ABSTRACT

Helping prospective youtubers choose video Content

YouTube is one of the most popular platforms for making money online. The biggest challenge that every youtuber face is **"which videos grab more user's attention(views)?**" more specifically, **what should be the video content so that the video can obtain more user views**. In order to answer this question, we have made big data analytics on Kaggle dataset (YouTube's trending video statistics) alternatively, we can also collect data in real-time using YouTube data API. Our Kaggle dataset contains information such as video_id, trending_date, **views**, likes, dislikes, title, channel, **tags**, category, etc. we provide a complete solution about choosing actual video content using both **tags column** (example: NBA| "Basketball"| "Sports") and **views column** of dataset as input and running an algorithm similar to word count algorithm in MR where mappers split the tags by '|' character, associate view count as initial counter and reducers reduce the similar tags. This algorithm outputs tag view counts and when visualized though tools like tableau youtubers can decide which video content to choose. We use spark framework to run this algorithm because spark provides faster data processing, minimal implementation code compared to Hadoop. Also, we can perform real-time data processing using spark.

Dataset : https://www.kaggle.com/datasnaek/youtube-new#USvideos.csv---trending

Team : Sai krishna uppala(suppala2)
Sampreeth bodireddy(sboddire)