

## Movie&TV Topic Modeling and Collaborative Recommender System

Question/Need:

- What is the question behind your analysis or model and what practical impact will your work have?

**Looking at Movie&TV reviews, we can see what people are feeling about the content and what they care most about the content/product. By analyzing the Amazon Movie&TV reviews, it may help Amazon knows better of their customers' needs and have a better idea introducing new products in the future.**

**Collaborative Recommender System is to better recommend customers the products they may like.**

- Who is your client and how will that client benefits from exploring this question or building this model/system?

**Amazon is my client and my analysis will help them to improve their business choices on movie&TV product.**

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data? Please include a link! (The link can be to the dataset you're downloading, the site you're scraping, etc.)

**Data is from <https://nijianmo.github.io/amazon/index.html> (Jianmo Ni, Jiacheng Li, Julian McAuley Empirical Methods in Natural Language Processing (EMNLP), 2019) under the 'Movies and TV' session. I used a subset of the dataset and later used a smaller fraction of the dataset because of running time and memory limit. I ended up using 1690 rows of reviews with total 69565 terms at this point.**

- What is an individual sample/unit of analysis in this project? In other words, what does one row or observation of the data represent?

**For topic modeling, one row of observation is a customer review.**

**For Recommender System, one row of observation is the customer ID, product they bought and the rating they gave.**

- What characteristics/features do you expect to work with? In other words, what are your columns of interest?

**My columns of interest are customer id, product ID, review, rating.**

- If modeling, what will you predict as your target?

**My target is the topics for topic modeling and recommendations for a certain customer for Recommender System.**

Tools:

- How do you intend to meet the tools requirement of the project?

**Text Processing, Tokenization: Regex, NLTK, TfidfVectorizer**

**Topic Modeling: NMF**

**Recommender System: Light FM**

**Visualization: Matplotlib, WordCloud**

- Are you planning in advance to need or use additional tools beyond those required?

**No**

MVP Goal:

- What would a [minimum viable product \(MVP\)](#) look like for this project?

**WordCloud of common terms of the reviews**

**List of topics**