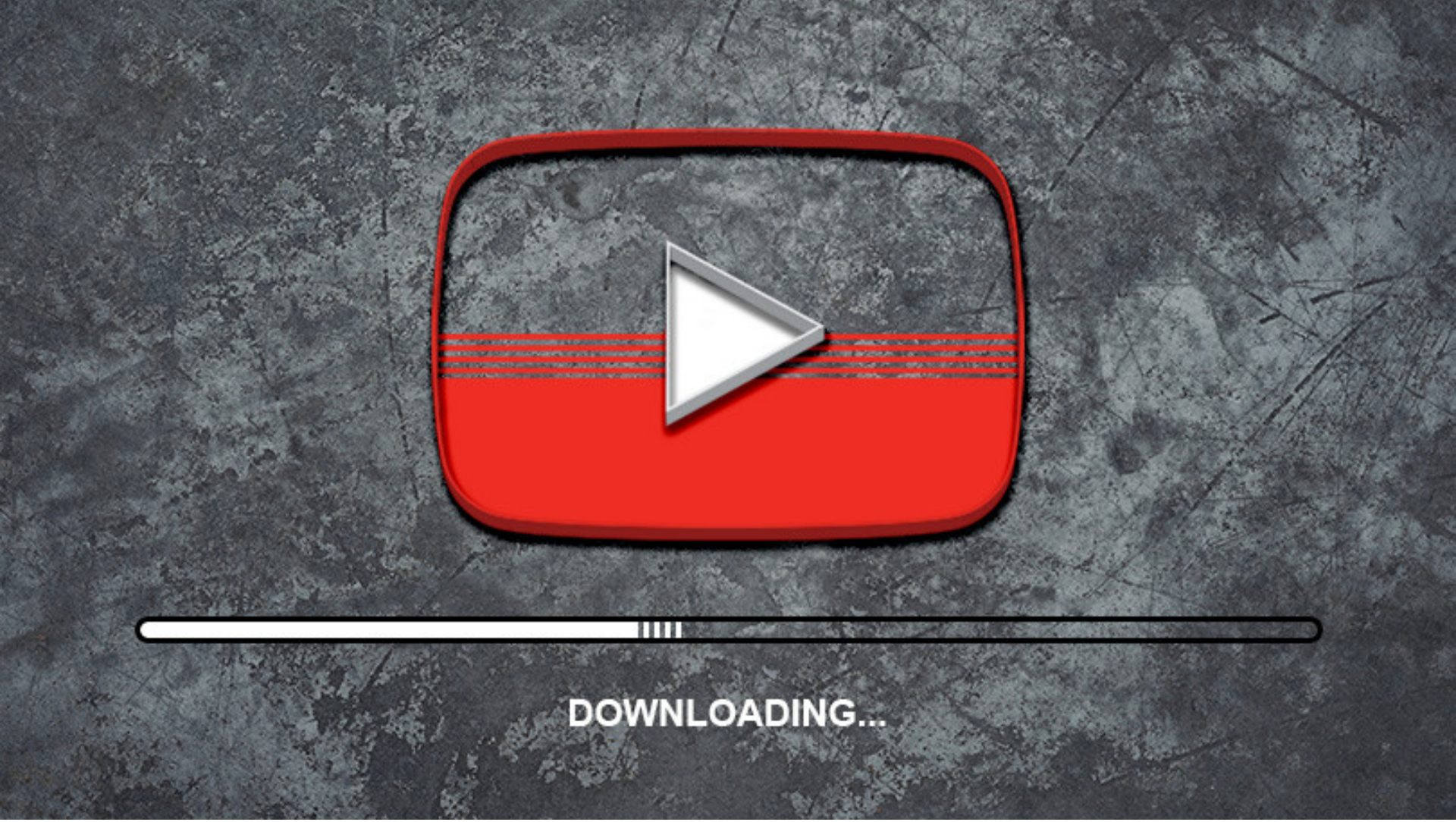


Predicting YouTube Video Popularity

Springboard Capstone Project

by Sara Peters

Summer 2019

Background

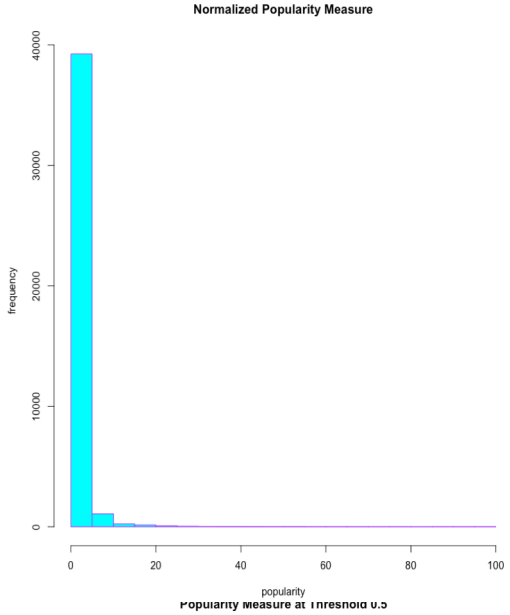
- Ranked as the second-most popular website in the world, the Google-owned video-hosting site YouTube claims that **one billion hours of user-generated content is watched on the broadcast site each day**.
- This global phenomenon is thanks in large part to the video site's *machine learning recommendation capabilities*.
- The record-breaking exposure channels can receive on the video platform has changed the game of marketing and afforded new career opportunities.
- "YouTubers," who are able to amass a huge following can take advantage of corporate sponsorships and sign deals to have product lines of their own.

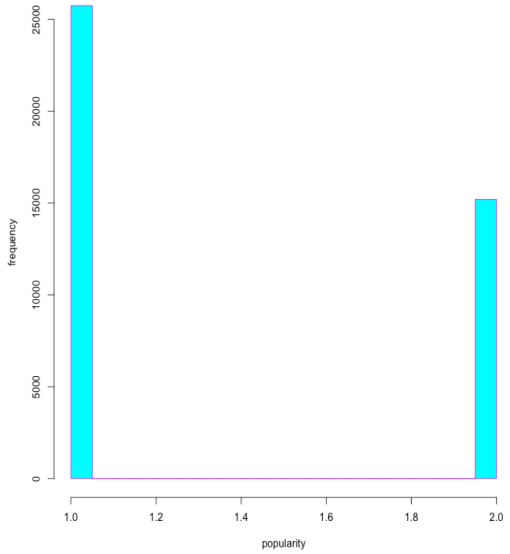
Question & Potential Clients

- Because there are a number of factors to consider when predicting what makes a video popular to one audience and not another, this project attempted to solve the following problem:
