



International Technology University

2711 N 1st St, San Jose, CA, 95134

Machine Learning

Project Plan
On

Sentiment Analysis on YouTube videos to Discover Pattern for Traffic Generation

Team Members	Professor
[94844] Priyanka deshमुख [94739] Dmitrii Dunin [95302] Eric Palma [95583] Alp Tureci	Dr. Patricia Hoffman Machine Learning Computer Science Department

Objective

This project helps to find out the most relevant attributes to predict if a video will be trending on the YouTube platform.

Background

Youtube is one of the most famous platforms for uploading and watching videos on the web where users continuously upload videos in real time. It is one of the most popular sites in social media, where users interact with sharing, commenting and rating videos. Ranking of video quality is based on how much likes that video have. According to this ranking the relevant videos listed on the search result. Sometimes irrelevant videos with low quality are ranked higher and display on search results. To minimise this issue, we use Sentiment Analysis strategy for ranking the videos. We do a Sentiment Analysis approach on user comments..

The effectiveness of the proposed scheme has been proved by a data driven experiment in terms of accuracy of finding relevant, popular and high quality video.

We have taken a dataset from kaggle. kaggle has a dataset of two countries UK and US. As for now, we will take US data. Our study focused on around 2500 videos from a database that had around 8000 data points.

Dataset attributes

Video_id
Comment
Likes
Dislikes
Tag
Title
Channel_title
Category_id
Comment_total
Views

Sentiment Analysis Background

Sentiment analysis is the process of examining a piece of text for opinions and feelings. There are innumerable real-life use cases for sentiment analysis that include understanding how consumers feel about a product or service, looking for signs of depression, or to see how people respond to certain ad and political campaigns.

The applications of sentiment analysis are broad and powerful. Sentiment analysis can be used to give valuable insights to businesses on how people feel about their brand. When applied to social media channels, sentiment analysis can give insights on what the people think about a certain topic or advertisement.

It's also a very important tool in the mental health field and can be implemented to identify signs of depression or anxiety in a person and inform family members.

Project Layout

Step 1: Clean videos dataset.

Step 2: Clean comments dataset.

Step 3: Apply decision trees model and compare with linear regression to videos dataset

Step 4: Apply Naïve Bayes model to comments dataset to perform sentiment analysis.

Step 5: Document elaboration.

Step 6: Review and submission of the project.

Team Member

Mem ber	Name	Dept	Student ID	Email
1	Priyanka Deshmukh	SWE	94844	deshmukhpriya2858@students.itu.edu
2	Eric Palma	CS	95302	palmaeric2254@students.itu.edu
3	Dmitrii Dunin	SWE	94739	dunindmitr2772@students.itu.edu
4	Alp Tureci	SWE	95583	turecialp2500@students.itu.edu

Project level responsibility

S.N.	Role	Member
1.	Leader	Priyanka D
2.	Team member	Dmitrii D
3.	Team member	Eric P
4.	Team Member	Alp T.

Project Schedule

Deadline	Goals
March 31	Elaborate and present project plan.
April 7	Clean videos and comments dataset.
April 4	Perform decision tree model and linear regression to videos cleaned dataset. Perform Naive Bayes model to comments dataset.
April 9	Elaborate the final document.
April 14	Present the final document

Task List

S.N.	Tasks	Responsible member
1.	Lead project communication	Priyanka D
2.	Elaborate project plan	Priyanka D, Eric P, Dmitrii D, Alp T
3.	Clean videos data set.	Eric P
4.	Perform decision tree model to videos dataset	Priyanka D
5.	Perform linear regression to videos	Eric P
6.	Perform Naive Bayes to comments dataset	Dmitrii D / Alp T
7.	Perform Sentiment Analysis on Youtube Comments.	Priyanka D, Eric P, Dmitrii D, Alp T

8.	Elaborate final project document	Priyanka D, Eric P, Dmitrii D, Alp T
----	----------------------------------	--------------------------------------

Communication and Tracking mechanism

- We will be communicating by a WhatsApp group, email, and hold meetings via Google Hangouts/Zoom.
- We will track the project according to the schedule proposed.

Question

This step is unique to each member of the team:

10) Give more details for your particular tasks. Define the interfaces between your tasks and your teammates' participation. If you need certain inputs from your teammates, are they well defined and on the schedule of tasks? Describe them in detail. Will you be able to finish your tasks in the scheduled time frame? Is the review schedule and tracking mechanism reasonable for you? This is your commitment to your team.

I will perform Naive Bayes on the comments dataset. Despite being considered as one of the simplest categorization mechanisms, Naive Bayes may be helpful in text categorization. We may perform analysis on presence of separate words to be a trigger of video success. Everyone picked the way of the analysis they think can perform best.

11) Good tasks are ones that you would like to work on, that you have some ideas about, and that are small enough for you to finish by the end of the semester.

We are not trying to build a “YouTube success formula” based on comments analysis, but may start small and find more patterns to improve our model. First step of this analysis can definitely be finished until the end of the trimester.

12) Do you have ideas about how to complete the task? Do you know how to solve the problem, write the code, or build the publication? If not, can you reformulate the

problem to make it easier? Do you need help from a teammate in performing your task?

Alp Tureci will help me to perform Naive Bayes and sentiment analysis. I have shared a few ideas of general natural language analysis which I learned from my former colleague who was researching natural language processing and particularly developed a model which can produce artificial poetry as if it was written by Russian classics.

13) Are there interfaces that need to be defined between your task and a team mate's task? If two members are writing code, are the modules well defined and separated with clear interfaces? Are the pieces of the presentation well defined so that the whole project will be presented and also that two people are not presenting the same material?

We have separated modules needed so we could work independently on our tasks and then combine and compare results to produce a final report. Even though two persons can work on the same task they will be working on different parts and can highlight their task from their own perspective