**Quantifying COVID-19 Content in the Online Health Opinion War Using Machine Learning**

Due to COVID-19 whole world stuck at one place which causes huge amount of losses in almost all fields and some peoples are utilizing this opportunity to spread misinformation about the vaccine and COVID-19. In this community two types of peoples are working such as anti-vaccine and pro-vaccine where pro-vaccine peoples try to spread true information about COVID-19 and vaccines and anti-vaccine peoples will always spread false information and discourage peoples from using vaccines and this peoples always hinder in the path of other vaccines such as measles.

To quantify such PRO and ANTI information author is utilizing Facebook public posts where users will post their view on COVID-19 and vaccines. To identify whether post is related to ANTI or PRO author is using various techniques and algorithm such as

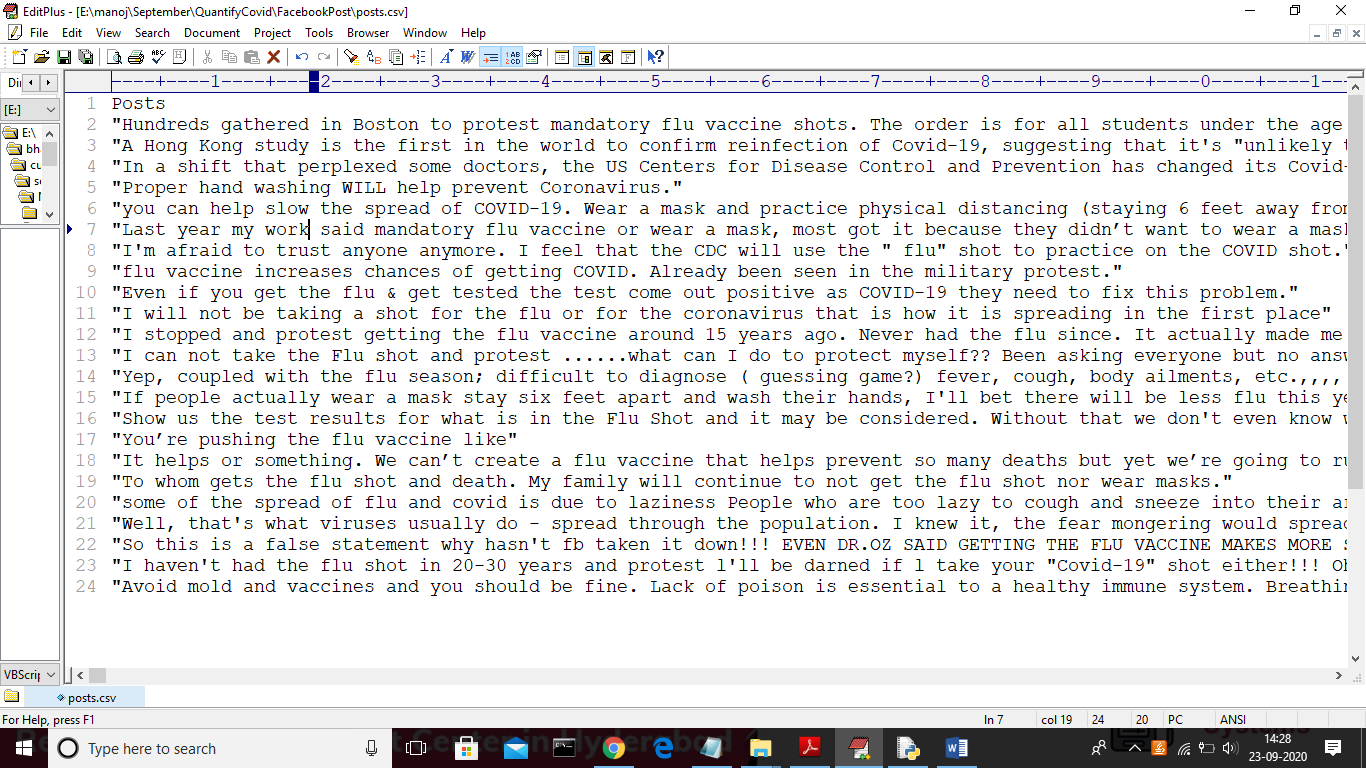
NLTK: natural language processing tool kit to remove stop words, URLS and special symbols from POSTS

Gensim API: using this API we will extract all text from post and then applying LDA (latent dirichlet allocation) algorithm from GENSIM API to extract all topics from Facebook POSTS.