  - **What factors affect how popular a YouTube video will be and, using such factors, can we predict popularity for any video?**
- Private individuals to large production companies have used YouTube to grow their audience base. Because of this large range of interests, there are several clients who would want to know the answer to this project's question.
  - entertainment content creators (movie studios or musical artists)
  - advertisers selling a product
  - home improvement channels
  - travel destinations/resorts
  - education companies making video lessons designed for kids

the Data Set

- 40,949 observations
- ***popularity*** measure (dependent, outcome variable)
  - combination of 3 factors: view, like, and comment count
  - categorical: popular (15,201) v. unpopular (25,748)
    - highly skewed metric
    - applied 0.5 threshold to create reasonably balanced classes with a ratio of 3:5

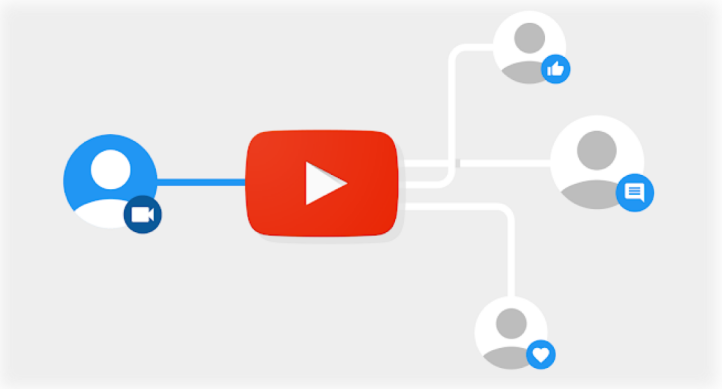




the Data Set

- 40,949 observations
- ***tags*** (independent, predictor variables)
  - 740 independent variables to represent the frequency of the tags used for at least 0.5% of the videos

