SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2024
ORGANIZATION AND MANAGEMENT SERIES NO. 210

UNCOVERING TOPICS IN YOUTUBE COMMENTS ON ELECTRIC VEHICLES USING LATENT DIRICHLET ALLOCATION

Marcin WYSKWARSKI

Institute of Economy and Informatics, Faculty of Organization and Management, Silesian University of Technology; marcin.wyskwarski@polsl.pl, ORCID: 0000-0003-2004-330X

Purpose: The purpose of this study is to analyse YouTube comments related to electric vehicles (EVs) to uncover the underlying topics discussed by users. By applying Latent Dirichlet Allocation (LDA), the study aims to gain a deeper understanding of public opinion and the key themes driving conversations about EVs online.
Design/methodology/approach: This research utilized the “scrapetube” package to gather YouTube video links related to EVs. Comments were then downloaded using the “youtube-comment-downloader” package. After preprocessing the comments, including removing duplicates and irrelevant content, Latent Dirichlet Allocation (LDA) was applied to identify and analyze the topics discussed in these comments. The analysis focused on comments in Polish to ensure contextual accuracy and relevance.
Findings: The study identified a diverse range of topics, including technical aspects, consumer concerns, environmental impacts, market dynamics, and charging infrastructure. These topics provide insights into the public's perception towards EVs.
Research limitations/implications: The analysis was confined to comments in Polish, which may not capture the full spectrum of global perspectives. Future research could expand the scope to include comments in other languages and from different platforms to provide a more holistic view. Additionally, while LDA effectively identified key topics, some nuances may be lost in the process of topic modeling. Further qualitative analysis could complement the findings.
Practical implications: The insights gained from this study can help shape policies that address consumer concerns about EV affordability and infrastructure. Manufacturers can focus on areas that need improvement, such as enhancing battery capacity and increasing the efficiency and range of EVs. Marketing strategies can be refined to highlight the benefits of EVs in everyday use.
Originality/value: This study demonstrates the utility of LDA in uncovering the underlying topics in YouTube comments about electric vehicles. It provides a comprehensive overview of consumer interests and concerns, highlighting areas that require attention from policymakers, manufacturers, and researchers. The methodology used in this study can serve as a blueprint for further research into public opinion on other emerging technologies.
Keywords: Electric Vehicles (EVs), Latent Dirichlet Allocation (LDA), Topic Modeling, YouTube Comments.
Category of the paper: research paper.

1. Introduction

The European Union has implemented a variety of initiatives aimed at fostering the adoption of electric vehicles (EVs). These initiatives include providing financial incentives, developing infrastructure, and crafting strategic plans designed to encourage consumers to transition to EVs (Hawkins et al., 2013; Pasaoglu et al., 2012; Sierzchula et al., 2014). Studies indicate that the widespread adoption of EVs is challenging to accomplish without government subsidies. Rahmani & Loureiro highlight the necessity of such financial support (Rahmani, Loureiro, 2018). Additionally, incentives are critical during the initial phase of introducing new innovations, as evidenced by the work of (Davies et al., 2016; Figenbaum et al., 2015; Trip et al., 2012).

EVs have the potential to significantly lower greenhouse gas (GHG) emissions and enhance urban air quality. This improvement in air quality directly benefits public health, as EVs produce only natural by-products instead of harmful exhaust fumes. By reducing pollutants such as nitrogen oxides and particulate matter, EVs contribute to cleaner, healthier environments in cities (Hannan et al., 2014; Ma et al., 2012; Van Mierlo et al., 2006). However, from the perspective of potential buyers, environmental benefits may hold less significance, as society inherently tends to be driven by individual and hedonistic motivations (Buenstorf, Cordes, 2008). As a result, purchasing decisions are often based more on personal benefits and convenience rather than environmental impact.

EVs are becoming a significant part of the automotive industry, with projections indicating they will make up over 30% of the light-duty vehicle market in the United States by 2030 (Wolinetz, Axsen, 2017). Given this growth, understanding public perception and sentiment towards EVs is crucial for both policymakers and manufacturers. YouTube as a social media platform is one of the main places where users share their opinions and experiences about various products, including electric vehicles. Its global reach and diversity of content make it a valuable source of data for analysing public opinion (Aydın, Yılmaz, 2021; Muhammad et al., 2019; Radescu, Muraru, 2019).

The field of NLP-based research on electric vehicles (EVs) is relatively new, but there have been efforts to leverage topic analysis, topic modeling, and sentiment analysis for classifying user reviews. For instance, Ha et al. conducted an analysis focusing on user experiences with electric vehicles, employing language transformer models and supervised topic classification. Their research primarily concentrated on consumer reviews of EV charging stations, aiming to extract insights into user sentiments and recurring themes within these reviews (Ha et al., 2021). Jiang et al. provided a broader comparison between consumer and media sentiment towards electric vehicles, aiming to identify differences and similarities in perceptions and discussions surrounding EV technology (Jiang, Everts, 2021). Asensio et al. utilized a CNN classifier based on the word2vec model to classify sentiment. Their research revealed that positive and negative sentiments in user reviews are distributed roughly equally (Asensio et al., 2008).

