



YouTube Trending Prediction

Bingwen Hu, Dau Cheng, Xinran Wang, Rui Wang

Background

- Youtube is an online video sharing and social media platform
- Contents are generated by individuals
- Youtubers can only type keywords in search box to get the videos content
- Sponsors have no idea where to advertise and maximize the benefits.

Target Users & Objectives

- Youtubers: Search for current and future trends

What kind of videos do people usually like to watch?

What topics are more likely to be popular in the future?

- Sponsors: Find influencers to corporate and advertising

Whose videos are more popular now so that I can corporate with?

What type of videos I should choose to advertise to promote my products?

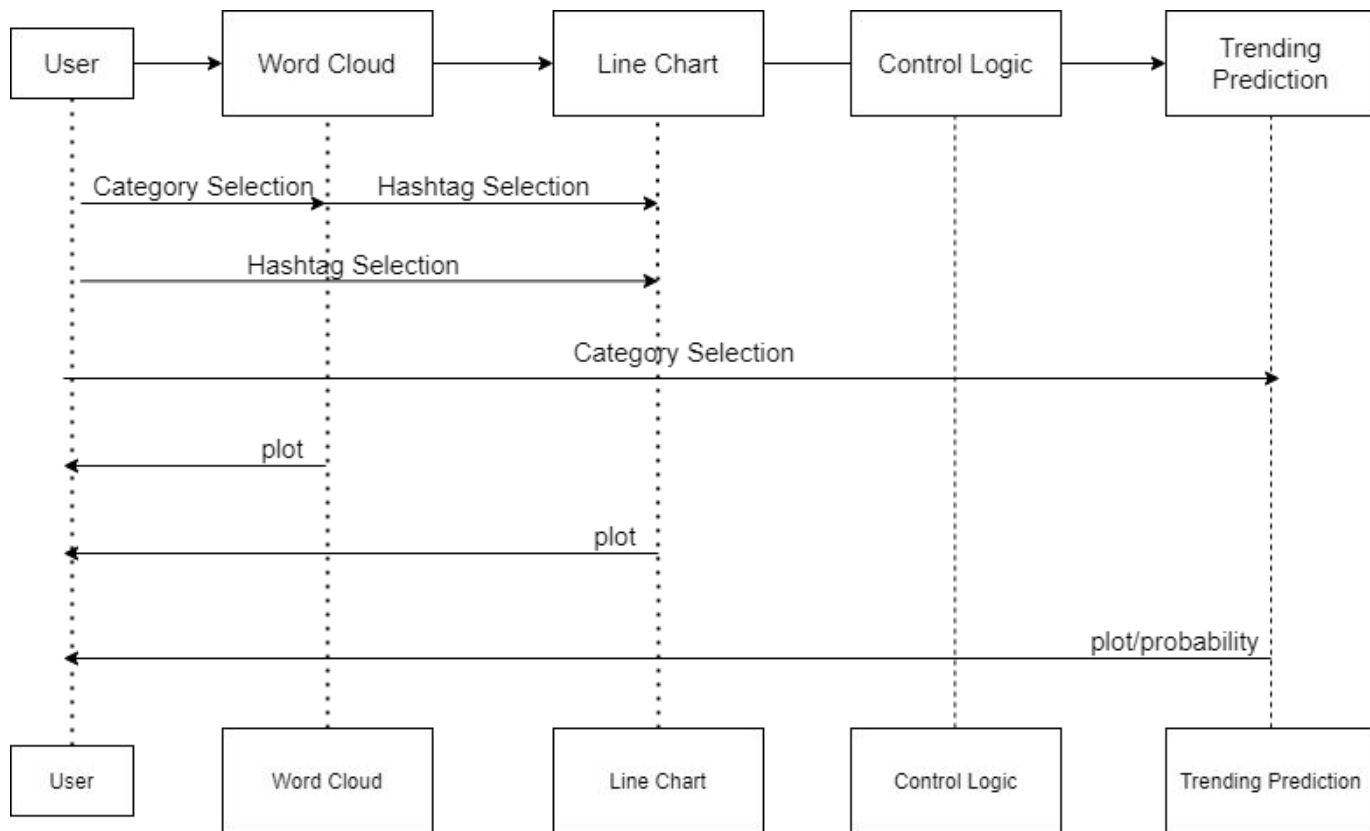
Data Source

- Data1: Youtube Trending Video Dataset

Data includes the video title, channel title, publish time, tags, views, likes and dislikes, description, and comment count.