Coherence metrics: using this metric we will check whether topic is related to ANTI or PRO. If more number of ANTI topics encountered then we can say peoples are talking in ANTI vaccine mode else PRO vaccine mode. Based on type we will create cluster to hold data of ANTI or PRO posts.

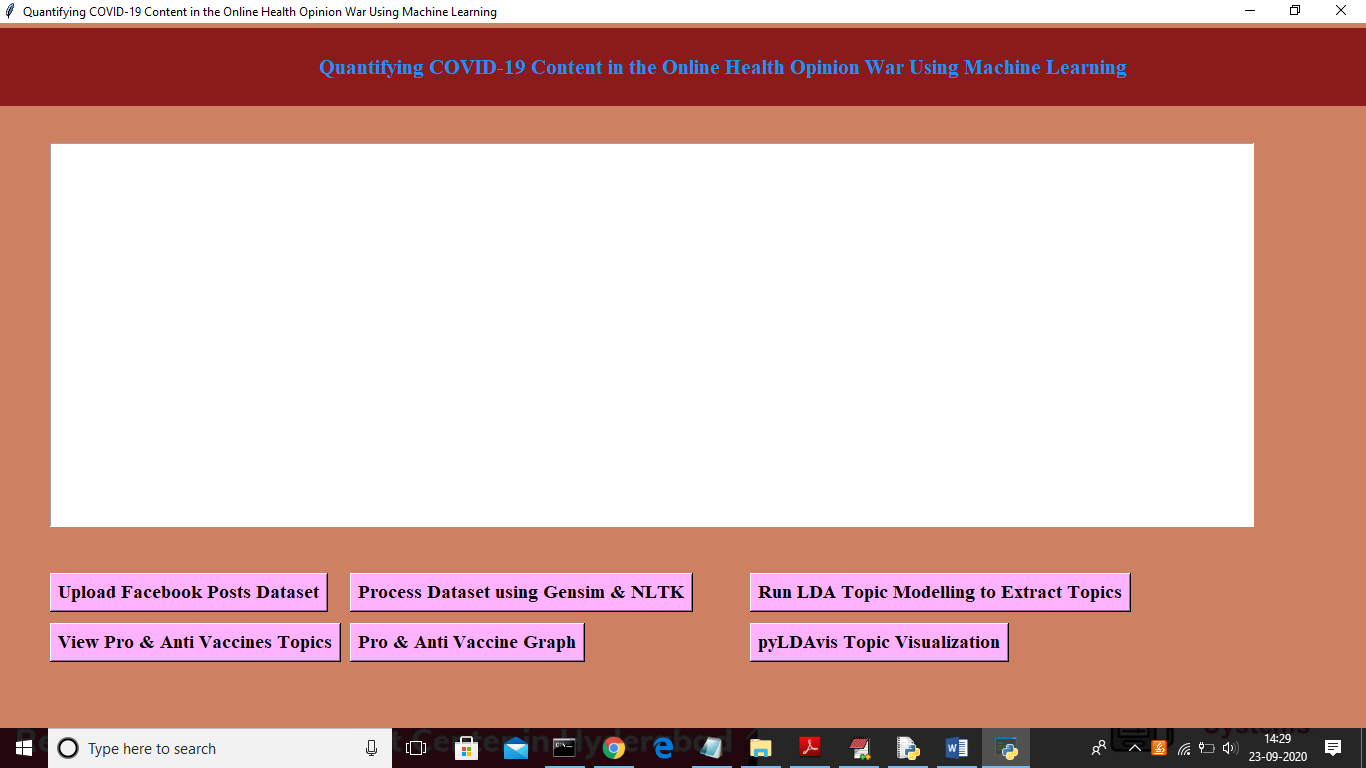
To implement this project we have extracted POSTS from Facebook and then apply above techniques to quantify COVID-19 ANTI or PRO vaccines cluster.

