Analysis of Sentiment in YouTube Comments for 2024 Political Campaigns for Governor of Puerto Rico

Author: Carlos A. Oriol Rivera
Advisor: Dr. Jeffrey L. Duffany
Department of Computer Engineering

Abstract

The analysis of public sentiment is an invaluable tool in political campaigns, enabling candidates and their teams to assess voter attitudes and adjust campaign strategies accordingly. This project focuses on collecting and analyzing video comments on YouTube from the months leading up to an election to determine the public's sentiment toward different candidates for governor in Puerto Rico. To make this sentiment analysis, the Text Blob Python library was used due to its ease of use and robust functionality. The project incorporates both client-side and server-side components, integrating modern web development frameworks and libraries to provide good user experience. This investigation outlines the data collection methodology, system architecture, analysis process, discussion of results, and coding logic used to achieve the analysis.

Introduction

In today's political landscape, understanding public sentiment has become essential for creating effective campaign strategies. Social media platforms, such as YouTube, play a very important role in shaping voter opinions, providing valuable insights for campaign staff and managers. This project focuses on analyzing public sentiment by examining YouTube comments related to governor candidates in Puerto Rico during the three months leading up to the 2024 election. By analyzing and diving into these comments, the researcher aimed to uncover trends and patterns in voter sentiment, offering a unique perspective on how people viewed the candidates and their campaigns in each timeframe.

Using the TextBlob Python library, a reliable tool for processing textual data, the researcher conducted sentiment analysis to categorize comments into positive, negative, or neutral sentiments. The project combined modern client-side and server-side technologies to ensure a user-friendly system for managing and visualizing sentimental data. This approach streamlined the analysis process, making it both efficient and accurate for handling large datasets.

The investigation outlines the entire process, from data collection and preparation to the system's architecture and sentiment analysis methodology. By detailing the technologies and steps involved, including data cleaning and categorization, the project provides a comprehensive look at how public sentiment on digital platforms can influence political campaigns and their strategies going forward.