- Data 2: Top search on Youtube

Design

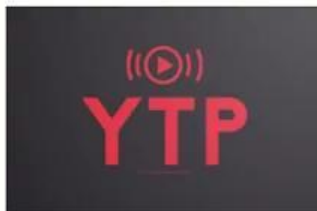


Demo

Youtube Trending Prediction

CSE 583 Final Project

By: [BingWen Hu](#), [Dau Cheng](#), [Ray Wang](#), [Xinran Wang](#)
Dec, 4, 2021 (Latest Update)



Introduction

YouTube is an online video sharing and social media platform, whose content is mostly generated by individuals, including YouTubers and corporate sponsors. YouTubers take advantage of the platform to share their skills and experience, while sponsors promote their products.

Technologies

- For project:
 - Python, pandas, numpy, matplotlib, scipy, etc
 - Git and shell scripting
 - Numerical methods for prediction, ML also tried
 - Tableau and github.io for visualization
 - GitHub Actions combined with flake 8 and unittest for testing and CI
- For Collaboration
 - GitHub, of course
 - GDrive
 - Discord

Challenges

- GitHub Actions with pylint:
 - 10/10 score of code style, otherwise terminate the CI
 - Exit codes without definition or documentation, nor instructions online
- Large dataset
 - Coping with git-lfs
 - Downscaled dataset/subset for unittest
- Relative path
 - Tricks of relative path sometimes can still get us lost
- ML models to make predictions
 - Groups of short time series
 - LSTM, GluonTS, not suitable for our dataset
 - Plain RNN maybe

Lessons Learned

- Technology:
 - Programming in Python: numpy, pandas, sklearn, altair...
 - Shell
 - Git

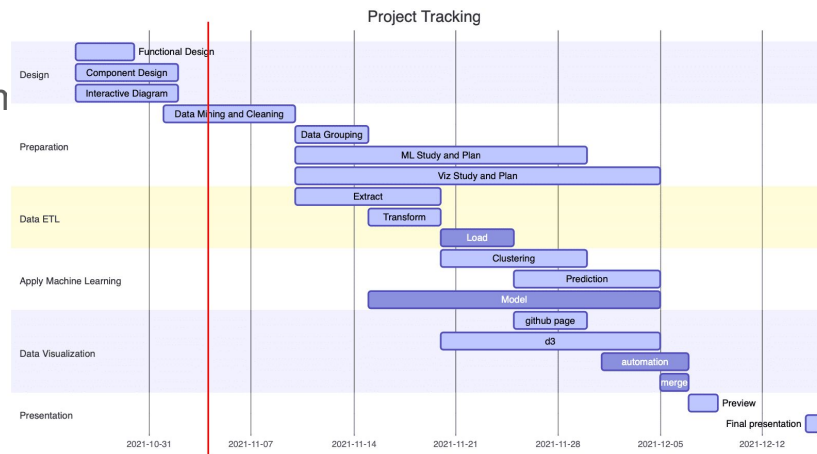


Lessons Learned

- Technology:
 - Programming in Python: numpy, pandas, sklearn, altair...
 - Shell
 - Git
- Workflow/pipeline:
 - Design: Functional Spec, Design Spec
 - Development: Jupyter notebook, package, module...
 - Test/CI: unittest, debug

Lessons Learned

- Technology:
 - Programming in Python: numpy, pandas, sklearn, altair...
 - Shell
 - Git
- Workflow/pipeline:
 - Design: Functional Spec, Design Spec
 - Development: Jupyter notebook, package, n
 - Test/CI: unittest
- Collaboration:
 - Github
 - Project tracking
 - Documentation



Thanks you!