Below is the screen shots of Facebook posts

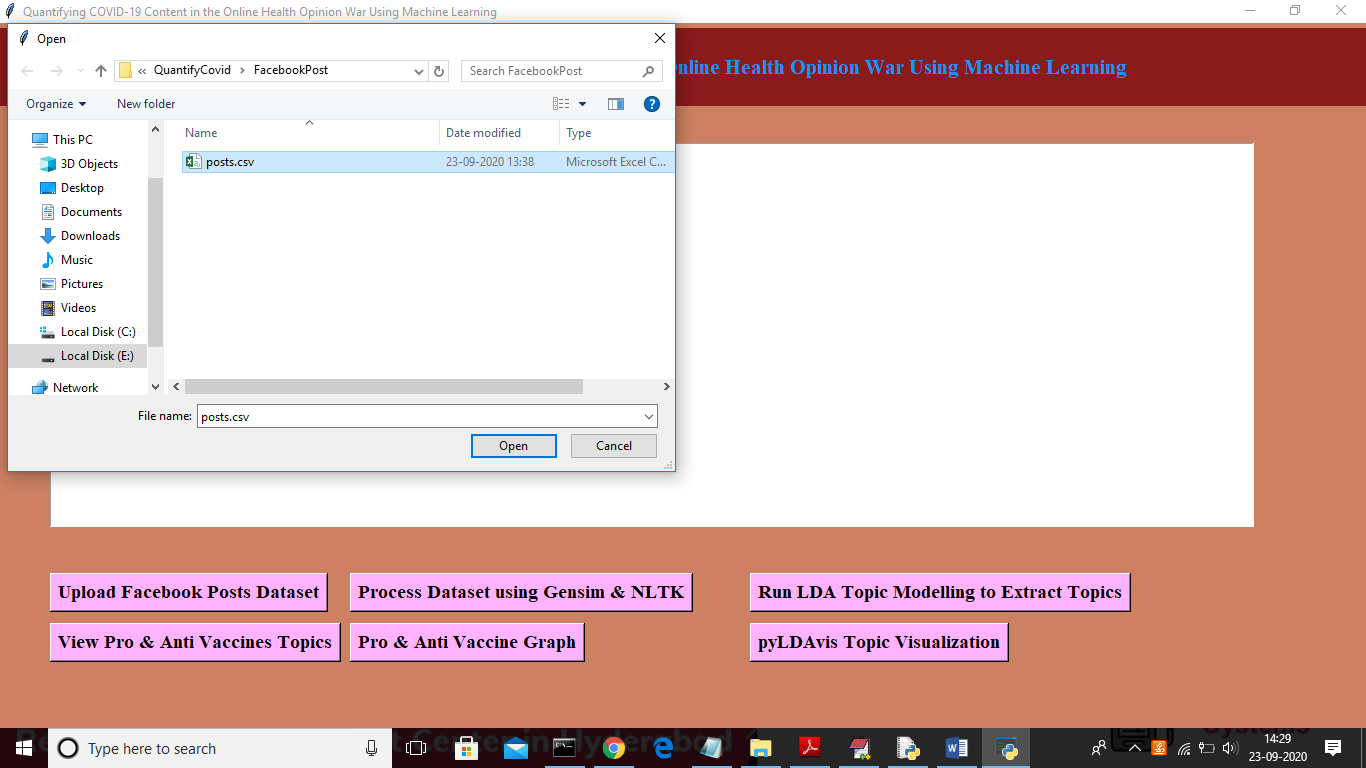


Screen shots

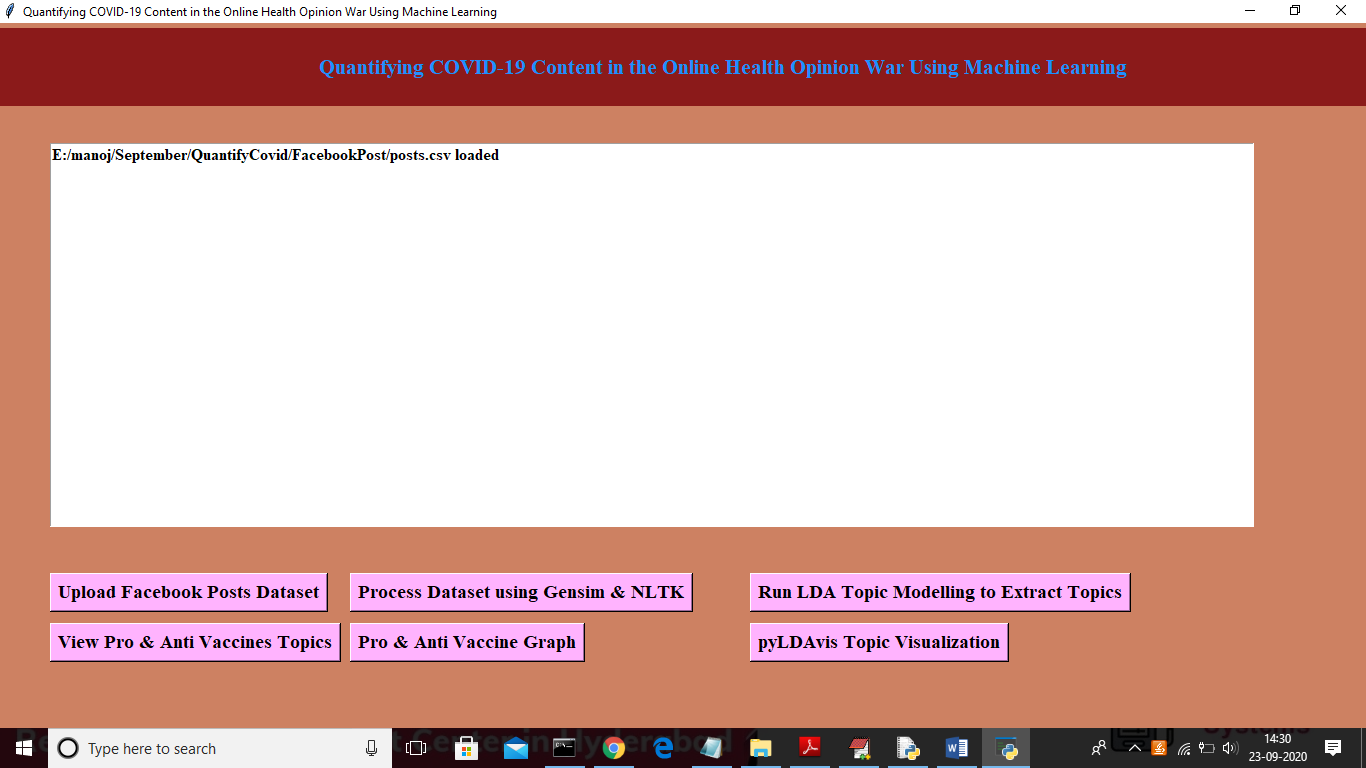
To run project double click on ‘run.bat’ file to get below screen



