YOUTUBE SUBSCRIBER PREDICTION

--- #YOUTUBE #PREDICTION

YIWEI ZHANG

06/10/2019

PROJECT OVERVIEW

Motivation:

- Provide stakeholders visions on potential product promotion opportunities with entry-level YouTubers
- Provides quantitative support to YouTube-related marketing decisions

Problem:

- User input information about a young YouTube channel
- Predict the subscriber amount of the channel in the next 2, 4, 5 and 7 years

DATA DESCRIPTION

- YouTube channel dataset from https://gitlab.com/thebrahminator/Youtube-View-Predictor/tree/master/datasets
- 3 M entries, each channel as an individual entry
- 27 Variables: subscriber count, channel view count, video count; Views / channel time, likes / dislikes, comments / views; channel time (in hour)
- New variable channel days calculated from data
- Split data into 4 cohorts based on channel days to train the different channels separately

MODELING AND SUCCESS CRITERIA

- Feature selection: Gradient Boosting, 14 variables in final dataset, 7 variables as user input
- Split data: 67% as training and 33% as validation, 10-fold cross validation on the training set of each cohort
- Modeling: K-Nearst-Neighbour with “minkowsi” distance metric, 5 neighbors for the first 3 cohort and 10 neighbors for the last cohort
- Success criteria:

Root-mean-square-error

on the validation set smaller than 10% of the range of the actual value

1.91% 3.26% 2.27% 2%

R^2

Explains a good amount of the variation (larger than 0.1)

0.11 0.36 0.67 0.27

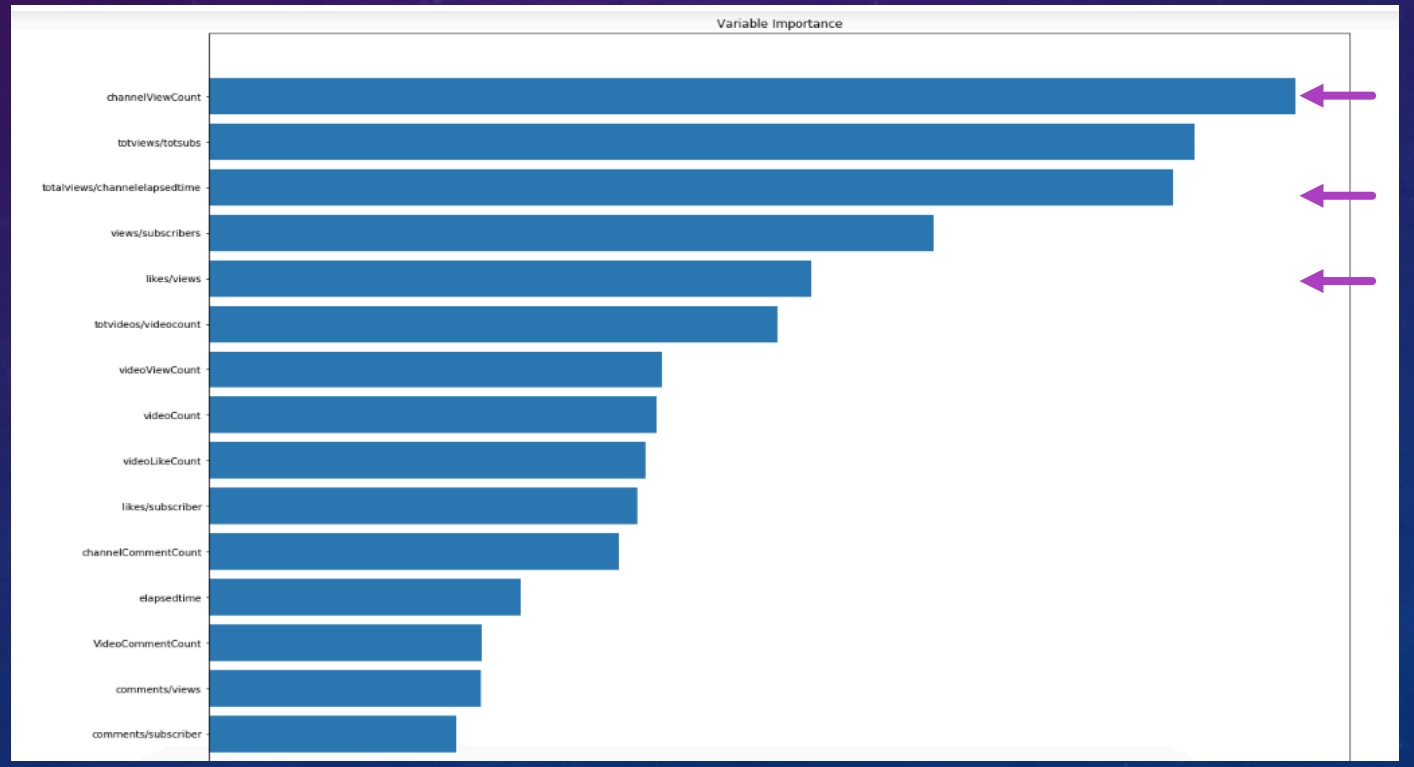
User satisfaction

Reported user satisfaction higher than 80%

To be tested

INSIGHT

- Important features selected by the Gradient Boosting Model. Variables were selected based on this result to fit the prediction model
- View count contributes the most to predict future subscribers



View count

View count / channel time

Like count / view count

THANK YOU!

- Questions?