

**COLLEGE OF COMPUTING AND INFORMATION SCIENCES**

**DEPARTMENT OF NETWORKS**

**BACHELOR OF SCIENCE IN SOFTWARE ENGINEERING (YEAR 2) RECESS TERM 2 (BSE2301)**

**GROUP 16**

**CONCEPT PAPER**

**PROJECT MEMBERS**

|  |  |  |  |
| --- | --- | --- | --- |
| **NAME** | **REG. NO** | **STD. NO** | **SIGNATURE** |
| KAGANZI TIMOTHY | 16/U/5134/EVE | 216007287 |  |
| KASOZI BENJAMIN | 16/U/ 5538/EVE | 216015602 |  |
| OKURUT JULIUS | 16/U/10950/PS | 216014096 |  |
| KUNYA NATHERN | 16/U/6372/EVE | 216006158 |  |

**PROJECT LEADER**

KAGANZI TIMOTHY

**SUPERVISOR**​: ISAAC MBABAZI

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE SOFTWARE ENGINEERING RECESS PROJECT BSE 2301

21​TH​ JUNE, 2018

**TRENDING YOUTUBE STATISTICS ANALYSIS**

YouTube is a free video sharing website that makes it easy to watch online videos. You can even create and upload your own videos to share with others. Founded in 2005, YouTube is arguable the biggest video platform worldwide featuring a wide variety of user-generated and corporate media content that include music videos, TV clips, as well as other video clips such as video blogs, short original videos, Let's Play gaming videos, instructional videos on everything from language learning to stain removal, as well as videos with educational content.

YouTube performs video ranking to its viewers using the different techniques that will be discussed later. The datasets provided in this case contain the YouTube viewer’s opinions about the videos in form of likes, dislikes, comments, shares among others

**BACKGROUND TO THE PROBLEM**

YouTube datasets were issued to us as a group to study and analyze how YouTube as a platform can help individuals, companies and other organizations to understand the users view about a certain product in market or his/her suggestion(s) concerning a topic of discussion via a certain YouTube channel. The problem here is for us to determine how YouTube uses this kind of data to be able to tell the most trending videos, categories them according to viewers’ likes and be able to predict the most probable video a particular viewer maybe interested at a particular time.

The datasets that we are going to use were extracted from some channels on YouTube and we are required to analyze them based on the different RStudio techniques for example regression, data visualization, neural networks among others.

These datasets were collected using the YouTube API. Data is included for the US, GB, DE, CA, and FR regions (USA, Great Britain, Germany, Canada, and France, respectively), with up to 200 listed trending videos per day.

Through this problem, we shall be able to discuss the following in particular

1. Sentiment analysis in a variety of forms
2. Categorizing YouTube videos based on their comments and statistics.
3. Training ML algorithms like RNNs to generate their own YouTube comments.
4. Analyzing what factors affect how popular a YouTube video will be.
5. Statistical analysis over time

**THE PROBLEM THE PROJECT IS ADDRESSING**

To be able to use the analysis software called R to understand what YouTube does to determine the year’s top-trending videos, how YouTube can help stakeholders to make decisions based on the analysis for example measuring user’s interactions (number of views, shares, comments and likes) and also to be able to use YouTube to know the top performers on the YouTube trending list.

**THE MAIN OBJECTIVE OF THIS PROJECT**

To understand the basics of data analysis in R based on the data extracted from different YouTube channels and how YouTube does its data analysis and make decisions on what is the most trending video.

**OBJECTIVES OF THE PROJECT**

To categorize YouTube videos based on the comments, views, shares, likes and dislikes.

To analyze what makes certain videos more trending or popular than others and how it affects YouTube.

To generate statistical information that is important for the various stakeholders.

To train models that will be able to generate comments based on various YouTube videos.

**OUTCOMES OF THIS PROJECT**

1. Most trending videos
2. Categorize videos according to views and likes
3. More knowledge on the use of Rstudio
4. Trained model that can generate comments

**METHODOLOGY TO BE USED TO ACCOMPLISH THE OBJECTIVES**

Sentiment analysis

Regression

**REFERENCES**

|  |  |
| --- | --- |
| [1] | Anon [Online]. url available[: https://www.webwise.ie/parents/what-is-youtube/](:%20https:/www.webwise.ie/parents/what-is-youtube/) |
| [2] | Anon [Online]. url available: <https://datanice.wordpress.com/2015/09/09/sentiment-analysis-for-youtube-channels-with-nltk/> |