In above screen click on ‘Upload Facebook Posts Dataset’ button to load dataset



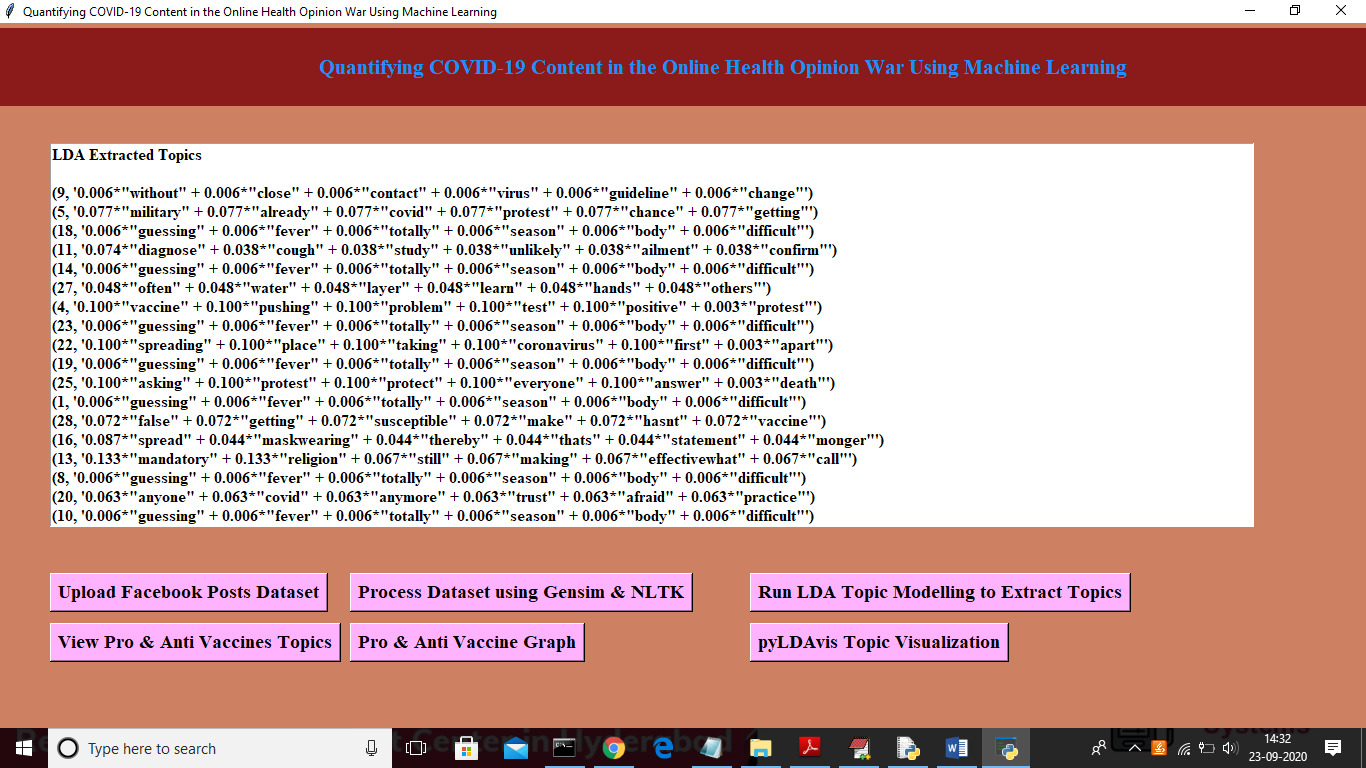
In above screen uploading ‘Posts.csv’ file and then click on ‘Open’ button to