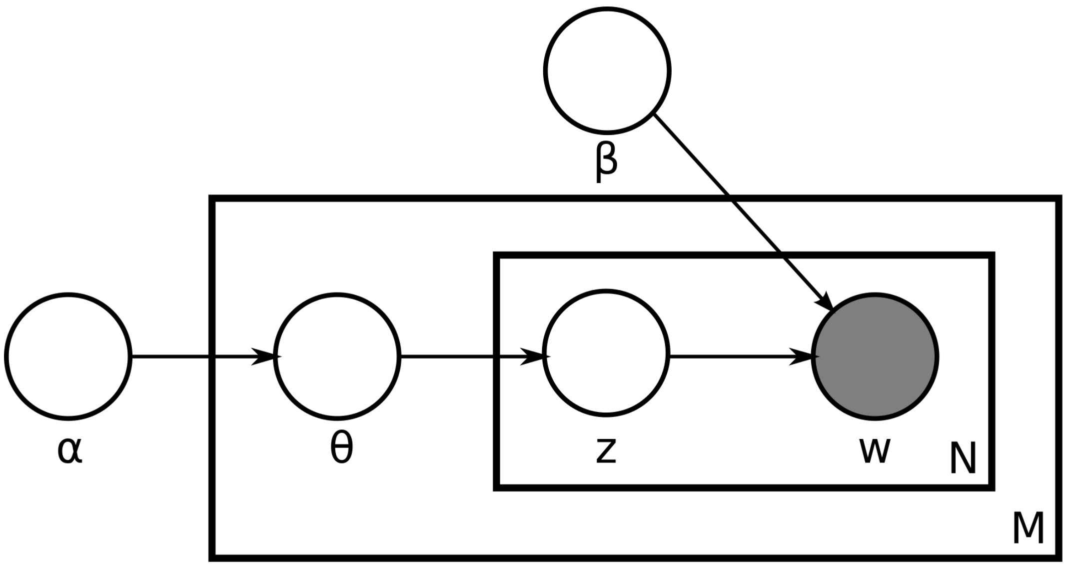

Figure 3.3: Graphical Representation of Latent Dirichlet Allocation (Blei et al., 2003)

For this study, the complexities above are captured through the `tm` package in `R` and were utilized to perform topic modeling. For this analysis, it was simplified (for demonstration purposes) to model only the top 3 topics on both threads and comments, and the β parameter and the number of words were used to infer topics for this analysis by plotting the model results in a bar plot for each topic.

Chapter 4

Text Mining and Discussion

As mentioned in the literature review and methodology process, this analysis will focus only on select subreddits where most Canadians post opinions on specific topics, such as discussions on electric vehicles. This section is structured as follows: a simple exploratory analysis of extracted Reddit data looking at trends and patterns followed by tokenization, a technique to split words in a sentence and paragraph as an individual observation, and then visualizing the most common words in both threads and comments. The last sections will showcase the result of sentiment analysis and topic modeling, where general sentiments and common topics from posts and comments are unveiled.

4.1 Exploratory Data Analysis

Reddit threads and comments data have timestamp information, which can be used to visualize the number of observations for each dataset over time. It was also valuable to visualize registrations of EVs throughout the year to have a sense of EV usage in Canada. The data from Statistics Canada was used to compare the number of threads/comments and EV registration over time. **Figure 4.1** below illustrates this trend, and it can be noticed that since 2015, discussions about EVs and EV registration have continued to grow up until now. Note that the data from Statistics Canada was limited to data from 2011 to 2022, hence the flat-line behavior shown in some portions of the graph. It is also worth noting that Reddit data was extracted in January 2024; therefore, the number of threads displayed on the graph only partially represents this year. Additionally, **Figure 4.2** below compares trends between the number of observations on the threads and comments dataset, in which both datasets follow the same behavior.

