

Exploring Text in ^{the} SIMPSONs



Project Description

For this project, I will be exploring a dataset containing information about the television show, *The Simpson's*.

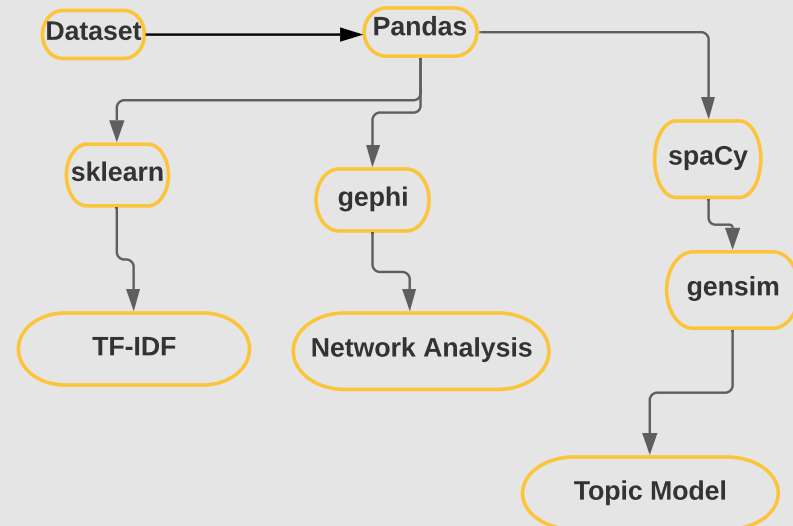
Justin Carlou "Jus" Lim | DigHum100 Dr. Adam Anderson | May 30, 2021

Questions for Analysis

Currently a work in progress, working on it with Prof. Anderson. I mostly have the models, but some ideas I was thinking about are:

1. What show themes are most significant?
2. Which characters talk to who, and how heavily?
3. What are the most common topics of *Simpson's* dialogue?

Tools & Methods



Dataset

In this project, I use Prashant Banerjee's "*The Simpson's Dataset*" posted on Kaggle. This dataset includes information about each line a character says in the show, where the characters say the line, and IMDb ratings for each episode. The dataset is split into four datasets: the character dataset, the episode dataset, the locations dataset, and the script lines dataset. Each were combined and formatted using the pandas library.

Interpretation of Results

From the TF-IDF, we can find which show themes are most relevant. From there, we can see how those show themes correspond with the network created between Simpson's characters. From there, we can see how the show themes correlate to the most prominent topics discussed in the dialogue— given the context of who talks to who in what settings.

TF-IDF Model

Network Analysis viz

Topic Model Viz

TF-IDF Interpretation

Network Analysis Interpretation

TM Interpretation



Works cited:

- <https://www.kaggle.com/ruchi798/sentiment-analysis-the-simpsons/notebook#Unigrams,-Bigrams-and-Trigrams>
- <https://www.kaggle.com/prashant111/the-simpsons-dataset>
- <https://towardsdatascience.com/topic-modelling-in-python-with-nltk-and-gensim-4ef03213cd21>