- To reframe the question as a machine learning problem, the following was used:



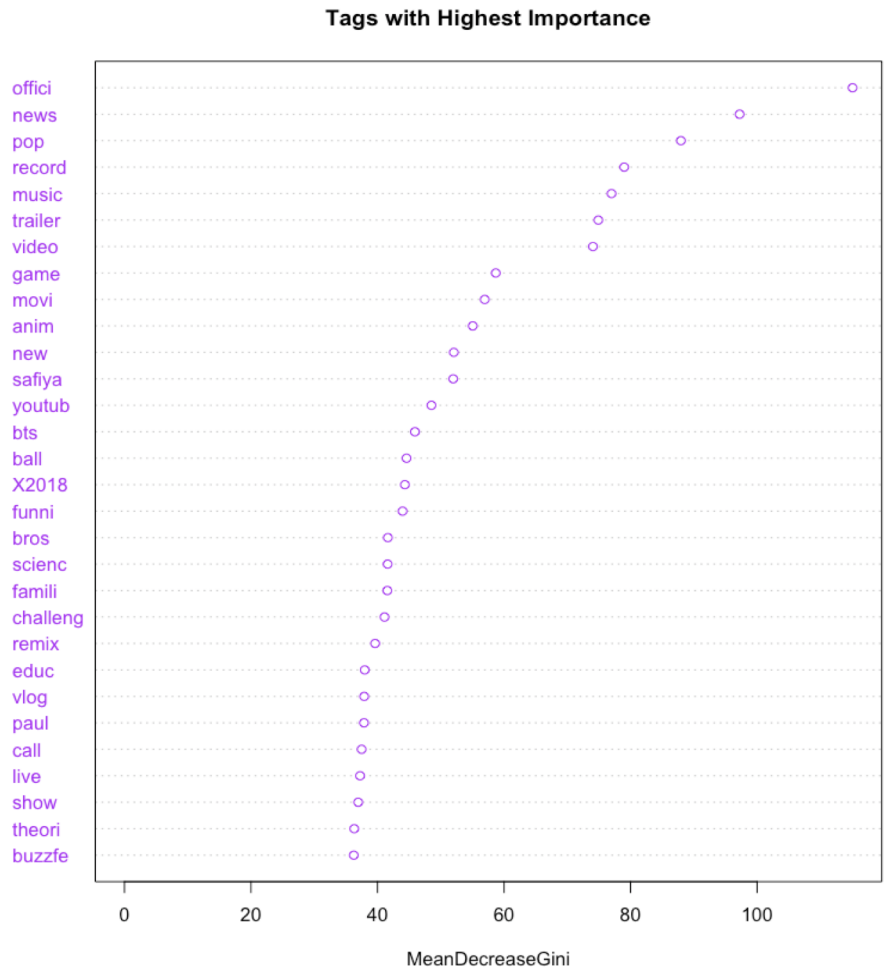
  - **Which tags predict popularity for a YouTube video?**

the Models

- 3 machine learning techniques were applied in addition to the baseline model
  - baseline = 63% accuracy
  - logistic regression = 71% accuracy & can differentiate between popular and unpopular videos pretty well
  - CART = 65% accuracy (even with cross-validation)
  - random forest = 91.5% & can differentiate between popular and unpopular videos very well
    - This model was the most accurate predictor of a video's popularity based on the tags used.