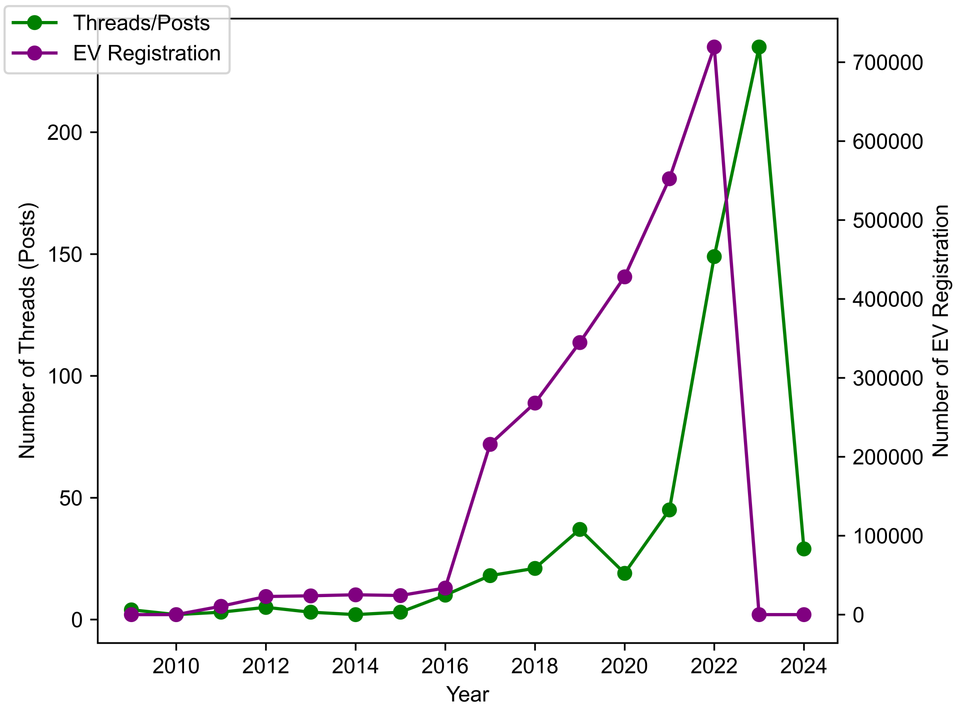

Figure 4.1: Reddit threads and EV registration in Canada from 2009 to Present

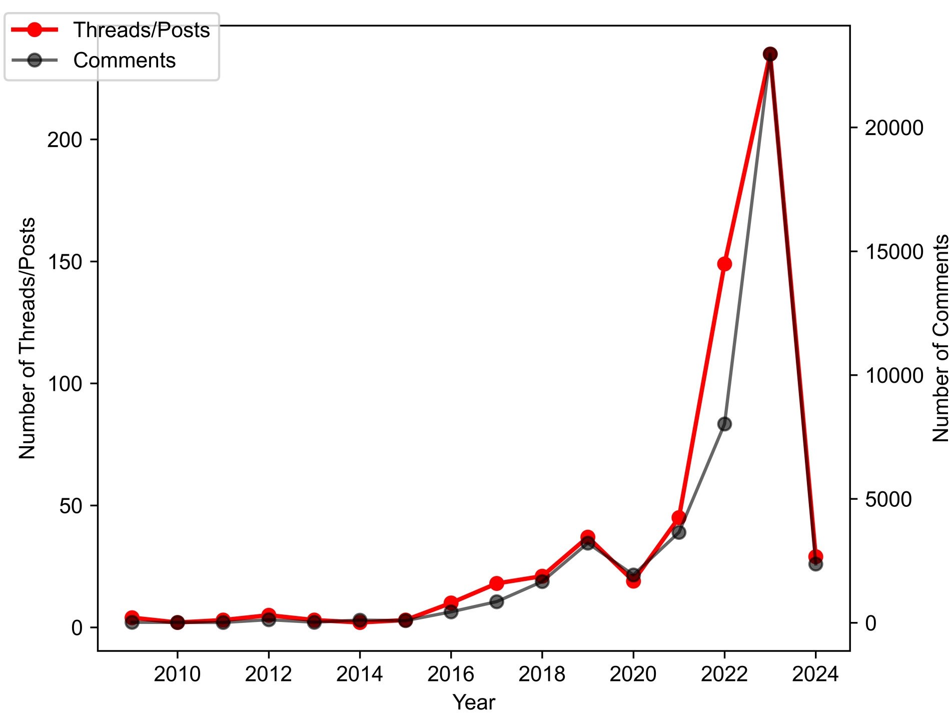

Figure 4.2: Reddit threads and comments about EVs from 2009 to Present

4.2 Tokenization and Word Frequency

Threads and comments datasets were processed such that each thread/comment was split into words, and each occupied a single row (tokenization). Once both datasets are tokenized, the stop words (e.g., the, a, is, are) are removed from this set and will skew the results of the subsequent analysis. Once cleaned, a bar chart of the most frequent words on threads and comments was plotted. A constraint was put on threads only to include a word count of at least 50 and comments with a word count of at least 2000. Note that there are far more comments than threads, so these constraints were implemented to visualize a reasonable amount of words in the bar charts. **Figures 4.3** and **4.4** illustrate the most common words appearing on threads and comments. For the threads dataset, the ten (10) most common words are *federal*, *trip*, *stations*, *gas*, *newsletter*, *drive*, *energy*, *ioniq*, *home*, *canadian* while the comments dataset's ten (10) most common words are *market*, *miles*, *4*, *charger*, *batteries*, *driving*, *1*, *ice*, *credit*, *power*. These might not hold much meaning now, but in a later section (**4.4 Topic Modeling**), the topic from a set of words (similar to **Figures 4.3** and **4.4**) will be used to infer common topics on threads and comments.