Suresha and Tawari using topic modeling and Valence Aware Dictionary tools to analyse 45,000 tweets, discovered that "Tesla" was one of the most frequently used hashtags associated with EVs (Suresha, Tiwari, 2021). Bhatnagar and Choubey conducted a sentiment analysis of tweets using TF-IDF scores and a Naive Bayes classifier, finding that the hashtag "Tesla" exhibited a more positive sentiment compared to other manufacturers (Bhatnagar, Choubey, 2021). Carpenter used data collected from user discussion forums related to EVs to identify key sentiments and topics of interest, finding that "range anxiety" and "price" were two of the most common barriers to EV adoption (Carpenter, 2015).

This article analyses YouTube comments on EV videos using Latent Dirichlet Allocation (LDA) to uncover the underlying topics discussed by users. By examining these comments, the goal is to gain a deeper understanding of public opinion and the key themes that drive conversations about EVs online. This paper is organized as follows: section 2 describes the research methodology, section 3 presents the results, followed by a discussion in section 4, and the conclusion in section 5.

2. Research methodology

Using the "scrapetube" package (Twersky, n.d.), a list of 37,317 YouTube video links related to electric vehicles (EVs) was obtained. To conduct video searches, the author employed 74 phrases associated with EVs. The phrases included the brand name and the model name of the car. These phrases were established based on information obtained from one of the online platforms dedicated to EVs (https://motorvolt.pl/). Subsequently, videos that did not contain the search phrase in the title and were not created by a "Polish" channel were excluded. This process yielded a list of 1152 videos.

In the next step, comments posted by users under each of these 1152 videos were downloaded on May 24, 2024, using the "youtube-comment-downloader" package. This package provides a straightforward script for downloading YouTube comments without the need to utilise the YouTube API (*Youtube-Comment-Downloader*, n.d.). A total of 128,851 comments were downloaded.

The next phase involved removing:

- comments written in languages other than Polish,
- comments that were identical to others posted by the same user (often advertising services, products, or job opportunities from a company); the content was treated as a character string and compared using the "==" operator,
- comments of zero length.

The content of the comments was then pre-processed. This process followed standard steps for text mining (Gładysz, 2012; Turban Efraim et al., 2014). Uppercase letters were replaced with lowercase letters. URLs, hashtags, emojis, user names, words not considered useful (such as stop words, conjunctions, or prepositions), and all characters except letters were removed from the content. The word count of each cleaned comment was then verified, and comments containing fewer than one word were excluded. After this process, it was found that some videos had no comments or the comments became empty after cleaning. Ultimately, 125,680 comments remained. These comments were associated with one of 992 videos.

The next step was to reduce words to their base forms (e.g., verbs to infinitives, nouns to nominative forms, etc.). This is particularly important for the Polish language, which is characterized by extensive inflection. For this purpose, the services developed by CLARIN-PL were used (Branco et al., 2023; Fišer et al., 2018; Fišer, Witt, 2022; Janz et al., n.d.). The processing pipeline included the following services:

- Any2txt: It converts text files (e.g., doc, docx, xlsx) into plain text.
- Speller2: It checks the spelling of the text, utilizing a tool known as Autocorrect (https://languagetool.org/pl/) for this purpose.
- Wcrft2: A basic morpho-syntactic tagger for the Polish language.
- WSD: A service designed for word sense disambiguation, specifically tailored for Polish texts. It employs plWordNet as a source of potential meanings, which organizes lexical units into synsets connected through lexico-semantic relationships. Each lexical unit encapsulates a lexical meaning and is defined by three elements: a lemma, a part of speech, and a sense identifier (Janz et al., 2017).

In the final part of the analysis, Latent Dirichlet Allocation (LDA) was employed. LDA is described as a generative probabilistic model for collections of discrete data, such as text corpora (Blei et al., 2003). It is a widely used algorithm for Topic Modeling. More on topic modeling can be found in (Abd-Alrazaq et al., 2020; Garcia, Berton, 2021; Mottaghinia et al., 2020). This algorithm assumes that each document is represented by a mixture of topics, and each topic is represented by a mixture of words. The authors used the 'topicmodels' package for R (Hornik, Grün, 2011) to generate abstract, hidden topics describing comments related EVs. The probability of each word corresponding to topic was estimated using the Gibbs Sampling method (Anupriya, Karpagavalli, 2015).

One of the challenges when using the LDA algorithm is the need to choose the number of topics (k) before running the algorithm. There are several metrics for estimating k for a series of fitted LDA models. All of these methods require fitting the LDA model multiple times to the same dataset with various candidate values of k. Four metrics were used to determine k:

- Arun2010: This metric measures the divergence between the distribution of topics in documents and the distribution of words in topics. A lower value of this metric indicates a more optimal model with a better fit (Arun et al., 2010).

