**The data incubator proposal:**

**‘#Metoo’ movement and the evolution of social perception on sexual assault**

**Uyen Nguyen**

**Project idea**

On October 17, actress Alyssa Milano posted a tweet with hashtag *Me too,* following the accusation of Harvey Weinstein on sexual assaults by Hollywood actresses. Following that, the social media has been flooded with millions of #Metoo and stories by victims of sexual abuse all over the world.

This problem has been existing in all societies around the world for a long time. The Harvey Weinstein incident also raised questions on why his victims did not report him any time sooner, even when they already gain success. Does social perception on sexual assault shapes how victims respond when things happen to them?

In this project, I proposed to use text analytics tools to explore the topics on how social perception on sexual harassment evolved through time.

**Method:**

1-Web scraping of public resources including posts with #Metoo on Twitter and Facebook, and newspaper articles reported on sexual violations since 1990.

2- Text analytics including high frequency words, bigram, and sentiment analysis

3- Topic modeling using non-negative metric factorization (NMF) to classify news articles body into topics to see if we can use them to gain insights social perceptions.

4-Along with text analysis, additional analysis on crimes related to sexual assaults data might give us insights into how they change through time, what are the

**Preliminary results**

Plot 1: By analyzing about 5000 tweets, downloaded by Twitter rested API method, I generated a wordcloud of 20 most frequent words and bigrams.

[(('story', 'men'), 254), (('sexual', 'harassment'), 239), (('men', 'womeni'), 228), (('sexual', 'assault'), 151), (('elizabeth', 'warren'), 130), (('clueless', 'men'), 125), (('warren', 'joins'), 110), (('joins', 'campaign:'), 110), (('campaign:', 'portrays'), 109), (('portrays', 'professor'), 108), (('sexual', 'predator'), 105), (('professor', 'praised'), 105), (('sen.', 'elizabeth'), 95), (('womeni', 'was\xe2\x80\xa6'), 88), (('praised', 'sexual'), 77), (('must', 'challenge'), 65), (('men', 'sexual'), 65), (('stay.', 'must'), 63), (('challenge', 'men'), 61), (('opinion', 'clueless'), 54)]

Plot 2: about 1400 articles downloaded using the New York Times API method using ‘sexual’ as search query. Those articles were then parsed, extracted for specific information, and saved into a csv file on my github. The

https://plot.ly/create/?fid=unguyen:5

**Broader Significances**

Sexual assault is a serious issue that needed to be addressed and cured. Insights gained from this project can be applied to raise social awareness on sexual assaults and to build educational models and effective preventive tools. The inspiration of this project came from a group call No Means No Worldwide in Kenya who teaches girls self-defense and educates boys on rape is wrong (3).

**Reference sources:**

(1) <https://www.washingtonpost.com/news/morning-mix/wp/2017/10/16/me-too-alyssa-milano-urged-assault-victims-to-tweet-in-solidarity-the-response-was-massive/?utm_term=.5b42504801d>7

(3) http://www.upworthy.com/kenyas-unique-approach-to-rape-prevention-should-have-the-rest-of-the-world-taking-note