Inho Ryu

Supercomputing Challenge

December 20, 2024

Modeling the Spread of COVID-19 through YouTube Data

About my project

This project addresses the challenge of tracking and predicting the spread of global diseases, specifically focusing on COVID-19, by utilizing social media data as a novel and accessible resource. COVID-19, one of the most significant global events in recent history, disrupted lives, economies, and healthcare systems worldwide. Traditional epidemiological models often rely on health records and testing data, which can be delayed or limited in coverage. Social media platforms, on the other hand, offer real-time insights into human behavior, information sharing, and trends.

By analyzing YouTube channel statistics, such as viewership trends and engagement metrics, this project aims to identify correlations between the growth or decline of these stats and significant pandemic events. The hypothesis is that changes in these metrics can reflect shifts in public behavior, interest, or concern, potentially serving as an early indicator for disease spread patterns or societal responses to pandemics.

How I will solve my problem

To solve this problem, the project employs computational techniques to gather, process, and analyze data from YouTube channels via Socialblade, a platform that provides detailed statistics on social media accounts. The approach involves the following steps. First, father

historical data on YouTube channels, including metrics like daily view counts, subscriber growth, and engagement rates. Data spanning pre-COVID, during COVID, and post-COVID periods will be collected for comparison. Then, using statistical and machine learning methods to build predictive models that estimate the expected growth or engagement metrics for each channel if COVID-19 had not occurred. Now analysis, comparing the predictions to actual data during the COVID-19 period to identify deviations and assess how these might correlate with pandemic-related events. Then visualization, developing an interactive dashboard or visualization tool that highlights these trends, enabling users to explore data from various channels and timeframes. The final tool will allow users to view historical and predicted trends, analyze discrepancies, and draw insights about how external events like pandemics impact social media engagement.

Progress

Significant progress has been made in several areas. Research has been conducted to understand the pandemic's timeline and its potential impact on digital behavior. Additionally, Socialblade's data collection capabilities and YouTube's API documentation have been explored to streamline the data-gathering process.

On the technical front, code has been written to automate data collection and preprocess the retrieved data for analysis. The data pipeline can handle multiple YouTube channels and formats the information for model training. Initial experiments with predictive modeling have also been conducted using machine learning libraries, focusing on time-series forecasting techniques.

Most of the current effort has been devoted to acquiring data from a variety of YouTube channels to ensure a comprehensive dataset.

Expected results

The project aims to forecast YouTube channel metrics accurately and identify significant deviations during the COVID-19 period. These deviations are expected to align with key pandemic events, validating the hypothesis that social media metrics can act as indicators of societal behavior during global crises.

If successful, the project will not only demonstrate the feasibility of using public social media data for tracking global events but also provide a foundation for similar analyses in future pandemics or crises.

Works Cited

"Social Media Research: Tools, Techniques and Outputs." *Pulsar Platform*, www.pulsarplatform.com/what-is-social-media-research-top-tools-and-techniques. Accessed 15 Dec. 2024.

"Subject Guides: Social Media Research: Home." *Home - Social Media Research - Subject Guides at American University*, subjectguides.library.american.edu/socialmediaresearch. Accessed 14 Dec. 2025.

Zimba O, Gasparyan AY. Social media platforms: a primer for researchers. Reumatologia. 2021;59(2):68-72. doi: 10.5114/reum.2021.102707. Epub 2021 Jan 16. PMID: 33976459; PMCID: PMC8103414.

"Use of Social Media as a Research Activity." *Penn IRB*, 9 Mar. 2023, irb.upenn.edu/homepage/social-behavioral-homepage/guidance/types-of-social-behavioral-research/use-of-social-media-as-a-research-activity/.

Stacy Loeb et al., Methodologies in Social Media Research: Where We Are and Where We Still Need to Go?. *JCO Oncol Pract* **18**, 533-535(2022). DOI:10.1200/OP.21.00871