load dataset and after loading dataset will get below screen



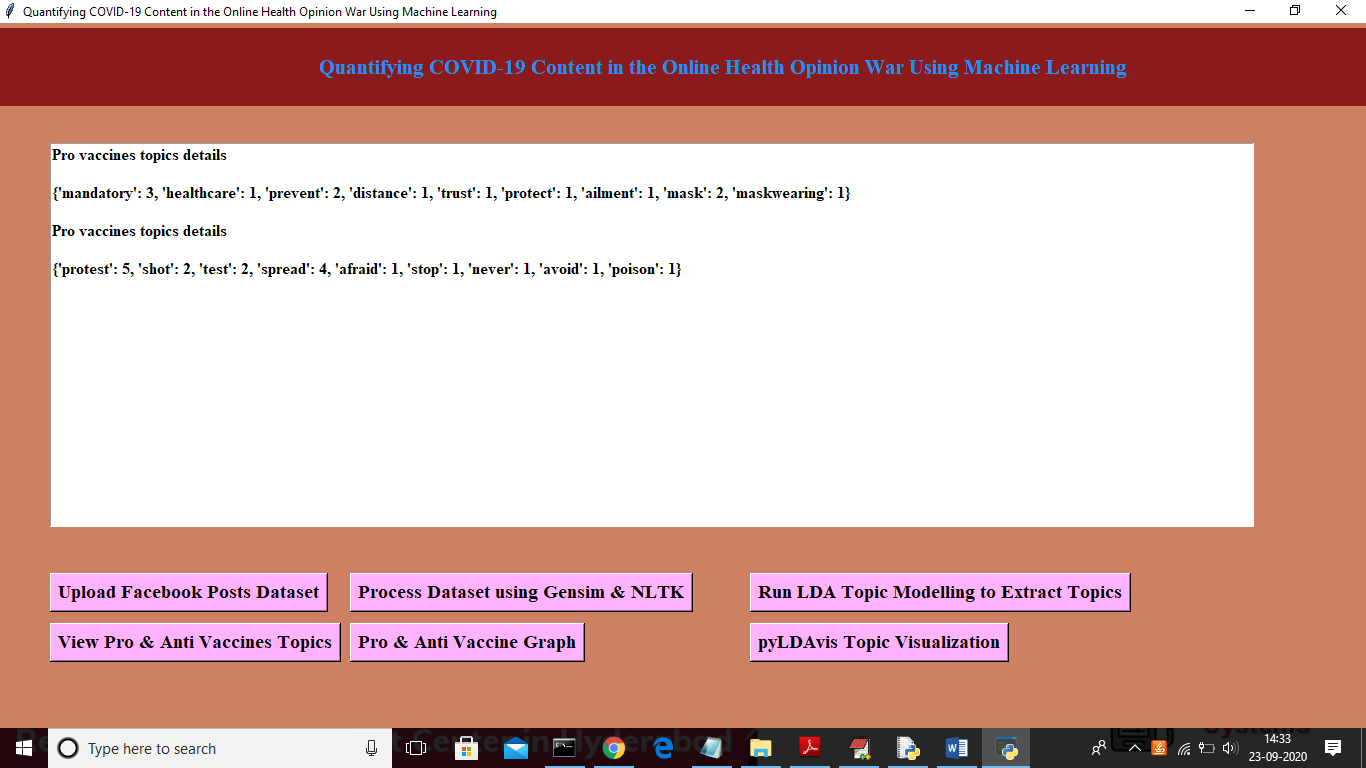
Now click on ‘Process Dataset using Gensim & NLTK’ button to read dataset and to clean dataset



