#MeToo Movement

Vaishnavi Srinivasan, Umang Mehta and Ahmad Al Marzook

Abstract—**Social media, especially Twitter, in the last decade, has given voices to the social movement with hashtags such as #ArabSpring, #BlackLivesMatter #LoveWins, #JeSuisCharlie and revolutionized how we approach social issues. In this project, we analyze #MeToo tweets to predict real-world outcomes as a direct consequence of sexual harassment that is plaguing the society. As the movement works towards changing perceptions and responses of sexual harassment, the tweets generated to support the movement are enormous with mixed opinions. Analyzing the #MeToo tweets, we found patterns that summarizes the viewpoint over time - negative polarity during an allegation and positive because of actions taken against offenders, along with fluctuating subjectivity which can be reflection of allegations and countenance. The change in subjectivity over time indicates the gravity of the situation and how much people are affected.**

Everyday a lot information is shared through twitter. Analyzing the tweets of this movement, we also intend to discover patterns. By exploring several tweets features we intend to understand what makes a tweet more widely shared than the others. We have gathered the tweets from October 15th, 2017 to March 6th, 2018. We made an explorative data analysis to uncover that tweets text have strong relationships with retweetability.

***Index Terms*— retweet; tweet; follower; social media; social movements, metoo, harassment**

I. Introduction

SOCIAL media has become a popular platform for people to voice their views and information that they consider important. It has become a stronghold for people to speak out their conundrums to a wider audience and seek help, or change. For #MeToo movement, this has brought to limelight the way sexual harassment has been going in the workplace for many years against primarily women. What began as a revelation of sexual harassment acts performed by noted celebrities in the Hollywood industry, has rapidly spread across other work places. With hashtags, such as #גמאנחנו, #QuellaVoltaChe, #وأنا كمان, #stilleforopptak, this movement has spread globally and is being used by women to come together on Twitter.

This movement which was ongoing for years, gained popularity with Harvey Weinstein's acts of sexual misconduct revelations and actress Alyssa Milano hashtag encouraging women to tweet their experiences. It serves to raise awareness and has been used for bringing to justice many offenders. Since the introduction of the hashtag, there have been constant tweets by users with intermittent spike following a revelation of an offender or when an offender has been penalized. With tweets classified using a lexicon, we intend to derive the subjectivity and polarity of the sentiments.

We begin the study by extracting tweets from 15th October 2017, when the hashtag gained momentum, to 6th March 2018 and perform data cleansing. For sentiment classification, we segregated the tweets into bi-weekly bins. Then we utilize three lexicons to determine the one that performs a better job of predicting the sentiment. We then calculate the subjectivity and polarity as defined in the paper "Predicting the Future with Social Media" by Sitaram Asur and Bernardo A. Huberman, The Barnaghi, P., Ghaffari, P., & Breslin, J. G. (2016). We then plot to see how the sentiments and polarity have changed over time and the real-world scenario behind such an outcome. Then group the tweets by users to understand the features of tweets, using the methods described in "Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network" by Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H. Chi and "Predicting the Political Alignment of Twitter Users" by Michael D. Conover, Bruno Gonc¸alves, Jacob Ratkiewicz, Alessandro Flammini and Filippo Menczer, that make it popular.

We discuss the problems encountered in predicting sentiment of the tweet and its popularity, any bias or noise introduced and the factors that influence them. We show that Vader lexicon perform better at predicting a tweets sentiment. For determining retweet rate, a classification problem, we consider the text of the tweets, and user features as described in "Predicting Popular Messages in Twitter" by Liangjie Hong Ovidiu Dan Brian D. Davison. We find that the text has better chance of being predicting retweet rate.

II. Dataset Description

We collected two datasets for the analysis. As part of content analysis for sentiment classification we gathered tweets scraping the web and collecting approximately 2 million tweets. For the tweet prediction, we have collected 75 thousand tweets using the Twitter API with user features.

Tweets are characterized as short messages with a limit of 140 characters (increased to 280 in September 2017) that serve to provide status update of a user. This length constraint has influenced the use of hashtags, retweets and mentions. They serve to group common messages for an intended audience with a specific topic.

A. 2 Million Dataset

The twitter API has several limitations with regards to tweet collection - the number of requests it can accept and the timeline of tweets retrieved. To bypass such limitations, we have used the GetOldTweets-python, a project developed by

Jefferson-Henrique and modified by David to scrape the web and collect the tweets containing the #MeToo. We collected a total of 2,035,518 tweets between the dates of 15th October 2017 and 6th March 2018. The distribution of tweets by date is shown in histogram in Fig. 1

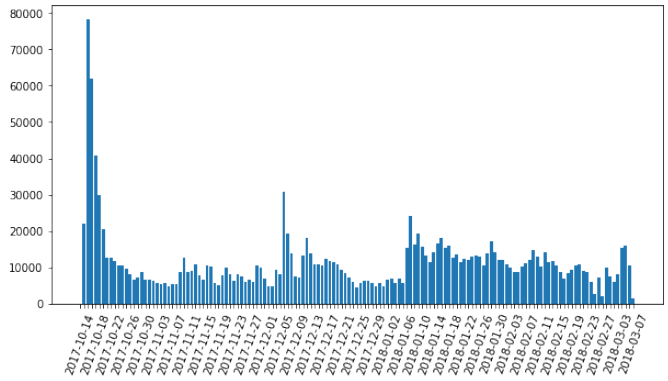

Fig. 1. Tweets count histogram starting from October 15, 2017

B. *Data Pre-processing*

For sentiment classification, we have performed the below pre-processing.

