**TECHNICAL ASSIGNMENT – 02**

The tools and methods I have used for this assignment are listed below:

1. Reddit’s developer API through which I scraped the subreddit “Netflix” to find the top 25 posts from last week. I used a filter to find the top 50 posts in this subreddit and extracted 25 out of them for analysis. The package used for accessing this API is “praw”.
2. The method used to find out the top words is Latent Dirichlet Allocation using topic modelling.

These top words suggest that most posts are relevant to “Netflix”. The top words suggest that posts are around streaming a netflix show. Since I have selected the top 50 threads and scraped 50 out of those, the top words are along the lines of a popular show “House of Cards” and when it is streaming and also some words suggest discussions about subscriptions and movie years.

A peculiar observation is that most reddit threads revolve around same discussions of House of Cards movies and subscriptions. These are directly related to subreddit netflix, we can maybe say that the users sway away from the topic by indulging in discussions of subscriptions instead of just shows.

So, based upon the results of LDA algorithm and context of the forum, I have concluded that online communities stay on topic. This is substantiated by the top words listed, which are mentioned in the appendix in this document. This is substantial proof that posts are relevant to the topic and the redditors make it a point to categorize and stay on topic specified as much as possible.

This could be the by product of enforcing used by moderators as most subreddit threads mention that uncategorized, violent or abusive language will be swiftly dealt with.

APPENDIX:

Topic 0:

Mean Netflix Posts: 0.04694558630097243

year edit streaming cards spacey movie watching shows like don new just time house watch

Topic 1:

Mean Netflix Posts: 0.02611225344383287

don edit like movie year watch streaming spacey house watching new shows cards time just

Topic 2:

Mean Netflix Posts: 0.026112253436054705

shows new don streaming spacey watching cards watch house edit time just year movie like

Topic 3:

Mean Netflix Posts: 0.12402891974030124

house cards spacey new year movie watch shows don streaming just like watching time edit

Topic 4:

Mean Netflix Posts: 0.0261122534309905

cards spacey time house like don just watching movie shows edit streaming year watch new

Topic 5:

Mean Netflix Posts: 0.026112253458562412

year shows movie cards just don house new spacey watch watching streaming like edit time

Topic 6:

Mean Netflix Posts: 0.026112253447062736

new just don watch like streaming spacey shows edit movie year cards time house watching

Topic 7:

Mean Netflix Posts: 0.026112253431729664

like movie spacey streaming house shows watching edit cards time year just new watch don

Topic 8:

Mean Netflix Posts: 0.026112253432904988

just watching don cards shows like new spacey edit house movie time watch year streaming

Topic 9:

Mean Netflix Posts: 0.15169704798493794

movie new watch just like edit time don year cards spacey shows watching streaming house

Topic 10:

Mean Netflix Posts: 0.026112253430844043

spacey just time watching movie like don house new edit year streaming watch cards shows

Topic 11:

Mean Netflix Posts: 0.026112253470185292

don year just time movie house new watching watch like cards shows edit spacey streaming

Topic 12:

Mean Netflix Posts: 0.026112253457772617

year spacey streaming watching just don edit house new cards shows like movie time watch

Topic 13:

Mean Netflix Posts: 0.026112253431153098

watch watching movie year just time like house cards streaming spacey don shows new edit

Topic 14:

Mean Netflix Posts: 0.02611225355140966

watch movie just watching year spacey don streaming cards like edit time shows house new

Topic 15:

Mean Netflix Posts: 0.026112253481739966

time year watching watch movie edit don new streaming shows just house spacey cards like

Topic 16:

Mean Netflix Posts: 0.02611225343450238

cards watch movie shows don new like house spacey watching edit streaming just year time

Topic 17:

Mean Netflix Posts: 0.026112253434123964

shows watching don edit streaming cards time house new like watch year movie just spacey

Topic 18:

Mean Netflix Posts: 0.2595323907706954

just streaming shows like don watching edit time spacey year cards new movie watch house

Topic 19:

Mean Netflix Posts: 0.026112253430223952

just shows time house year new movie cards like watch streaming edit don spacey watching

REFERENCES:

Latent Dirichlet allocation. (2019, January 26). Retrieved March 03, 2019, from <https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation>

Netflix. Retrieved March 04, 2019, from <https://www.reddit.com/r/netflix/top/?t=month>

PRAW: The Python Reddit API Wrapper¶. (n.d.). Retrieved March 02, 2019, from <https://praw.readthedocs.io/en/latest/>