- CaoJuan2009: This metric calculates the density of word co-occurrences within topics. Lower values of this metric indicate higher quality and more distinct topics (Cao et al., 2009).
- Deveaud2014: This metric is based on the normalized pointwise mutual information (NPMI) of word pairs within topics. Higher values suggest more coherent topics that are easier to interpret (Deveaud et al., 2014).
- Griffiths2004: This metric evaluates the likelihood of the data under the model, with higher values indicating that the model better explains the observed data (Griffiths, Steyvers, 2004; Ponweiser, 2012).

3. Results

To determine the optimal number of topics for the Latent Dirichlet Allocation (LDA) models, the extrema were analyzed: maximum values in the case of Deveaud2014 and Griffiths2004, and minimum values in the case of Arun2010 and CaoJuan2009. From Figure 1, it can be observed that the optimal number of topics according to the applied metrics is 22 (Deveaud, 2014), 98 (Griffiths, 2004), 100 (Arun, 2010), and 90 (CaoJuan, 2009). Therefore, it can be assumed that the optimal number of k lies within the range of 22 to 100.

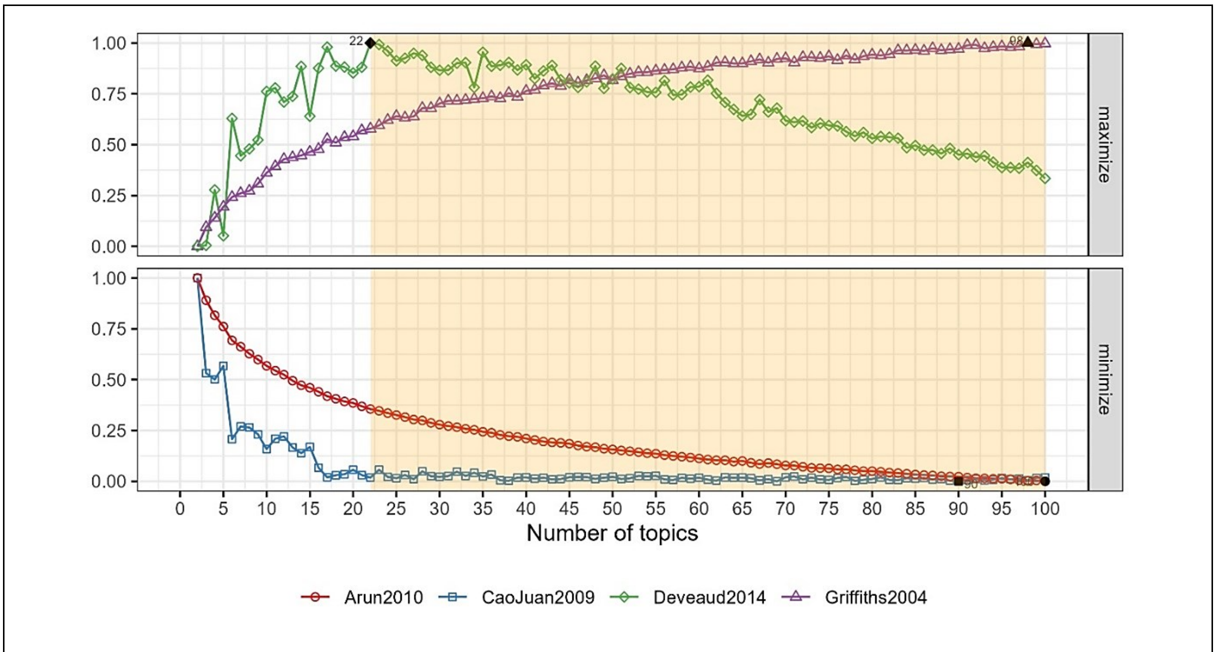

Figure 1. Metrics of LDA model quality for different numbers of topics.

Sources: original research.

Based on these values, LDA models were trained with the number of topics (k) set sequentially from 22 to 100. In the following step, the author conducted a manual evaluation of each generated model. The evaluation involved attempting to interpret the generated topics.

The identified topics were presented as word clouds. Each cloud contained the 15 most probable words assigned to the topic. Each cloud was generated with the same settings, and their images were not cropped. The larger the word, the higher the probability of the word occurring in the given topic. If the author was unable to interpret a given topic and assign a label (e.g., "Car model comparison"), the model was rejected. The model with the number of topics set to 22 was positively evaluated. Only in this case was the author able to interpret all identified topics and assign a single label to each topic, allowing for clear interpretation and presentation of the results.

Figure 2. LDA topics displayed as word clouds with labels – part 1.

Source: original research.