Methodology

Data Collection

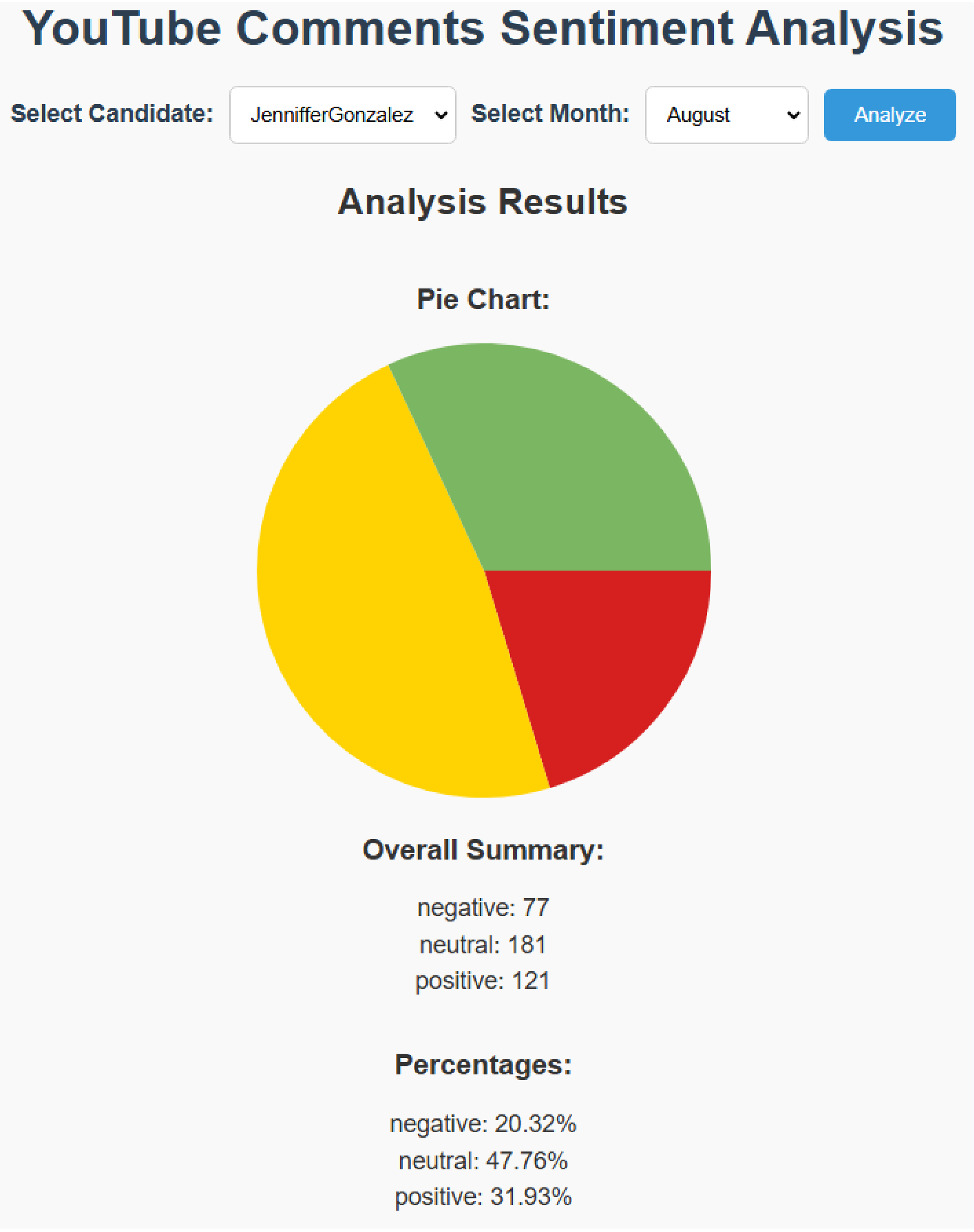
The data was collected through Google video searches. First the investigator searched for the candidate and added "site:youtube.com" at the end to only include YouTube videos. Then filtered the results by month (August, September, October 2024). Finally, extracted video IDs located in the URL of each video. The IDs were organized into CSV files categorized by candidate and month. This structured approach ensures accurate and efficient data analysis.

System Architecture

The client is a single-page application built using ReactJS [1]. It provides a user-friendly interface where users can select a candidate and a month from dropdown menus and initiate the analysis via a button. After clicking the Analyze button, it will deactivate and show a loading text. Finally, when the analysis ends, it will show the results below the input fields. The backend server is a REST API designed in the Python programming language [2] and implements a POST endpoint named 'analyze' that was implemented using the Flask framework [3]. When triggered, this endpoint processes the selected dataset, cleans the comments, and performs sentiment analysis using various libraries.

Logic

The first thing we do through the code is locate the YouTube video IDs desired located in the data collected CSV files. After that, the code loops through each if the ids in the selected file and connect to the google API [4] to extract the comments for each video and do the Sentiment Analysis on every one of them. To analyze the comments, the code must first translate the text to English. To do that the researcher wrote a function that detected the language of each comment and translated it, if necessary, before doing the sentiment analysis. After that, the investigator used the TextBlob library to do the sentiment analysis of the comment [5]. Finally, the researcher used the result of the sentiment analysis to structure the data for the front-end client to present it