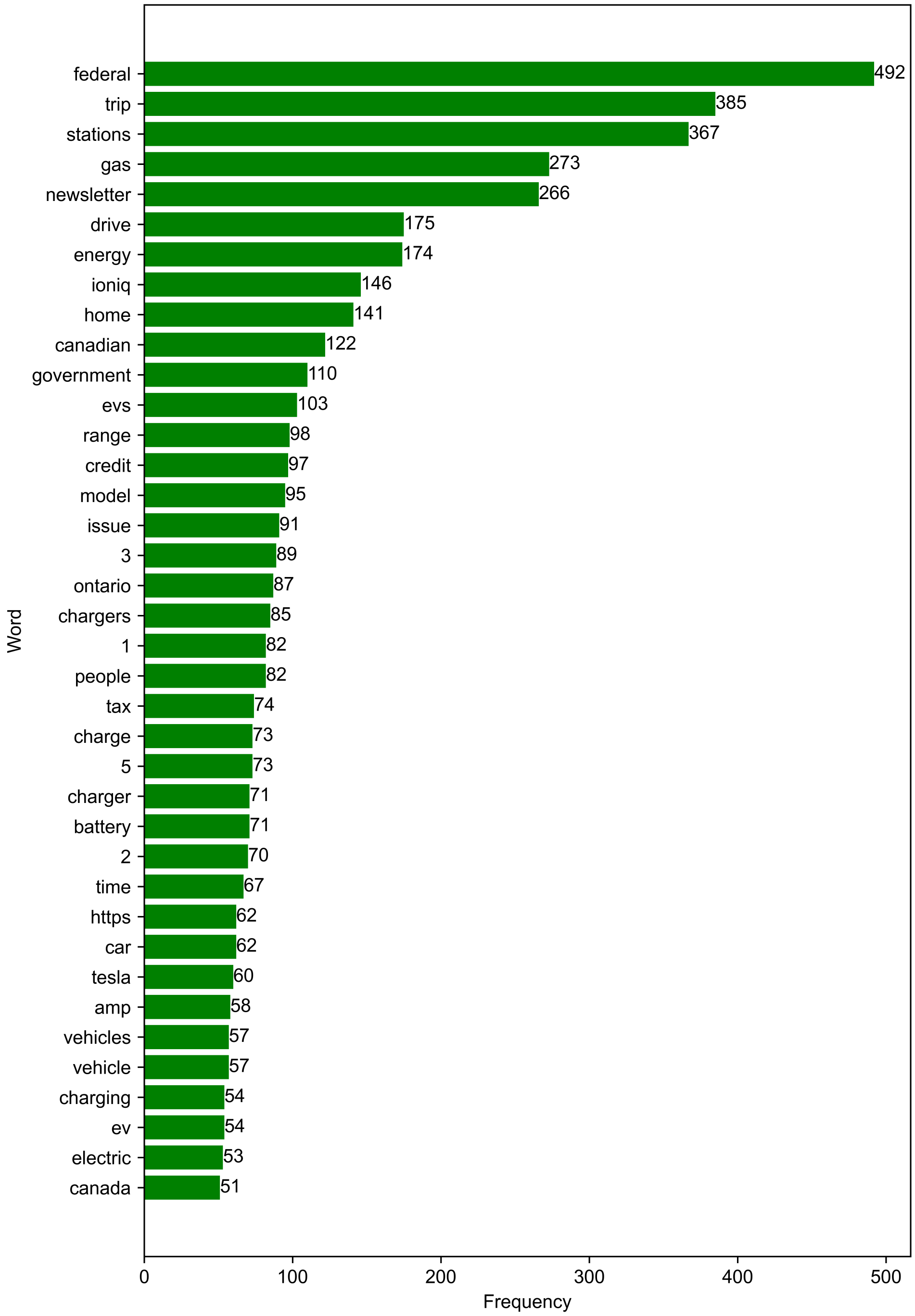

Figure 4.3: Word Frequency - Threads Data

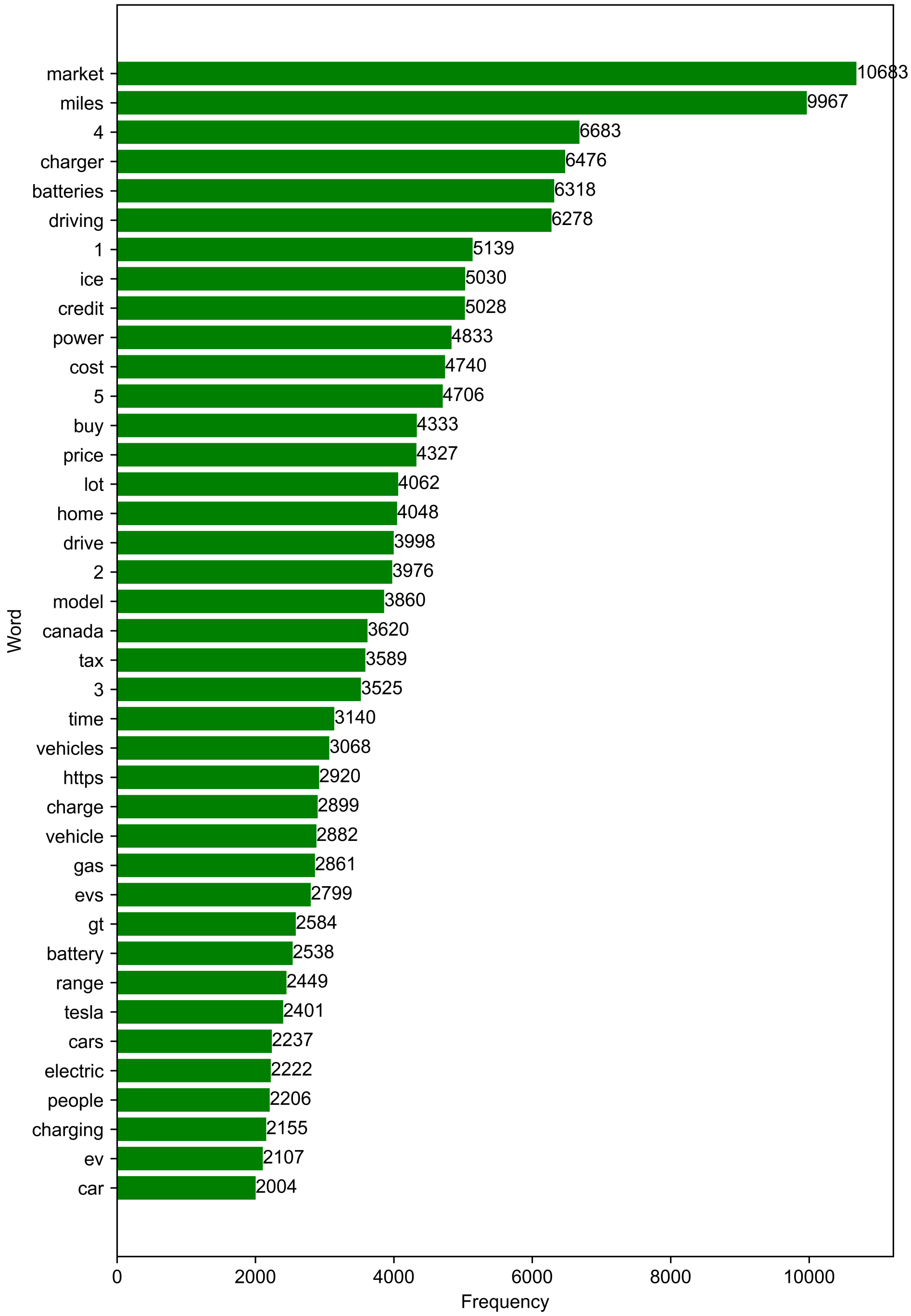

Figure 4.4: Word Frequency - Comments Data

4.3 Sentiment Analysis

The general sentiment for each thread and comment can be determined using the tokenized words from the threads and comments dataset. Using the **AFINN lexicon dictionary**, each word was scored with a positive or negative number. Once scored, the words were grouped according to their unique identifier (post or comment ID), and scores were averaged to get the overall thread/comment sentiment. The resulting polarity of individual threads/comments is shown in **Figures 4.5** and **4.6**, where each thread/comment is plotted with their average sentiment. The plots are very noisy (especially **Figure 4.6**), and they cannot show the proportion of threads/comments that were positive, negative, or neutral. To supplement this graph, another bar plot showing these proportions is illustrated in **Figures 4.7** and **4.8**, where 69% and 61% of the threads and comments data, respectively, are classified as *positive*. While a large proportion of threads and comments are on the positive scale, a significant amount of threads and comments have negative sentiment (28% and 33%, respectively), which confirms that while EVs have a positive impact as a sustainable transportation mode, it may be confronted with some challenges such as affordability and performance under harsh winter conditions (Basso et al., 2022; Transport Canada, 2022). It is also worth noting that the *neutral* category is for threads/comments with an average sentiment of zero (0). This happens when words from the dataset do not correspond to a word in the lexicon dictionary, leading to a polarity score of zero (0). The *neutral* category does not necessarily mean that the thread/comment under analysis has a neutral sentiment, but more so, the average sentiment value equates to zero (0).