The topics identified by LDA are presented as word clouds in Figures 2 and 3. The assigned labels for these topics, which were created by the author, are placed below the word clouds. The clouds consist of words with the highest probabilities, representing the most significant terms for each topic according to the LDA model. These words are in their original form, i.e., in Polish, reduced to their base forms. It was decided not to translate them into English due to the fact that some words might have different meanings depending on the context.

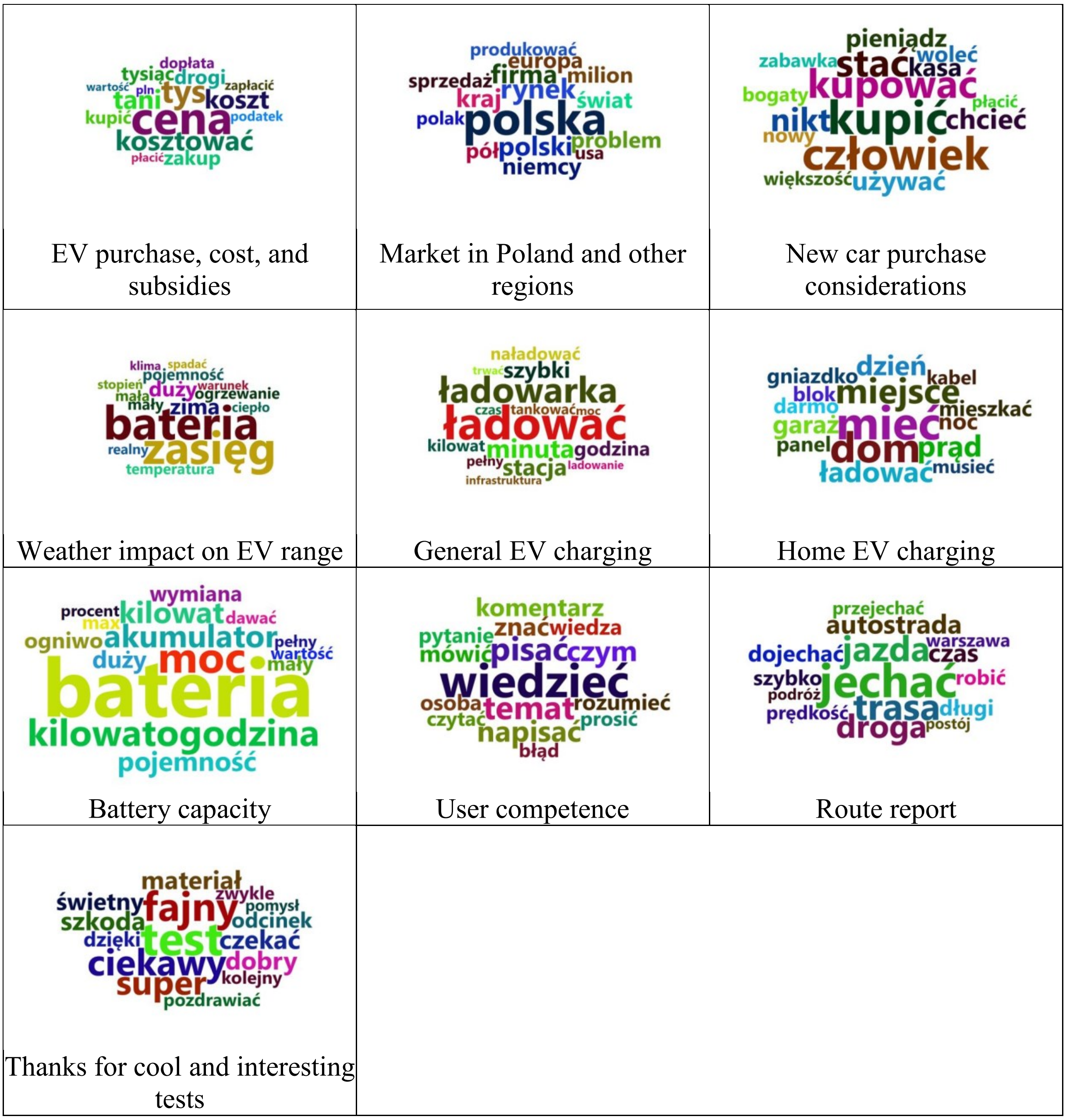

Figure 3. LDA topics displayed as word clouds with labels – part 2.

Source: original research.