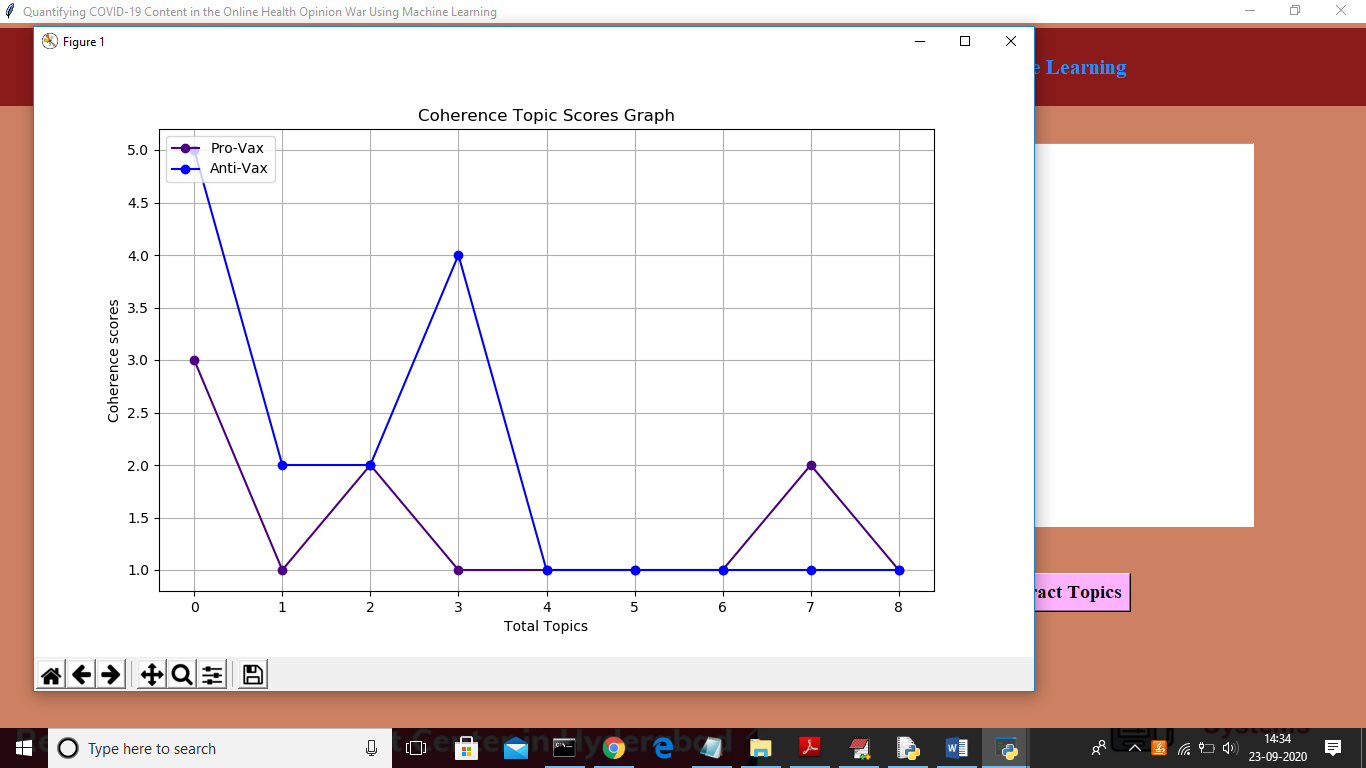
In above screen after cleaning will get above text from all posts and now click on ‘Run LDA Topic Modelling to Extract Topics’ button to extract topics from all posts