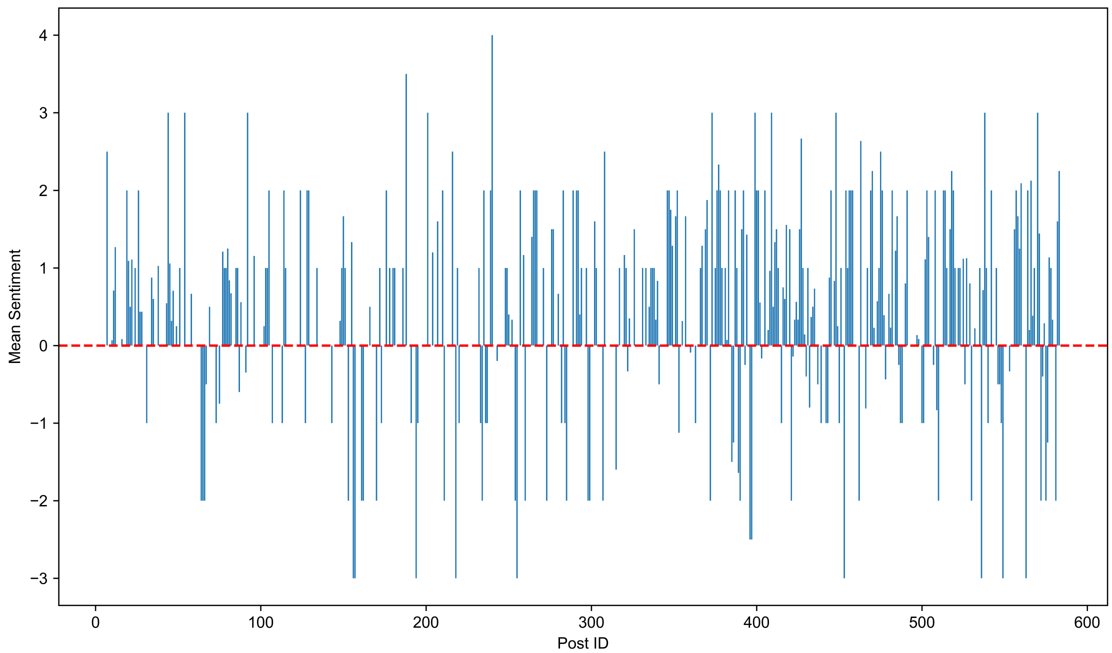

Figure 4.5: Average Sentiment per **Threads**

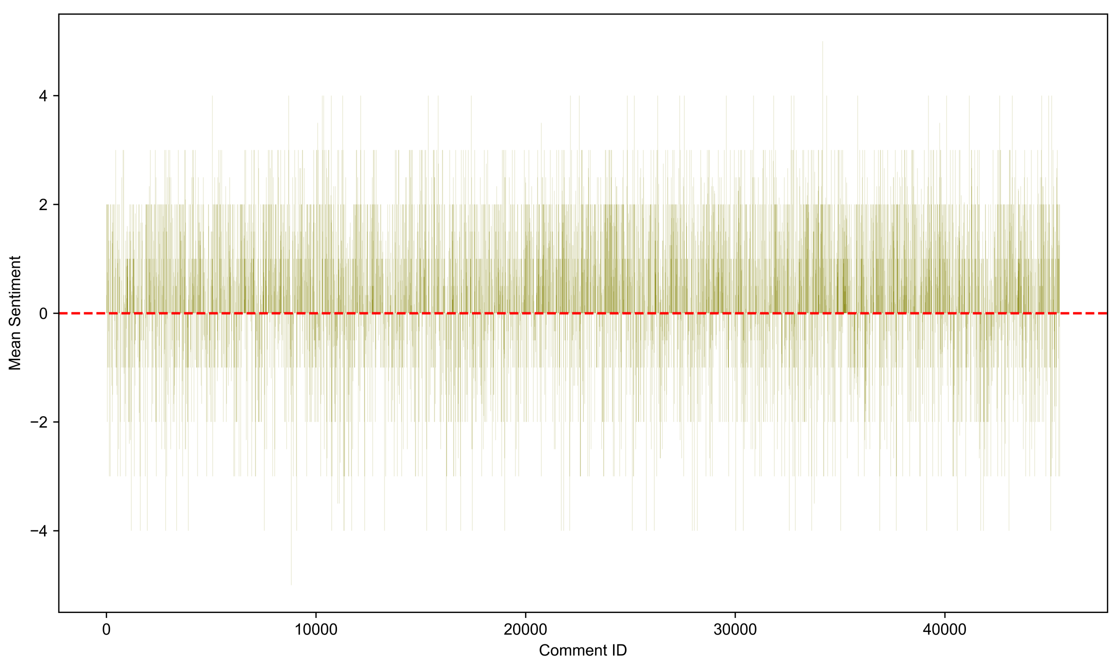

Figure 4.6: Average Sentiment per **Comments**

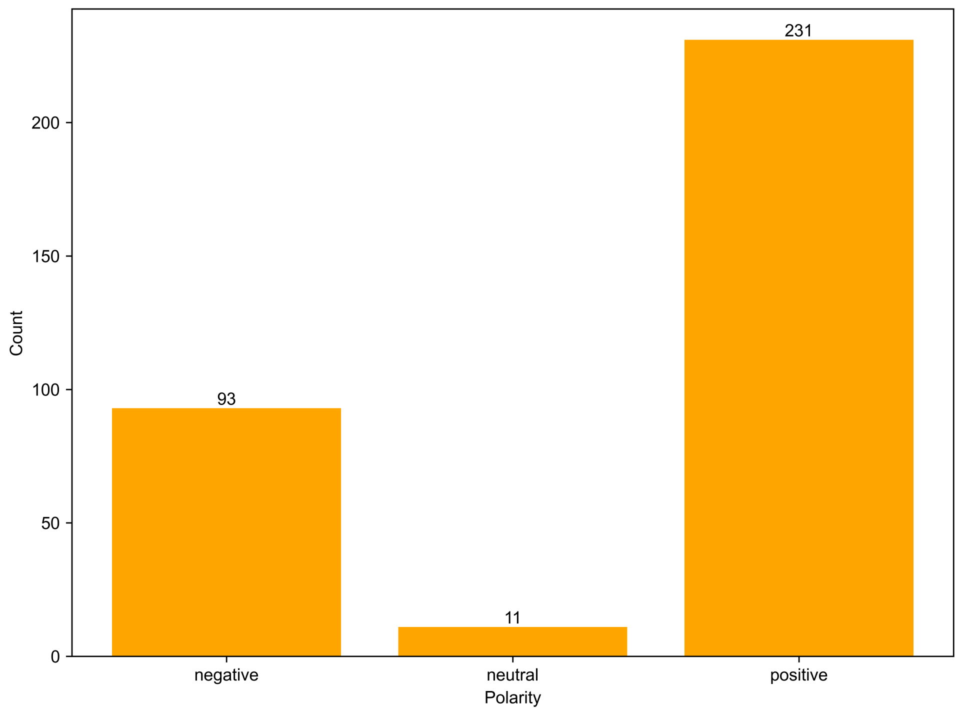

Figure 4.7: Proportion of **Threads** by Polarity

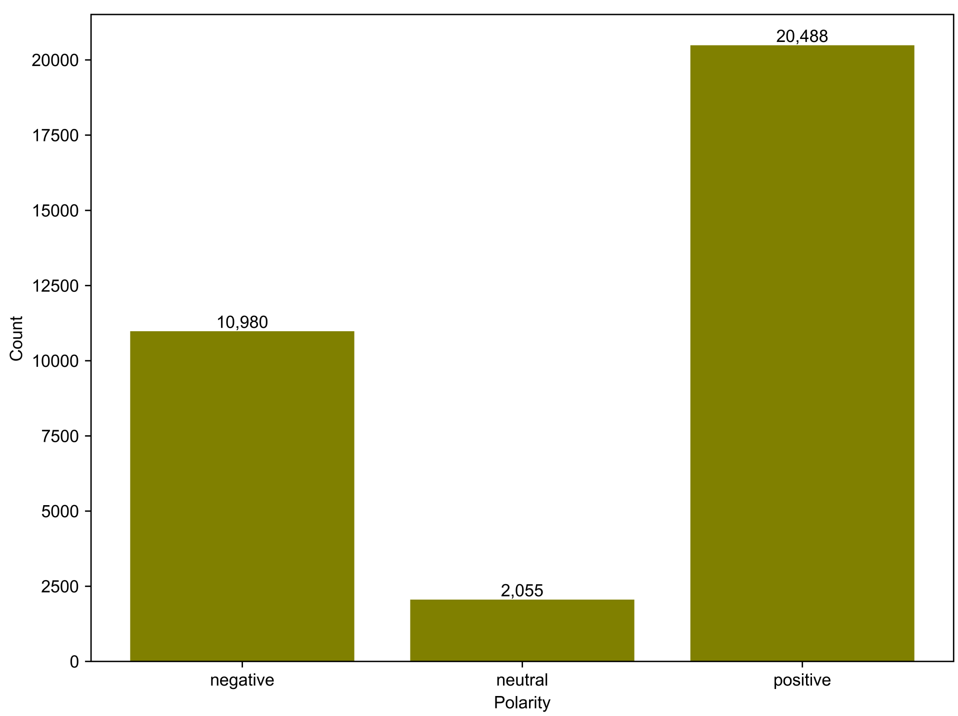

Figure 4.8: Proportion of **Comments** by Polarity

4.4 Topic Modeling

Using the tokenized words from **Section 4.2**, common topics from the threads and comments discussions can be inferred using the `LDA()` function in `R`. The function generates per-topic-per-word probabilities called the beta (β) parameter, as shown in the LDA process in **Figure 3.3**, which is used to infer common topic(s) on the dataset. Figures 4.9 and 4.10 illustrate the corresponding β values of the top 15 words. Topics can be inferred from the threads dataset (**Figure 4.9**) as follows:

- Topic T1 - discussion on Canada's EV charging infrastructure and charging times
- Topic T2 – discussion on EV ownership in Canada and its incentives
- Topic T3 – discussion on some EV model issues available in Canada

Lastly, comments dataset common topics were also inferred using **Figure 4.10**, and each topic was as follows:

- Topic C1 – discussion on ownership of different EV models in Canada
- Topic C2 – discussion on EV ownership in Canada and its incentives
- Topic C3 – discussion of Canada's EV charging infrastructure, charging time, and price of owning

As observed, both threads and comment topics are focused on the infrastructure, ownership, efficiency, and cost-effectiveness of EVs in Canada. It can also be noted that some of the topics on both data sets are the same (Topic T1 and Topic C3, Topic T2 and C2). This information can be helpful for policymakers and manufacturers to improve existing and future infrastructure as well as improve EV models.

