Social Media Comments Analysis

Project Report

Anil Kumar Kochera
School of Computing and Engineering
University of Missouri – Kansas City
Kansas City, MO
akkw32@umsystem.edu

Naveen Tiragati
School of Computing and Engineering
University of Missouri – Kansas City
Kansas City, MO
ntdpm@umsystem.edu

Vijay Agavinti
School of Computing and Engineering
University of Missouri – Kansas City
Kansas City, MO
vactp@umsystem.edu

Kalyan Kumar Saiempu
School of Computing and Engineering
University of Missouri – Kansas City
Kansas City, MO
ks6hp@umsystem.edu

ABSTRACT

Youtube is a great place for many content creators to upload their videos and allows viewers to comment, like, and share the video. When there are an immense number of comments on a video, it is practically not possible to reply to each and every one of them by the creator, so we provide an ML solution that replies to comments automatically. Also, this corpus of comments provides a very interesting opportunity to grab the overall sentiment of the comment section and ideas for future videos.

Motivation and Significance

Founded in 2005, Youtube has grown to become the second-largest search engine in the world (behind Google) that processes more than 3 billion searches per month. [[1]] It is, however, generally a myth how the Youtube algorithm works, and what makes a video get views and be recommended over another. In fact, YouTube has one of the largest and most sophisticated industrial recommendation systems in the existence [[2]]. For new content creators, it is a challenge to understand why a video gets video and others do not. There are many "myths" about the success of a Youtube video [[3]], for example, if the video has more likes or comments, or if the video is of a certain duration. It is also worth experimenting and looking for "trends" in the topics that Youtube channels are covering in a certain niche.

We have decided to gain some insights on this topic which might be useful for other new content creators. The scope of this small project is limited to data science channels. Therefore, this project will explore the statistics of around 10 most successful data science Youtube channels.

Objectives

Within this project, we would like to explore the following:

1. Getting to know Youtube API and how to obtain video data 2. Analyzing video data and verifying different common "myths" about what makes a video do well on Youtube, for example: Does the number of likes and comments matter for a video to get more views? 3. Does the video duration matter for views and interaction (likes/ comments)? 4. Does title length matter for views? 5. How many tags do good-performing videos have? What are the common tags among these videos? 6. Across all the creators I take into consideration, how often they upload new videos. On which days in the week? 7. Explore the trending topics using NLP techniques 8. Which popular topics are being covered in the videos (e.g. using word cloud for video titles)? 9. Which questions are being asked in the comment sections in the videos? These are the objectives we will primarily focus on.

Approaches

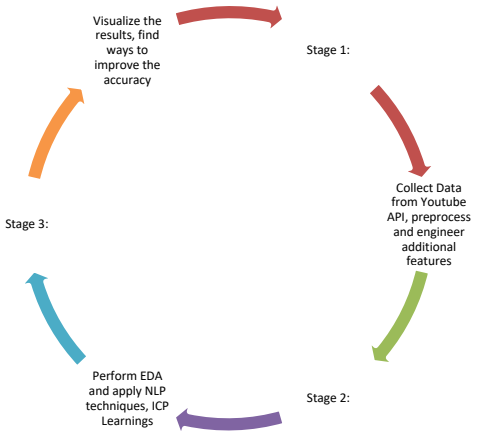

Fig: 1 – Approach Lifecycle

We decided to obtain the video metadata via Youtube API for the top 10-15 channels in the data science niche (this includes several small steps: create a developer key, request data and transform the responses into a usable data format) Preprocess data, and engineer additional features. 3. Perform EDA and apply NLP techniques, and ICP Learnings. 4. Show the conclusions in a visually appealing dashboard.

Results

GitHub:
https://github.com/vijay9237865/Social_media_Comments_Analysis

Planned Work

Our project is currently in Stage 1 (from fig 1). We are working on collecting the data from the Youtube API and creating a corpus to proceed to the next stages.

Conclusion

In this project, we have explored the video data of the 9 most popular Data Science/ Data analyst channels and are working on revealing many interesting findings for anyone who is starting out with a Youtube channel in data science or another topic.

Acknowledgment

Our team would like to thank Professor. Yugyung Lee, our TA Duy H. Ho for teaching and assisting us in ICPs which helped us learn incrementally.

References

[1] https://www.mushroomnetworks.com/infographics/youtube---the-2nd-largest-search-engine-infographic/

[2] https://dl.acm.org/doi/10.1145/2959100.2959190

[3] https://vidiq.com/blog/post/5-youtube-algorithm-myths-youtubers-need-to-know-about/