The labels of the identified topics, along with short descriptions and interpretations are as follows:

- Car model comparison:
  - Description: Discussions comparing different car models.
  - Interpretation: Users frequently compare different electric car models to evaluate their advantages and disadvantages. Such comparisons help potential buyers make informed purchasing decisions.
- Car quality:
  - Description: Conversations about the build and quality of cars, especially premium brands.
  - Interpretation: The quality of build, especially for premium brands, is a significant topic of discussion. Users pay attention to the materials used in production and the overall durability of the vehicles.
- Trim levels:
  - Description: Dialogues about different equipment versions and specifications.
  - Interpretation: Equipment versions and technical specifications are often discussed, indicating that buyers are attentive to the various options available in each model.
- Car appearance:
  - Description: Opinions on the design and aesthetic appeal of cars.
  - Interpretation: The appearance of cars, including their design and aesthetics, is an important factor for consumers. Discussions on this topic can influence brand perception and attractiveness.
- EV as a daily driver:
  - Description: Considerations about using EVs for everyday commuting.
  - Interpretation: Users analyze whether electric vehicles are practical for everyday use, including range, charging convenience, and overall functionality in daily life.
- Powertrain specifications:
  - Description: Technical details about engines, gearboxes, and drivetrains.
  - Interpretation: Technical specifications of powertrains, including engines and gearboxes, are crucial for understanding the performance and reliability of vehicles.
- Fuel types:
  - Description: Discussions on various fuel types including LPG, gasoline, diesel, electric, and hybrid.
  - Interpretation: Consumers compare different fuel types, analyzing the benefits and drawbacks of each, which can influence their decisions about transitioning to electric vehicles.

- Car usability:
    - Description: Debates on the ease of use and practicality of cars.
    - Interpretation: The ease of use and practicality of cars are key topics that can affect user satisfaction and purchasing decisions.
- Car maintenance:
    - Description: Issues related to the maintenance and repair of cars.
    - Interpretation: Maintenance and repair issues are important for owners, affecting operating costs and overall vehicle satisfaction.
- EV viability and future of internal combustion engine:
    - Description: Debates on the practicality of EVs and the future of traditional internal combustion engines.
    - Interpretation: Discussions about the practicality of electric vehicles and the future of internal combustion engines reflect changing trends in the automotive industry and consumer expectations.
- Eco and non-eco energy sources:
    - Description: Comparisons between environmentally friendly and non-environmentally friendly energy sources.
    - Interpretation: Comparisons of environmentally friendly and non-environmentally friendly energy sources highlight the importance of sustainable development and the impact of energy choices on the environment.
- Energy consumption at high speed:
    - Description: Concerns about energy efficiency and consumption at high speeds.
    - Interpretation: The energy efficiency of vehicles at high speeds is a significant topic, influencing the range and performance of electric vehicles.
- EV purchase, cost, and subsidies:
    - Description: Discussions on the financial aspects of buying EVs, including costs and available subsidies.
    - Interpretation: Purchase costs and available subsidies are key factors affecting consumers' decisions to buy electric vehicles.
- Market in Poland and other regions:
    - Description: Insights into the EV market dynamics in Poland and other regions.
    - Interpretation: Analysis of the electric vehicle market in Poland and other regions helps understand the differences in technology adoption and the impact of local policies on the market.
- New car purchase considerations:
    - Description: Factors influencing the decision to buy a new car.
    - Interpretation: Factors influencing the decision to buy a new car include both financial and practical aspects, helping consumers make informed choices.

- Weather impact on EV range:
  - Description: The effect of weather conditions on the driving range of EVs.
  - Interpretation: The impact of weather conditions on the range of electric vehicles is a crucial issue that can affect practicality and user satisfaction. This is particularly significant in extreme weather conditions where energy consumption increases due to the need for heating in cold weather or cooling in hot weather. Such factors can significantly influence the driving range and overall efficiency of electric vehicles, making it a key concern for potential buyers and current users.
- General EV charging:
  - Description: General issues and experiences related to EV charging.
  - General issues related to electric vehicle charging, such as the availability of charging stations and charging time, are critical for EV users.
- Home EV charging:
  - Description: Specific discussions about the practicality and setup of home charging for EVs.
  - Charging electric vehicles at home is an important aspect, influencing convenience and operating costs. Homeowners have the advantage of access to their own power source, making it easier to charge their vehicles overnight. In contrast, individuals living in apartments or multi-unit dwellings may face challenges finding accessible and reliable charging stations, which can affect their ability to conveniently charge their vehicles.
- Battery capacity:
  - Description: Concerns and discussions about the battery capacity of EVs.
  - Interpretation: Battery capacity is a key factor affecting the range and performance of electric vehicles, which is frequently discussed by users.
- User competence:
  - Description: Competence and experience of users in discussing and handling EV-related topics.
  - Interpretation: The competence and experience of users in discussing and handling electric vehicle-related topics can affect the quality and value of these discussions. Often, YouTube viewers accuse others of lacking competence on the subject, leading to unnecessary conflicts and disputes. This dynamic can detract from constructive conversations and diminish the overall value of the discourse.
- Route report:
  - Description: Reports from journeys undertaken using EVs.
  - Interpretation: Reports from specific journeys undertaken with electric vehicles provide practical information on the real-world use of EVs. They cover topics such as charging the vehicle on the road, the time needed for charging, and using chargers

from specific companies. These firsthand accounts offer valuable insights into the challenges and benefits of using EVs for long-distance travel, enhancing the overall understanding of the practicality and convenience of electric vehicles.