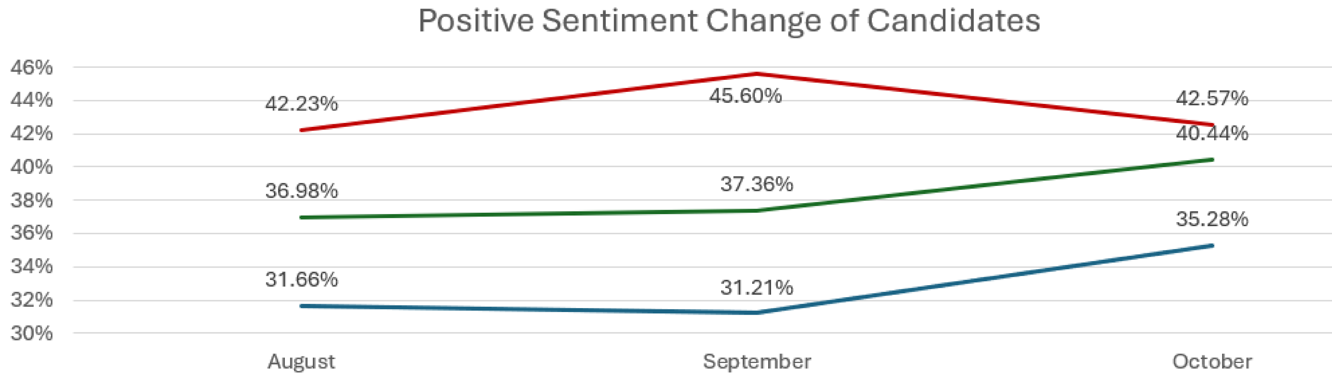
Results and Discussion

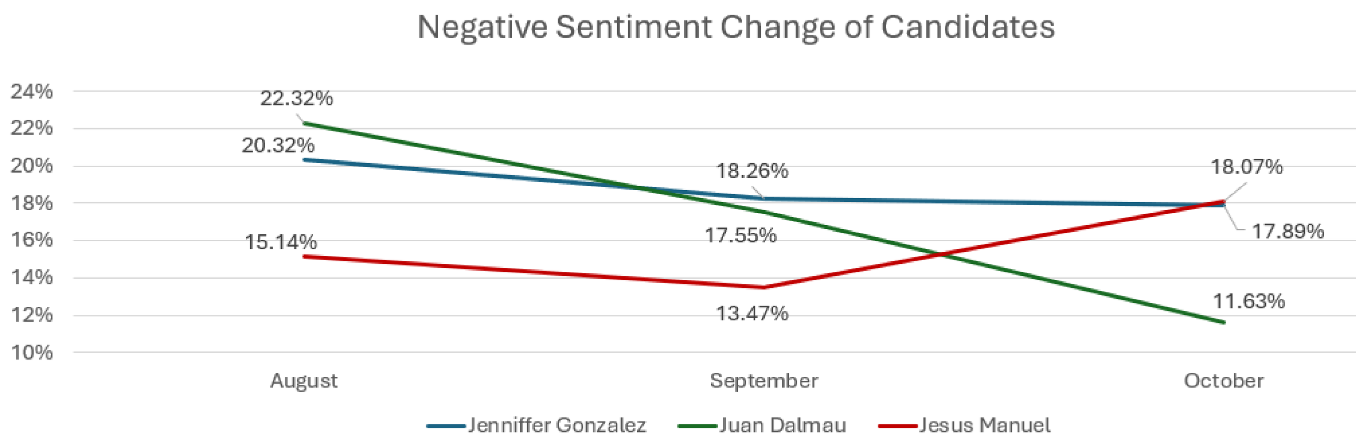
Jenniffer González started in August with a relatively balanced sentiment distribution, with 31.66% positive, 20.32% negative, and 48.02% neutral. In September, her positive sentiment dropped slightly to 31.21%, and her negative sentiment fell to 18.26%, while neutral sentiment rose to 50.53%, suggesting the public was becoming less polarized and more indifferent about her. By October, González regained positive sentiment, rising to 35.28%, while negative sentiment decreased further to 17.89%. Neutral sentiment fell slightly to 46.82%, indicating a growing favorable perception and a decline in criticism, as more people shifted away from neutrality.

Juan Dalmau showed consistent improvement across all months. In August, his sentiment was 36.98% positive, 22.32% negative, and 40.70% neutral. By September, his positive sentiment increased to 37.36%, while negative sentiment fell significantly to 17.55%, and neutral sentiment rose to 45.09%. This indicates a decline in criticism and growing appeal. In October, Dalmau's positive sentiment jumped further to 40.44%, with negative sentiment dropping to 11.63% and neutral sentiment rising to 47.92%. This steady increase in favorability and reduction in criticism highlights his growing momentum as a leading candidate.

