

Science Fair Project

By Eric Xie

ericxie6@gmail.com

Why I chose this area of science

I was interested in machine learning because I've always been interested in computers and coding, but I've never actually done anything useful with coding; I've only made games and solved made-up competition problems. I am looking to transition to using programming for more practical purposes using machine learning to solve real world problems.

The problem or challenge

My goal was to find which features of a video would cause it to end up on the trending page of YouTube

The hypothesis (or prediction)

There will be certain word prevalent in trending video features. These word will include popular topic such as celebrity names or “official” or “music”

Research

I have watched many Youtube videos, especially during quarantine. I've noticed that certain videos always make it on the trending page, specifically music videos and movie trailers. I'm curious to see whether there are specific words in text parts of these videos (title, description, tags) that help the, make it on the trending page.

My testing method

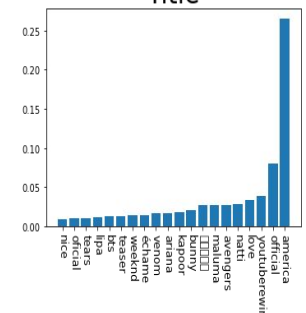
Each scientist uses different methods of experimentation

What methods did you use in your experiment?

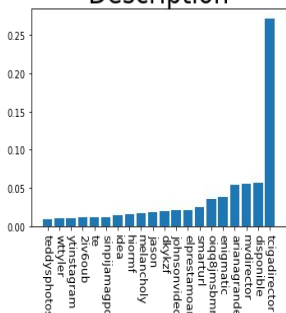
- I used a decision tree regressor to train my classifier
- After that, I took the 20 most word that were the most impactful to the training and put them on a graph
- I repeated these step using different dependent variables (views, views/likes, likes/dislikes)

Experiment data

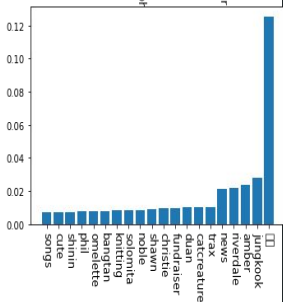
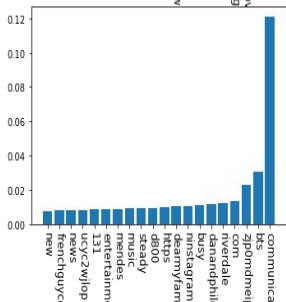
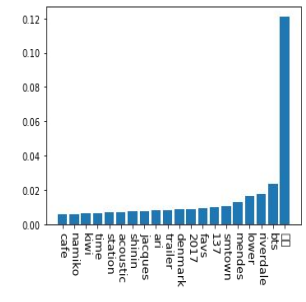
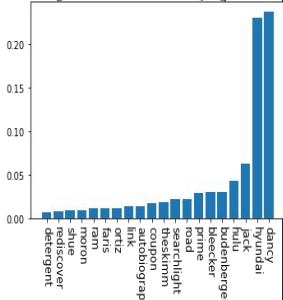
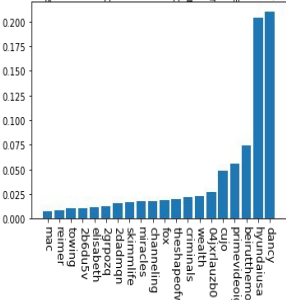
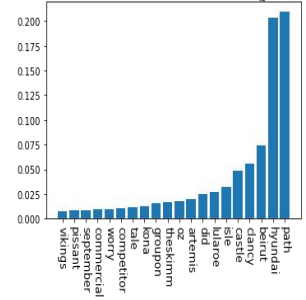
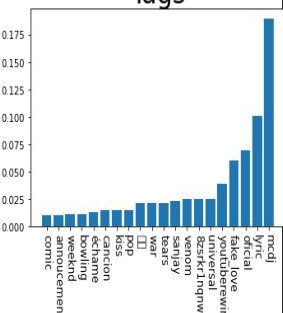
Title



Description



Tags



Aha!

My discoveries

What did you learn after testing?

1. All the test resulted in one or two top words that were far more important than the rest
2. The top title words of the trending videos that attracted the most views were “official”, “america”, “youtuberewind”, and “love”. These words are popular topics
3. Interestingly, the description has more obscure words
4. The tags are similar to the title with popular topics. The top ones are all music related, suggesting that music videos tend to trend

My discoveries cont.

The dependent variable in this case was views per like

1. The first thing I notice is that all the features top words include Hyundai. This shows that hyundai has been very successful with their advertising in the united states
2. The top word for both description and tags was “dancy”. I’m not sure what this word means. Maybe it refers to a dance or someone’s last name

My discoveries cont.

The dependent variable in this case was likes per dislike

1. All features' top word's were k-pop related
2. The empty squares are character that can't be rendered by the font; this can include chinese and korean characters
3. In general, k-pop is the most liked topic on trending videos

Conclusion

There are definitely words in the features of a Youtube video that help it trend. I think the words reflect what the population is interested in at the moment. For example, maybe in a few months, youtube rewind won't be relevant anymore. Maybe a new k-pop group enters the scene and causes a rise and k-pop related videos.

What will I do next?

Another media platform, TikTok, has been gaining lots of popularity recently. I want to look into what will make a video go viral on TikTok