- Thanks for cool and interesting tests:
  - Description: Appreciation for informative and engaging test videos on EVs.
  - Interpretation**:** Appreciation for interesting and valuable vehicle tests shows the important role of content creators in educating and informing consumers.

Each of these topics reflects different aspects of consumer interest and concern, providing a comprehensive overview of the public discourse on electric vehicles. By examining these discussions, we can better understand the key issues that matter to consumers, from technical specifications and practical usability to environmental impact and economic considerations.

4. Discussion

The analysis of YouTube comments on electric vehicles (EVs) using Latent Dirichlet Allocation (LDA) has yielded significant insights into the public discourse surrounding EVs. This study identified a range of topics discussed by users, offering a comprehensive overview of consumer interests and concerns. These findings can inform policymakers, manufacturers, and researchers about the key themes that drive conversations about EVs online.

Key findings are as follows:

- Diverse range of topics: The identified topics cover a broad spectrum, including technical aspects, consumer concerns, environmental impacts, market dynamics, and charging infrastructure. This diversity reflects the multifaceted nature of public discussions on EVs. Previous studies have also highlighted the wide range of consumer interests in EVs, such as user experiences and technical specifications (Ha et al., 2021; Jiang, Everts, 2021).
- Consumer concerns: The analysis revealed common concerns among consumers, such as the viability of EVs, the future of internal combustion engines, and financial aspects related to purchasing EVs. These findings are consistent with other research showing that range anxiety and cost are significant barriers to EV adoption (Bhatnagar, Choubey, 2021; Carpenter, 2015).
- Environmental impact: The prominence of topics related to eco and non-eco energy sources indicates significant interest in the environmental implications of different energy types. This highlights the importance of sustainability in consumer decision-making, aligning with research that emphasizes the environmental benefits of EVs (Hannan et al., 2014; Ma et al., 2012; Van Mierlo et al., 2006).

- Market dynamics: Discussions about the EV market in Poland and other regions, as well as considerations for purchasing new cars, point to the economic and regional factors influencing EV adoption. Studies have shown that regional policies and market conditions significantly impact EV adoption rates (Sierzchula et al., 2014; Wolinetz, Axsen, 2017).
- Charging Infrastructure: The focus on EV charging, both general and home-specific, suggests that charging infrastructure remains a critical area of concern for potential EV buyers. This finding is supported by research indicating that the availability and convenience of charging stations are crucial for EV adoption (Ha et al., 2021; Suresha, Tiwari, 2021).

It should be noted that the topics identified by LDA did not solely pertain to electric vehicles but also encompassed other aspects of YouTube activity and the broader automotive industry. For example, some comments referred to the activities of content creators on YouTube, expressing gratitude for interesting and valuable vehicle tests (label: Thanks for cool and interesting tests). Other topics included discussions about various energy sources, ranging from fossil fuels to renewable energy sources. Additionally, there were conversations comparing different car models and deliberating the pros and cons of traditional combustion engines versus electric engines. These findings highlight the broader context in which discussions about EVs occur, reflecting a wide range of interests and concerns among users.

5. Conclusion

The study demonstrates the utility of LDA in uncovering the underlying topics in YouTube comments about electric vehicles. The identified topics offer a snapshot of consumer interests and concerns, highlighting areas that require attention from policymakers, manufacturers, and researchers. Addressing these concerns can contribute to broader adoption and acceptance of electric vehicles, ultimately supporting the transition to a more sustainable transportation system.

By leveraging the insights gained from this analysis, stakeholders can make informed decisions that better align with public sentiment and address the practical challenges associated with EV adoption. This approach not only enhances the understanding of consumer perspectives but also fosters a more consumer-centric development of EV technologies and policies.

The findings underscore the importance of addressing issues such as EV viability, cost, and charging infrastructure. Policymakers can utilize this information to design targeted incentives and infrastructure projects that more effectively address these consumer concerns. Manufacturers can focus on enhancing battery capacity, efficiency, and range to better meet consumer expectations and improve the overall appeal of EVs.

Future research should consider expanding the linguistic scope to include comments in other languages and from various platforms to provide a more comprehensive view of global perspectives on EVs. Additionally, further qualitative analysis could complement the quantitative findings, capturing nuances that may be lost in the process of topic modeling.

In summary, this research highlights the effectiveness of LDA in analysing large datasets of user-generated content to uncover key topics in discussions about EVs. The insights gained can inform future policies, manufacturing strategies, and research directions, contributing to the broader discourse on sustainable transportation and helping to accelerate the adoption of electric vehicles.