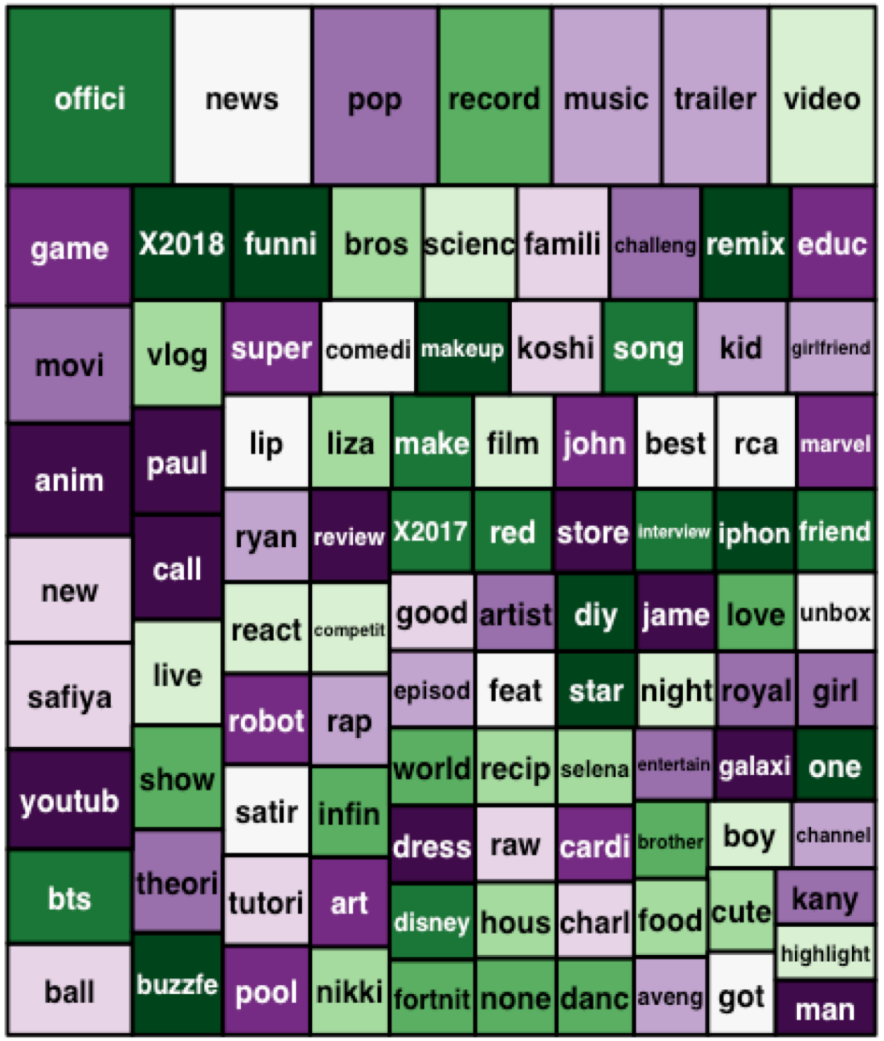
The findings

- The tags that are most important when it comes to predicting video popularity are displayed in the graph.
  - Note: Some of the tags have been stemmed during the pre-processing stage.
  - “offici” = “official” or “officially”



The findings

- The 100 most important tags used to predict popularity can be seen in the TreeMap.
  - Note: The bigger the box, the more important the tag is in terms of predicting a video's popularity.



Client Recommendations

- Multiple variations of the same tag don't need to be included. YouTube's search algorithm can filter the results accordingly to provide accurate search queries.
  - Example: Don't include both "game" and "games".
- Because the top 100 tags are mostly related to the top two video categories included in the data set (*Entertainment* & *Music*), clients should look for the most generic tags in the list to use in their YouTube video postings.
  - Suggestions: "video", "channel", "store"

Client Recommendations

- Positive attributes seem to be featured quite a bit and would be applicable to many channels.
  - Suggestions: “new”, “super”, “best”, “good”
- Depending on the video’s content, including the year could be useful to create a popular video. Both “2017” and “2018” were included in the top 100 tags list.
  - Example: “Top Songs of 2019” or “Official Trailer (2019)”