

**COLLEGE OF COMPUTING AND INFORMATION SCIENCES**

**DEPARTMENT OF NETWORKS**

**BACHELOR OF SCIENCE IN SOFTWARE ENGINEERING (YEAR 2) RECESS TERM 2 (BSE 2301)**

**CONCEPT PAPER**

**PROJECT MEMBERS (Group 16)**

|  |  |  |  |
| --- | --- | --- | --- |
| **NAME** | **REG. NO** | **STD. NO** | **SIGN** |
| 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

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

**BACKGROUND TO THE PROBLEM**

Youtube data sets were issued to us as a group to study and analyze how Youtube uses this kid of data sets to be able to tell the most trending videos, categories them according to viewers’ likes, views, shares and comments, thus be able to predict the most trending video, as well, be able to tell the kind of videos a viewer maybe interested in viewing, for example, learning tutorials, football videos among others.

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 to be able to use Youtube to know the top performers on the Youtube trending list.

**THE MAIN OBJECTIVE OF THIS PROJECT**

To be able to determine how Youtube predicts the most trending videos as well as what a viewer may want to view whenever they get to Youtube. 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 generate statistical information that is important for the various stakeholders.

To train models that will be able to generate trending videos according to viewers’ views, comments, likes and shares

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

**OUTCOMES OF THIS PROJECT**

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

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

Neural networks

Regression

**REFERENCES**

1. <https://www.webwise.ie/parents/what-is-youtube/>

**CONTACT INFORMATION**

Okurut Julius 0704476718 [okurut.julius@cis.mak.ac.ug](mailto:okurut.julius@cis.mak.ac.ug)