References

1. Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hai, M., Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: A surveillance study. *Journal of Medical Internet Research, 22(4), e19016*. https://doi.org/10.2196/19016
2. Anupriya, P., Karpagavalli, S. (2015). *LDA based topic modeling of journal abstracts.* International Conference on Advanced Computing and Communication Systems, 1-5. https://doi.org/10.1109/ICACCS.2015.7324058
3. Arun, R., Suresh, V., Veni Madhavan, C.E., Narasimha Murthy, M.N. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6118 LNAI(PART 1),* 391-402. https://doi.org/10.1007/978-3-642-13657-3_43
4. Asensio, O.I., Alvarez, K., Drorc, A., Hollauere, E.W.C. (2008). Evaluating popular sentiment of electric vehicle owners in the United States with real-time data from mobile platforms. *Annali Dell'Istituto Superiore Di Sanita, 44(2),* 135-144.
5. Aydın, E., Yılmaz, E. (2021). YouTube as a Source of Information on Echocardiography: Content and Quality Analysis. *Acta Cardiologica Sinica, 37(5),* 534-541. https://doi.org/10.6515/ACS.202109_37(5).20210514A
6. Bhatnagar, S., Choubey, N. (2021). Making sense of tweets using sentiment analysis on closely related topics. *Social Network Analysis and Mining, 11(1),* 1-11. https://doi.org/10.1007/S13278-021-00752-0/TABLES/2
7. Blei, D., Ng, A., Jordan, M. (2003). Latent Dirichlet Allocation Michael I. Jordan. *Journal of Machine Learning Research, 3*, 993-1022.
8. Branco, A., Eskevich, M., Frontini, F., Hajič, J., Hinrichs, E., Jong, F. de, Kamocki, P., König, A., Lindén, K., Navarretta, C., Piasecki, M., Piperidis, S., Pitkänen, O., Simov, K., Skadiņa, I., Trippel, T., Witt, A., Zinn, C. (2023). The CLARIN infrastructure as

an interoperable language technology platform for SSH and beyond. *Language Resources and Evaluation*, 1-32. https://doi.org/10.1007/S10579-023-09658-Z/FIGURES/1

9. Buenstorf, G., Cordes, C. (2008). Can sustainable consumption be learned? A model of cultural evolution. *Ecological Economics, 67(4),* 646-657. https://doi.org/10.1016/J.ECOLECON.2008.01.028
10. Cao, J., Xia, T., Li, J., Zhang, Y., Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing, 72(7-9),* 1775-1781. https://doi.org/10.1016/J.NEUCOM.2008.06.011
11. Carpenter, T. (2015). *Measuring and Mitigating Electric Vehicle Adoption Barriers.* University of Waterloo. https://uwspace.uwaterloo.ca/handle/10012/9096
12. Davies, H., Santos, G., Faye, I., Kroon, R., Weken, H. (2016). Establishing the Transferability of Best Practice in EV Policy across EU Borders. *Transportation Research Procedia, 14*, 2574-2583. https://doi.org/10.1016/J.TRPRO.2016.05.350
13. Deveaud, R., SanJuan, E., Bellot, P. (2014). Accurate and effective Latent Concept Modeling for ad hoc information retrieval. *Document Numerique, 17(1),* 61-84. https://doi.org/10.3166/DN.17.1.61-84
14. Figenbaum, E., Assum, T., Kolbenstvedt, M. (2015). Electromobility in Norway: Experiences and Opportunities. *Research in Transportation Economics, 50*, 29-38. https://doi.org/10.1016/J.RETREC.2015.06.004
15. Fišer, D., Lenardič, J., Erjavec, T. (2018). *CLARIN's Key Resource Families.* https://aclanthology.org/L18-1210
16. Fišer, D., Witt, A. (2022). *CLARIN: The infrastructure for language resources,* 1-810. https://doi.org/10.1515/9783110767377/EPUB
17. Garcia, K., Berton, L. (2021). Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing, 101,* 107057. https://doi.org/10.1016/j.asoc.2020.107057
18. Gładysz, A. (2012). Zastosowanie metod eksploracyjnej analizy tekstu w logistyce. *Logistyka, no. 3,* 643-651.
19. Griffiths, T.L., Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America, 101(SUPPL. 1),* 5228-5235. https://doi.org/10.1073/pnas.0307752101
20. Ha, S., Marchetto, D.J., Dharur, S., Asensio, O.I. (2021). Topic classification of electric vehicle consumer experiences with transformer-based deep learning. *Patterns, 2(2),* 100195. https://doi.org/10.1016/j.patter.2020.100195
21. Hannan, M.A., Azidin, F.A., Mohamed, A. (2014). Hybrid electric vehicles and their challenges: A review. *Renewable and Sustainable Energy Reviews, 29,* 135-150. https://doi.org/10.1016/J.RSER.2013.08.097