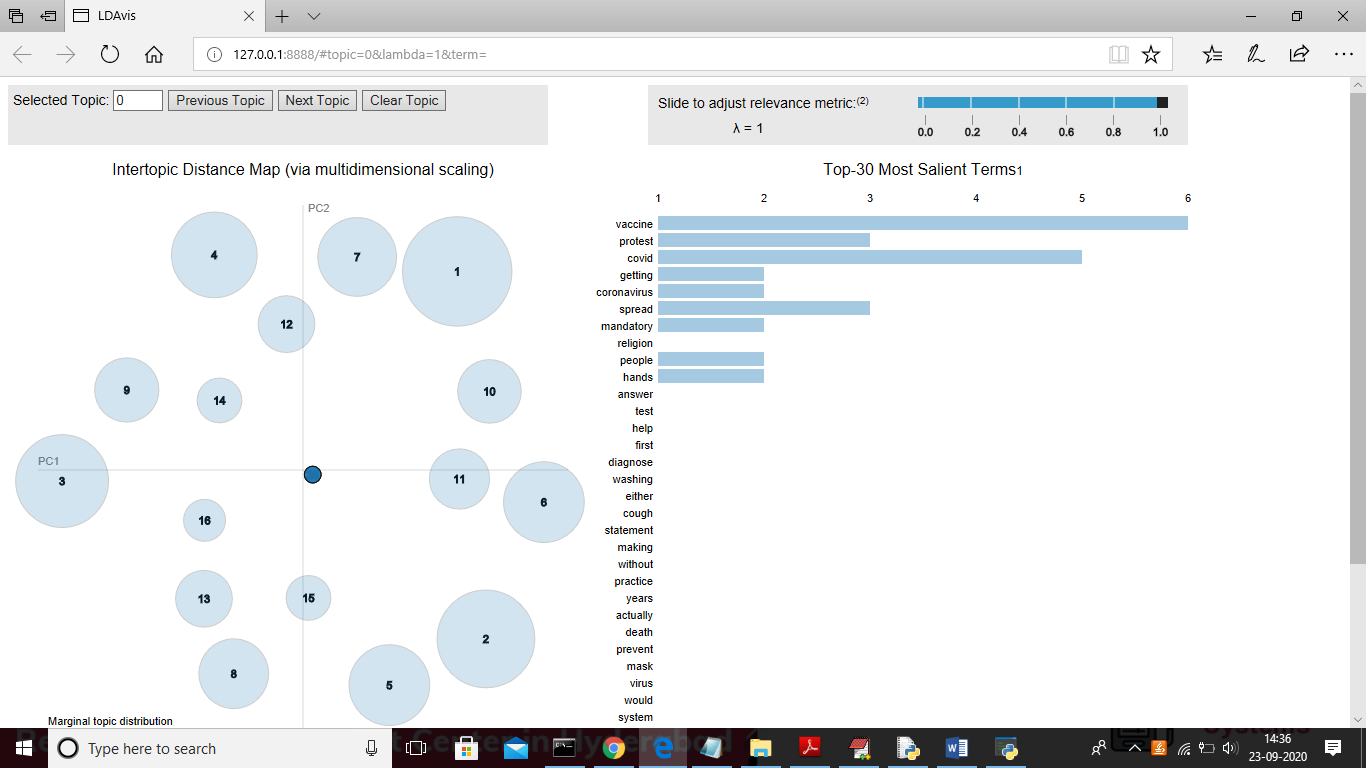
In above screen we can see TOPIC and its probability score for each word from POSTS and now click on ‘View Pro & Anti Vaccines Topics’ button to view all topics with coherence score