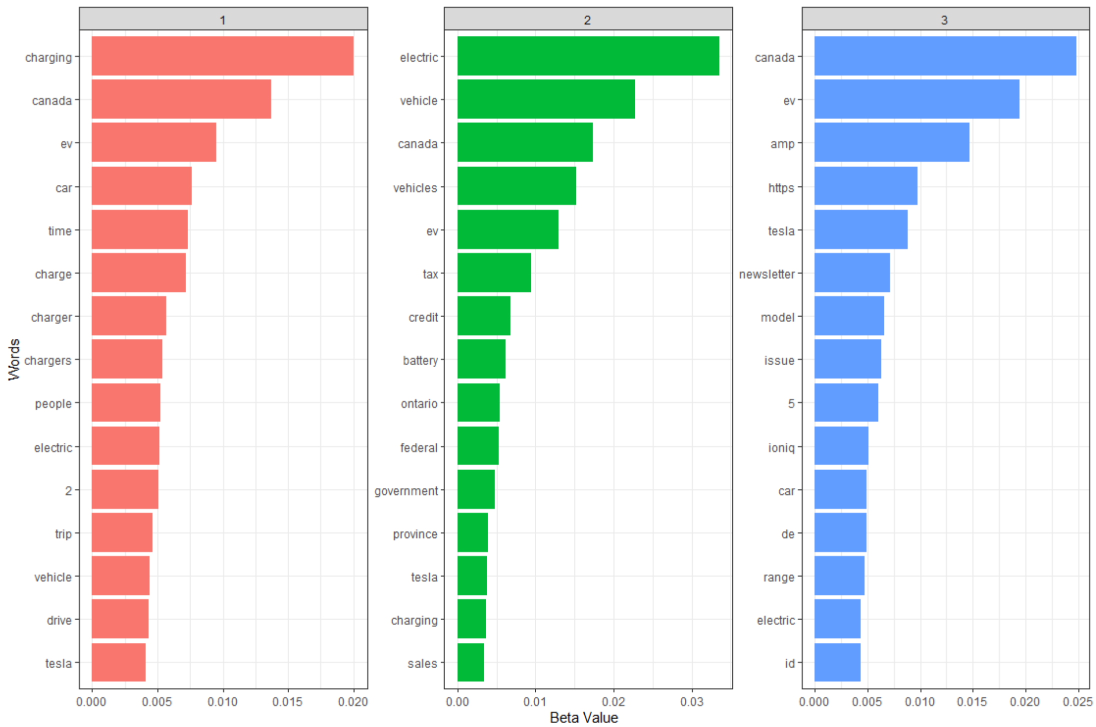

Figure 4.9: Per Topic Per Word Probabilities - Threads Data

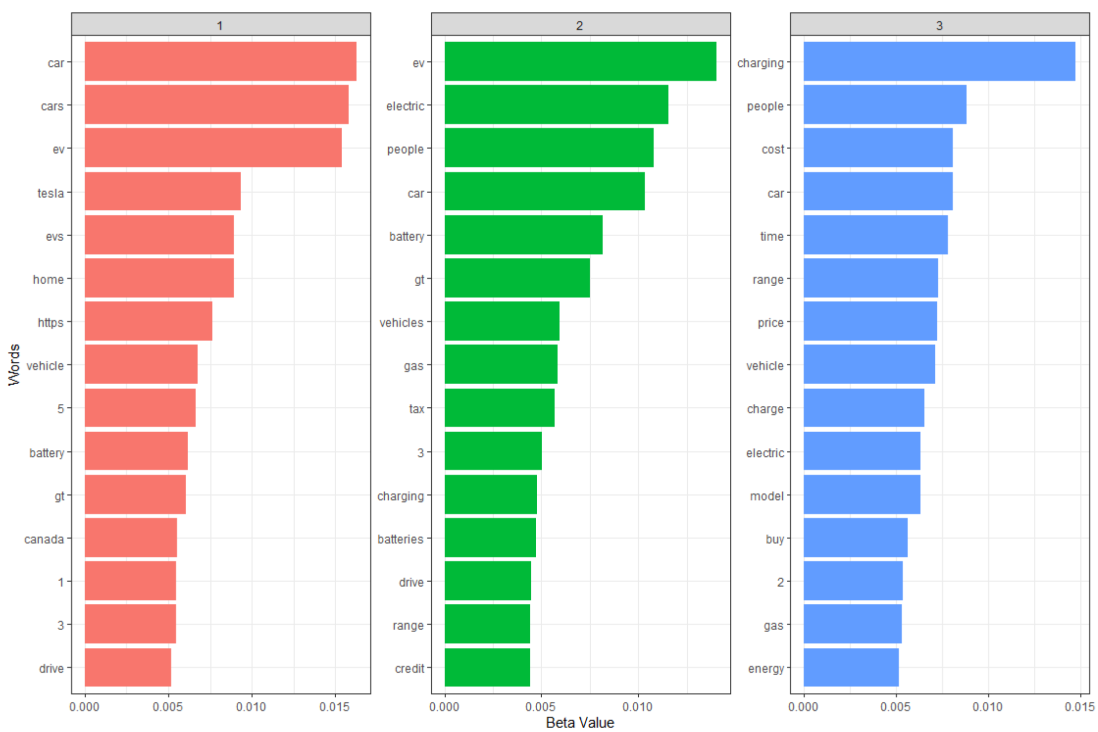

Figure 4.10: Per Topic Per Word Probabilities - Comments Data

Chapter 5

Conclusion and Future Work

The objective of this study was to use Reddit's thread and comment data to perform basic text-mining techniques in unveiling sentiments regarding electric vehicles (EVs) in Canada and to discover common topics discussed on the social media platform. Analysis showed that since 2015, there has been an increase in discussions about EVs on Reddit and EV registrations in Canada. This trend is forecasted to continue to grow in the coming years. Sentiment analysis was also conducted to determine the average sentiment of each thread and comment, and it was found that the majority of the discussions online were **positive**; however, some threads/comments also showed negative sentiments. Lastly, topic modeling was also conducted using the Latent Dirichlet Allocation (LDA) method to generate three (3) topics based on the tokenized words. The common themes on these topics were the Canadian EV infrastructure, charging facilities, vehicle efficiency and affordability, vehicle model issues, and incentives. This study will hopefully be helpful to vehicle manufacturers and policymakers to consider what has been said on social media sites regarding EVs to improve existing practices and plan for the future of sustainable mobility. The analysis conducted in this study has some limitations. For the sentiment analysis, the words are only compared to a lexicon dictionary and do not account for the usage of a word in a sentence, which can skew the overall sentiment of a thread/comment. Future studies should consider a more robust classification algorithm accounting for word usage or using advanced natural language processing (NLP) techniques. Another limitation is that the dataset used was not geotagged; therefore, it was not verified whether the threads/comments originated from Canada. Future studies must consider having geolocation information on the data to confirm thread/comment origin.

References

[1] Suresha, H. and Tiwari, K. (2021). *Topic modeling and sentiment analysis of electric vehicles of twitter data.* Asian Journal of Research in Computer Science, 13-29.

[2] Raamkumar, A., Tan, S., & Wee, H. (2020). *Measuring the outreach efforts of public health authorities and the public response on facebook during the covid-19 pandemic in early 2020: cross-country comparison.* Journal of Medical Internet Research, 22(5), e19334.

[3] Dubois, E., Gruzd, A., & Jacobson, J. (2023). *Journalists' use of social media to infer public opinion: the citizens' perspective..*

[4] Zúñiga, H., Molyneux, L., & Zheng, P. (2014). *Social media, political expression, and political participation: panel analysis of lagged and concurrent relationships.* Journal of Communication, 64(4), 612-634.

[5] Zhou, Z., Cai, Y., Su, S., Tang, X., & Cao, Y. (2020). *Electric vehicles scale evolution model considering social attributes.* IEEE Access, 8, 168784-168792.