Jesús Manuel Ortiz consistently maintained high levels of positive sentiment, starting in August with 42.23% positive, 15.14% negative, and 42.63% neutral. In September, his positive sentiment increased to 45.60%, with negative sentiment slightly declining to 14.47%, while neutral sentiment dropped to 40.93%. This indicates polarization, with more people shifting from neutral to positive perceptions. However, in October, Ortiz's positive sentiment fell slightly to 42.57%, and negative sentiment rose to 18.07%, while neutral sentiment remained stable around 39.36%. This suggests some emerging criticism or controversies.

Overall, Juan Dalmau emerged as the candidate with the most consistent improvement, particularly in reducing negative sentiment and increasing positive views, making him the most positively perceived candidate in October taking in consideration the data collected from YouTube comments. Jenniffer González showed signs of recovery after a very neutral phase, with a notable improvement in positive sentiment toward the end of the period. Jesús Manuel Ortiz, while maintaining strong overall favorability, faced a slight rise in negative sentiment in October, suggesting some potential issues at the end of his campaign leading to the election.





Conclusions and Future Work

This project shows how modern web development tools and powerful sentiment analysis libraries can work together to dig into the political sentiment around specific candidates. Using the TextBlob library along with other frameworks and technologies, the researcher was able to handle large datasets of comments, uncovering interesting patterns and trends in how people feel about candidates. Specifically, this project analyzed the sentiment trajectory for each candidate based on YouTube comments, which gave the researcher a good sense of how public opinion shifted over time on this platform.

There's room to make this tool even better. In the future, others, including the researchers, could expand the analysis to include more social media platforms like X (formerly known as Twitter) and Facebook. These platforms have a wider variety of conversations and interactions that could give different people a more accurate picture of voter sentiment. Of course, doing this would require more time, resources, and a bigger budget to access and process all that extra data. With these upgrades, this tool could become even more precise and provide deeper insights into how people feel about political candidates at a given timeframe across multiple platforms. However, it is important to recognize that sentiment analysis is not a predictor of election outcomes. In the case of this project, the candidate with the most positive sentiment did not win the election. While sentiment analysis offers valuable insights into public perception and shifts in opinion, it does not account for other important factors such as voter turnout and the influence of offline events. Sentiment analysis should be viewed as a complementary tool that campaigns can use to analyze public reception, identify key issues, and refine messaging and campaign strategies. By using this technology, political campaigns can enhance their engagement efforts on online platforms and better understand how voters are reacting to their candidates in real-time.

Acknowledgements

I would like to express my sincere gratitude to Dr. Jeffrey L. Duffany for his guidance, feedback, and support throughout the project. I'm also thankful to the Polytechnic University of Puerto Rico for providing the resources and knowledge necessary for this project.

References

[1] Meta Platforms, Inc. (2024, 12, 05). React v19. [Online]. Available: https://react.dev/blog/2024/12/05/react-19. [Accessed: January 6, 2025].
[2] Python Software Foundation. (2025, 12, 03). Python 3.13.1 documentation. Python: [Online]. Available: https://docs.python.org/3/. [Accessed: January 6, 2025].
[3] Pallets. (2024, 11, 13). Flask Framework Documentation: Pallets. [Online]. Available: https://flask.palletsprojects.com/en/stable/. [Accessed: January 6, 2025].
[4] Google for Developers. (2022, 07, 21). YouTube Data API v3: Google for Developers: [Online]. Available: https://developers.google.com/youtube/v3/docs. [Accessed: January 6, 2025].
[5] Loria, S. (2024, 02, 15). TextBlob: Simplified Text Processing. TextBlob: [Online]. Available: https://textblob.readthedocs.io/en/dev/. [Accessed: January 6, 2025].