****

CSC - 501

REPORT (ASSIGNMENT – 2)

SPATIO-TEMPORAL DATA

**SUBMITTED BY - GROUP B**

|  |  |  |
| --- | --- | --- |
| Kapoor, Nikita |  |  |
| Malik, Mona |  | V00935224 |
| Pattamatta, Suseela |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| Vemula, Sannath Reddy |  | V00949217 |

**Submitted To -** Prof. Sean Chester

[schester@uvic.ca](mailto:schester@uvic.ca)

**SECTION - 2**

**Visualization**

**INSIGHT #1 :** **Number of Tweet from different Troll Account and its variation over the time**

**INSIGHT #2 :** **Left Troll Vs Right Troll**

**2.1 Number of Tweet by Right Troll on daily Basis for the year 2016, 2017,2018**

**2.1 Number of Tweet by Light Troll on daily Basis for the year 2016, 2017,2018**

**2.3 Popular Hashtag’s used by Right Troll**

**2.4 Popular Hashtag’s used by Light Troll**

**2.5 Popular Retweets by Right Troll**

**2.6 Popular Retweets by Left Troll**

right Troll and Left Troll are the meat of the agency’s trolling campaign. Right Trolls behave like “bread-and-butter MAGA Americans, only all they do is talk about politics all day long,” Linvill said. Left Trolls often adopt the personae of Black Lives Matter activists, typically expressing support for Bernie Sanders and derision for Hillary Clinton, along with “clearly trying to divide the Democratic Party and lower voter turnout.

**Topic Modelling:**

LDA model(Latent Dirichlet allocation) is built with 5 different topics where each topic is a combination of keywords and each keyword contributes a certain weightage to the topic.

In the Visualization each bubble on the left-hand side plot represents a topic. The larger the bubble, the more prevalent is that topic. Right-hand shows the salient keywords that form the selected topic and their weightage.

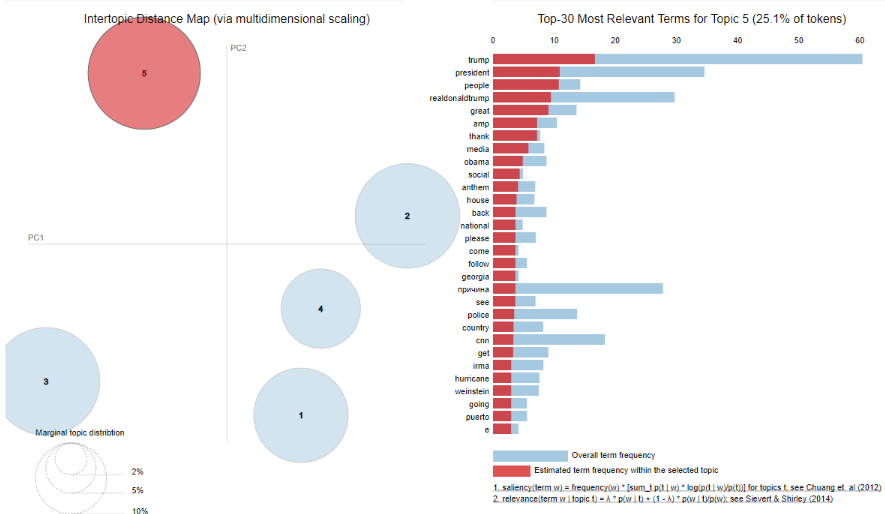
Model perplexity and [topic coherence](https://rare-technologies.com/what-is-topic-coherence/) provide a convenient measure to judge how good a given topic model is.

Perplexity: -7.870451016183864

Coherence Score: 0.5757121463422921

Cons – This model takes a lot of time to build and visualize. The work is done only on 1000 tweets to get the computational measures.

Reference - <https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/>

I

For only 1000 tweets

