

Netflix recommender project call

26th May 2021

Present: Michael Hoang (MH), Rucha Sharma (RS), Vivek Kotecha (VK)

Where we are at (current skills, reasons for participation, availability)

MH

Has carried out the project in R and built a content-based recommender system (a lot of work!). Would like others to build a similar project in Python to complement his project.

Offers a good opportunity for collaboration and as a community-built unguided project.

Currently has time to help with writing up articles and assist with data cleaning/wrangling steps.

RS

More comfortable working on the earlier steps, less experience of machine learning (ML) aspects.

Wants to show she can collaborate, this project offers a good opportunity to do so, and reinforce existing skills.

Has time available to dedicate to project over coming weeks.

VK

Is currently at the start of the ML courses for Python but keen to learn whilst on this project. Has a statistics/econometrics background so some aspects of ML are 'there' but not applied in python yet.

Wants to practise collaboration on Github, help build up some unguided projects for the community, and to reinforce python skills.

Self-employed so availability fluctuates. Minimum that can be given is usually 5-7 hours a week, can rise to 15-20 hours depending on work for that week.

Brief overview of project stages

Using dataset: <https://www.kaggle.com/shivamb/netflix-shows>

For **full details** on the stages and the overall project, please see MH's great presentation: https://drive.google.com/file/d/1_aZGb3XLdnJyVBP9o78lYKgclJ3X2lGe/view

1st stage – foundational level

Data cleaning; simple visualizations; answer a few research questions

2nd stage – intermediate level

For us to decide, possibilities: more advance visualizations e.g. interactive dashboard; combine in a second dataset for analysis (e.g. tv/film ratings from [IMDB](https://www.imdb.com/));

3rd stage – advanced level

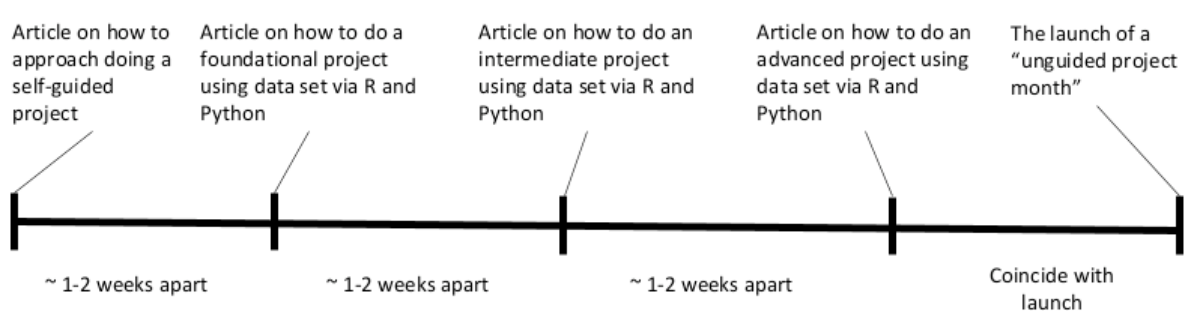
Build a content-based recommender using the Netflix data.

4th stage – not expected to be done but if we had time after the project is complete

Build a hybrid recommender system e.g. combine content ratings with user ratings for each show/film; enable the content recommender to take in user-downloaded Netflix viewing data.

Key for all stages is that the work builds on and references the prior stage, is not too time consuming or overwhelming, and scalable.

Articles



MH:

Best to have done some/most of the analysis in advance of publishing the articles as this avoids large gaps between them (they are meant to form a coherent series).

Important to document work as we progress so can be referenced in our articles.

Each successive article should reference prior ones to encourage readers to progress from the start and not just jump in.

Proposed timeline

2 week intervals

Coding

w/c 7th June – stage 1 analysis drafted

w/c 21st June – stage 2 analysis drafted

w/c 5th July – touch base on progress on recommender system

End July – all 3 stages done

Articles

Mid-June – kick-off article

Subsequent articles – 2 weeks later (possibly separate articles for python and R)

Next steps

VK:

Set up Github repository (<https://github.com/vivek-kotecha/netflix-recommender>)

)

Circulate draft meeting notes, upload to github
Explore the data set.

RS:

Start stage 1 and commence dataset cleaning.

MH, RS:

Work on introduction article