

NLP_Project_Repository

My project focuses on testing topic modelling techniques on picture and hashtag data curated from Tumblr.

NOTEBOOK NOTES:

- Original txt file of README is left in the folder for reference.
- Topic modelling cannot work on individual hashtag strings, better suited for the entire dataframe
- LSA was successfully able to extract the relevant images
- LDA is not able to perform in the same way (since query will just be string and cosine similarity cannot really work with bag of words).