CIS 407 Sem Career/Internships

Sentiment Analysis Report

Author:

Jason Page

jpage2

Report

Introduction

In the age of digital media, the sentiment of online content plays a crucial role in shaping public opinion and influencing trends. Social media platforms, particularly video-sharing sites like YouTube, host vast amounts of user-generated content, making them rich sources of data for sentiment analysis. This report details the development of a sentiment analysis tool designed to scrape and analyze comments from YouTube videos. The tool utilizes natural language processing (NLP) techniques to categorize comments into positive, neutral, and negative sentiments, providing insights into audience reactions.

Background

Sentiment analysis, a subfield of NLP, involves identifying and categorizing opinions expressed in text to determine the writer's attitude towards a particular topic. It has applications in various domains such as marketing, customer service, and political analysis. This project aims to leverage sentiment analysis to assess the emotional tone of YouTube comments. By using APIs to scrape comments and machine learning algorithms to analyze them, this tool can help content creators, marketers, and researchers understand audience engagement and sentiment trends over time.

Implementation

The implementation of this project involves several key components: a web scraper, sentiment analysis module, and a command-line interface for user interaction. The scraper.py module contains functions for retrieving comments from a specified YouTube video using YouTube's API. The sentiment.py module includes functions to analyze these comments using NLP techniques to classify them into positive, neutral, and negative sentiments. The main script, main.py, orchestrates the workflow by accepting a video ID as a command-line argument, invoking the scraper to collect comments, and then analyzing these comments to output the sentiment results.

Performance Results and Discussion

The performance of the sentiment analysis tool was evaluated by testing it on a diverse set of YouTube videos, covering various genres and topics. The tool effectively scraped comments and categorized them accurately, demonstrating robustness in handling different types of user inputs. The results revealed interesting patterns, such as higher positive sentiment in tutorial videos compared to news videos, which had a more balanced sentiment distribution. However, the accuracy of sentiment classification could be further improved by refining the NLP models and incorporating more advanced techniques like deep learning.

Conclusion

This project successfully developed a sentiment analysis tool for YouTube comments, providing valuable insights into audience sentiment. The combination of web scraping and NLP techniques proved effective in analyzing large volumes of user-generated content. Future enhancements could include expanding the tool's capabilities to other social media platforms and integrating more sophisticated sentiment analysis algorithms. Overall, this tool serves as a useful resource for content creators and researchers aiming to understand and leverage audience feedback.