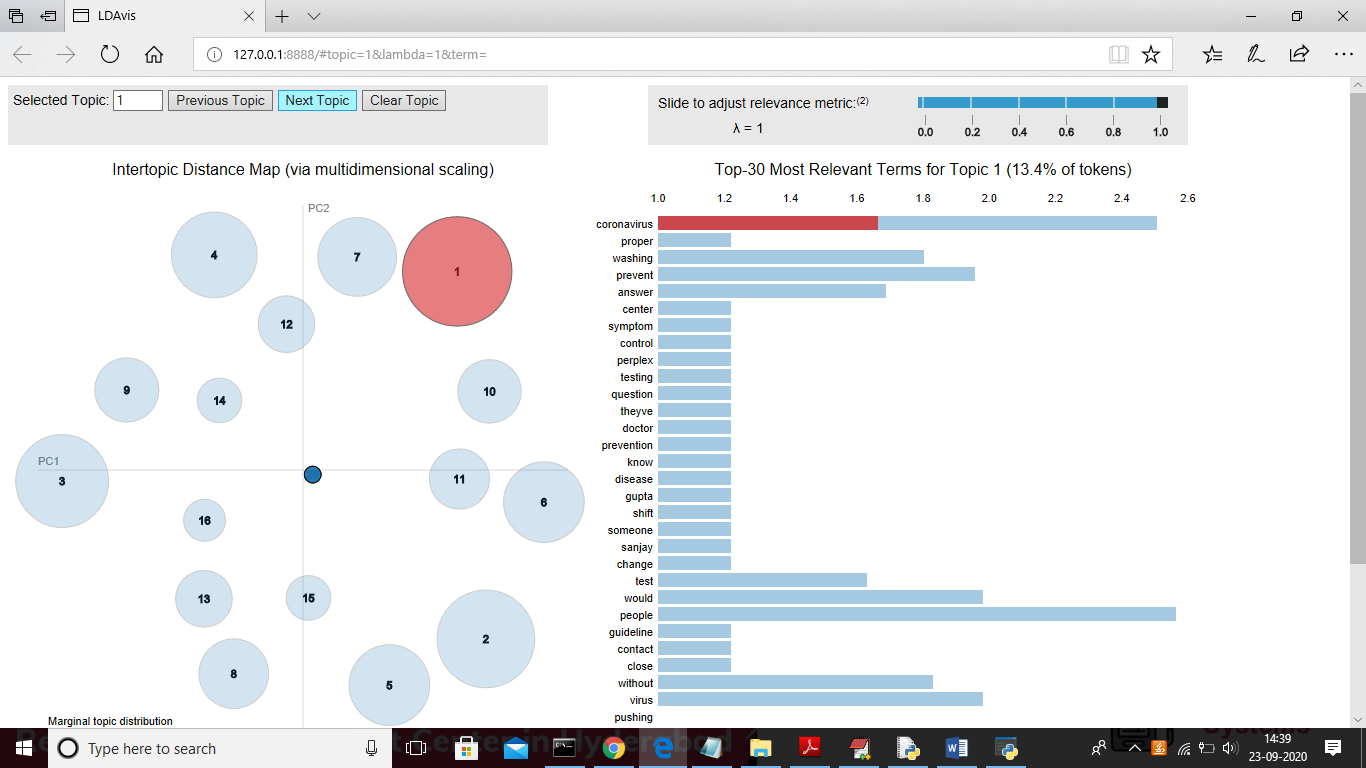
In above screen we can see all topics of PRO and ANTI with count value and now click on ‘Pro & ANTI Vaccine Graph’ button to get below graph and to quantify number of peoples are in favour of PRO or ANTI



In above graph x-axis represents number of topics and y-axis represents coherence score and in above graph blue line refers to ANTI and indigo colour line refers to PRO vaccine and from above graph we can conclude more peoples are discussing ANTI topics about vaccine. Now click on ‘pyLDAvis Topic Visualization’ to get visualization of all topics in browser



In above graph large circle refers that topic occurs more number of time and if occur less number of times then its circle will be in small size. In right size we can see all topics from that posts and u can click on ‘Next Topic’ button from top side of window to get next topic visualization. In above graph each circle represents 1 topic



In above screen for topic 1 coronavirus topic appear more times in all posts. Above graph u can see in browser only