22. Hawkins, T.R., Singh, B., Majeau-Bettez, G., Strømman, A.H. (2013). Comparative Environmental Life Cycle Assessment of Conventional and Electric Vehicles. *Journal of Industrial Ecology, 17(1),* 53-64. https://doi.org/10.1111/J.1530-9290.2012.00532.X
23. Hornik, K., Grün, B. (2011). Topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software, 40*, 1-30. https://doi.org/10.18637/jss.v040.i13
24. Janz, A., Kocoń, J., Piasecki, M., Zaśko-Zielińska, M. (n.d.). *plWordNet as a Basis for Large Emotive Lexicons of Polish CLARIN-PL View project Liner2-A Customizable Framework for Automatic Text Annotation (NER, TimeX, Events) View project.* Retrieved from: https://www.researchgate.net/publication/322684200, June 5, 2022.
25. Janz, A., Kocoń, J., Piasecki, M., Zaśko-Zielińska, M. (2017). *plWordNet as a Basis for Large Emotive Lexicons of Polish*.
26. Jiang, X., Everts, J. (2021). *Making sense of electrical vehicle discussions using sentiment analysis on closely related news and user comments.* https://arxiv.org/abs/2112.12327v4
27. Ma, H., Balthasar, F., Tait, N., Riera-Palou, X., Harrison, A. (2012). A new comparison between the life cycle greenhouse gas emissions of battery electric vehicles and internal combustion vehicles. *Energy Policy, 44,* 160-173. https://doi.org/10.1016/J.ENPOL.2012.01.034
28. Mottaghinia, Z., Feizi-Derakhshi, M.R., Farzinvash, L., Salehpour, P. (2020). A review of approaches for topic detection in Twitter. *Journal of Experimental and Theoretical Artificial Intelligence*, 1-27. https://doi.org/10.1080/0952813X.2020.1785019
29. Muhammad, A.N., Bukhori, S., Pandunata, P. (2019). *Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes – Support Vector Machine (NBSVM) Classifier.* International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), 199-205. https://doi.org/10.1109/ICOMITEE.2019.8920923
30. Pasaoglu, G., Honselaar, M., Thiel, C. (2012). Potential vehicle fleet CO2 reductions and cost implications for various vehicle technology deployment scenarios in Europe. *Energy Policy, 40(1)*, 404-421. https://doi.org/10.1016/J.ENPOL.2011.10.025
31. Ponweiser, M. (2012). *Latent Dirichlet Allocation in R. May*, 2-21. https://research.wu.ac.at/en/publications/latent-dirichlet-allocation-in-r-3
32. Radescu, R., Muraru, V. (2019). Study Platform for Complex Data Analysis of Telecommunications and Social Network Applications Using Business Intelligence. In: I. Roceanu, D. Belgian, I.A. Stefan, A. Moldoveanu, S.T. Matu (Ed.), *New technologies and redesigning learning spaces, Vol. I* (pp. 358-365). Carol i Natl Defence Univ Publishing House. https://doi.org/10.12753/2066-026X-19-048
33. Rahmani, D., Loureiro, M.L. (2018). Why is the market for hybrid electric vehicles (HEVs) moving slowly? *PLOS ONE, 13(3),* e0193777. https://doi.org/10.1371/JOURNAL.PONE.0193777

34. Sierzchula, W., Bakker, S., Maat, K., Van Wee, B. (2014). The influence of financial incentives and other socio-economic factors on electric vehicle adoption. *Energy Policy, 68,* 183-194. https://doi.org/10.1016/J.ENPOL.2014.01.043
35. Suresha, H.P., Tiwari, K.K. (2021). Topic Modeling and Sentiment Analysis of Electric Vehicles of Twitter Data. *Asian Journal of Research in Computer Science, 12(2),* 13-29. https://doi.org/10.9734/AJRCOS/2021/V12I230278
36. Trip, J.J., Lima, J., Bakker, S. (2012). *Electric mobility policies in the North Sea Region countries. Report for E-Mobility NSR.* https://www.yumpu.com/en/document/view/7530898/electric-mobility-policies-in-the-north-sea-region-e-mobility-nsr
37. Turban Efraim, Sharda Ramesh, Delen Dursun (2014). *Business Intelligence and Analytics: Systems for Decision Support (10th ed.).* Pearson Education Limited.
38. Twersky, C. (n.d.). *scrapetube*. Retrieved from: https://pypi.org/project/scrapetube/, June 13, 2023.
39. Van Mierlo, J., Maggetto, G., Lataire, P. (2006). Which energy source for road transport in the future? A comparison of battery, hybrid and fuel cell vehicles. *Energy Conversion and Management, 47(17),* 2748-2760. https://doi.org/10.1016/J.ENCONMAN.2006.02.004
40. Wolinetz, M., Axsen, J. (2017). How policy can build the plug-in electric vehicle market: Insights from the REspondent-based Preference And Constraints (REPAC) model. *Technological Forecasting and Social Change, 117,* 238-250. https://doi.org/10.1016/j.techfore.2016.11.022
41. youtube-comment-downloader (n.d.). Retrieved from: https://pypi.org/project/youtube-comment-downloader/, June 13, 2023.