[6] Long, Z., Axsen, J., & Kormos, C. (2019). *Consumers continue to be confused about electric vehicles: comparing awareness among canadian new car buyers in 2013 and 2017.* Environmental Research Letters, 14(11), 114036.

[7] Barkenbus, J. (2020). *Prospects for electric vehicles.* Sustainability, 12(14), 5813.

[8] Catton, J., Walker, S., McInnis, P., Fowler, M., Fraser, R., Young, S., & Gaffney, B. (2019). *Design and analysis of the use of re-purposed electric vehicle batteries for stationary energy storage in Canada.* Batteries, 5(1), 14.

[9] Mierlo, J. and Maggetto, G. (2007). *Fuel cell or battery: electric cars are the future.* Fuel Cells, 7(2), 165-173.

[10] Nayak, S. and Bohre, A. (2022). *Status of electric vehicles charging methods.* International Journal of Engineering Science and Technology, 14(3), 132-143.

[11] Yu, T., Rita, P., Moro, S., & Oliveira, C. (2021). *Insights from sentiment analysis to leverage local tourism business in restaurants.* International Journal of Culture, Tourism and Hospitality Research, 16(1), 321-336.

[12] SS, R. A., M, P., & Kalinathan, L. (2023). *Image caption generation for blind users of social media websites.*

[13] He, W., Zha, S., & Ling, L. (2013). *Social media competitive analysis and text mining: a case study in the pizza industry.* International Journal of Information Management, 33(3), 464-472.

[14] Rahman, S., Jahan, N., Sadia, F., & Mahmud, I. (2023). *Social crisis detection using twitter based text mining-a machine learning approach.* Bulletin of Electrical Engineering and Informatics, 12(2), 1069-1077.

[15] D, R. and Babu, J. (2017). *Public opinion mining on social media.* Ijarcce, 6(3), 934-937.

[16] Xie, R., Chu, S. K. W., Chiu, D. K., & Wang, Y. (2021). *Exploring public response to covid-19 on weibo with lda topic modeling and sentiment analysis.* Data and Information Management, 5(1), 86-99.

[17] Arifin, R. and Purnama, D. A. (2023). *Identifying customer preferences on two competitive startupproducts: an analysis of sentiment expressions and textmining from twitter data.* Jurnal Infotel, 15(1), 66-74.

[18] Budi, T. S. W., Wen, S. C., & Ratnawati, K. (2017). *Using social media mining to understand public opinion towards destination image.* Asia Pacific Management and Business Application, 5(3), 137-160.

[19] Ordenes, F. V. and Zhang, S. (2019). *From words to pixels: text and image mining methods for service research.* Journal of Service Management, 30(5), 593-620.

[20] Sick, N., Preschitschek, N., Bröring, S., & Leker, J. (2015). *Market convergence in the field of stationary energy storage systems.*

[21] Hesse, H. C., Schimpe, M., Kucevic, D., & Jossen, A. (2017). *Lithium-ion battery storage for the grid—a review of stationary battery storage system design tailored for applications in modern power grids.* Energies, 10(12), 2107.

[22] Thanh, N. T., Naumann, M., Nam, T. C., & Jossen, A. (2018). *Techno-economic evaluation of energy storage systems built from ev batteries – prospective revenues in different stationary applications.* E3S Web of Conferences, 64, 03003.

[23] Hodges, J., Simonsen, M., & Ottwein, J. K. (2022). *Gifted education on reddit: a social media sentiment analysis.* Gifted Child Quarterly, 66(4), 296-315.

[24] Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). *Studying reddit: a systematic overview of disciplines, approaches, methods, and ethics.* Social Media + Society, 7(2), 205630512110190.

[25] Gozzi, N., Tizzani, M., Starnini, M., Ciulla, F., Paolotti, D., Panisson, A., & Perra, N. (2020). *Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis.* Journal of medical Internet research, 22(10), e21597.

[26] Zeng, X., Li, M., El-Hady, D. A., Alshitari, W., Al-Bogami, A. S., Lü, J., & Amine, K. (2019). *Commercialization of lithium battery technologies for electric vehicles.* Advanced Energy Materials, 9(27).

[27] Slovic, A. D., Oliveira, M. A. d., Biehl, J., & Ribeiro, H. (2015). *How can urban policies improve air quality and help mitigate global climate change: a systematic mapping review.* Journal of Urban Health, 93(1), 73-95.

[28] Panchal, S., Dinçer, İ., Agelin-Chaab, M., Fraser, R., & Fowler, M. (2016). *Experimental temperature distributions in a prismatic lithium-ion battery at varying conditions.* International Communications in Heat and Mass Transfer, 71, 35-43.

[29] Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach.* O'Reilly Media.

[30] Pang, B. and Lee, L. (2008) *Opinion Mining and Sentiment Analysis.* Foundations and Trends® in Information Retrieval, 2, 1-135.

[31] Liu, B. (2012) *Sentiment Analysis and Opinion Mining.* Synthesis Lectures on Human Language Technologies, 5, 1-167.

[32] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent Dirichlet allocation.* Journal of Machine Learning Research, 3, 993-1022.

[33] Basso, Rafael, Kulcsár, Balázs and Sanchez-Diaz, Ivan, (2021), *Electric vehicle routing problem with machine learning for energy prediction*, Transportation Research Part B: Methodological, 145, issue C, p. 24-55.