- Removed hashtags, mentions, url and 'rt'
- Removed the stop-words
- Added parts of speech tagging
- Stemmed the words and lemmatized the words

Many users, spam twitter by sending out the same tweets multiple times. We have removed duplicates in tweets to remove such a noise.

C. *75K dataset*

To perform exploratory data analysis in understanding the tweet features that are associated with retweeting we have used Twitter API to collect 75 thousand tweets on 26th April 2018 containing the #MeToo.

D. *Tweet Features*

For the 75K dataset, we extracted the features listed below to build a retweet model. The features are concerned with text of the tweet and the tweet's user.

- Retweet – number of times a tweet has been retweeted
- Follower – users following tweet author
- Friends – users who are following a tweet author
- CreatedSince –days since account is active
- Statuses – tweets tweeted by an author
- Favorite – tweets favorited by an author

III. Literature Review

We perform sentiment classification using a Lexicon and interpret its score to categorize a tweet to positive, negative and neutral. We have used three Lexicons to determine which one works for our dataset. We have used Vader, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text" by Hutto, C.J. & Gilbert, E.E. (2014), AFinn, developed by Finn Årup Nielsen, "A new ANEW: evaluation of a word list for sentiment analysis in microblogs", and TextBlob, a Python (2 and 3) library for processing textual data developed by Steven Loria.

"Social media and social movements: Facebook and an online Guatemalan justice movement that moved offline" by Summer Harlow (2012) and "Tweets and the streets: social media and contemporary activism" PAOLO GERBAUDO are some of the articles we used for conceptual studies. The Guatemalan justice movement, just like the current sexual harassment movement moved from the world of online chatter to real world consequences.

We are expecting the initial sentiments of the sexual harassment to be negative, an expression of several factors such as post-traumatic stress or depression suffered by the victims, flippant attitude and bullying of the victims, or lax or irresponsible attitude expressed by media or law enforcement body. A transition towards positive sentiment is expected as more transformations and several reforms are expected by governing bodies which can be attributed to influencers and supporters, who are responsible for actionable results.

Then, the subjectivity and polarity are visualized using a bi-weekly bin, we have used the methodology defined in "Predicting the Future with Social Media" by Sitaram Asur and Bernardo A. Huberman, The Barnaghi, P., Ghaffari, P., & Breslin, J. G. (2016).

IV. Experiment Design And Method

We first look at the classification performed by various lexicons, then move on to calculating subjectivity and polarity.

A. *Sentiment Analysis*

The preprocessed tweets are classified using each lexicon. All the lexicons work on a unigram data. TextBlob performs pattern analyzer, while Afinn ranks each work based on a dictionary and Vader is built on tweet corpus and uses context. All the classifiers derive a compound score. A score of greater than 0 is labelled 'Positive', less than zero 'Negative' and equal to 0 as 'Neutral'.

Random 50 tweets are then labelled manually and compared with the classifiers output. The Vader classifier gives a pretty good accuracy as compared to Afinn and TextBlob. Results are displayed in Table I.

TABLE I
CLASSIFIER SCORE ANALYSIS

	Vader	Afinn	TextBlob
Labels Match with Manual Label	37	28	38
Accuracy	74	56	78

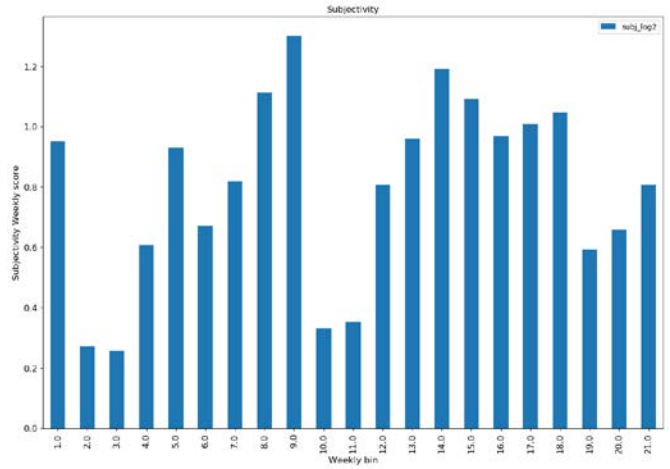

Fig. 2. Subjectivity Values in continuous bi-weekly bins

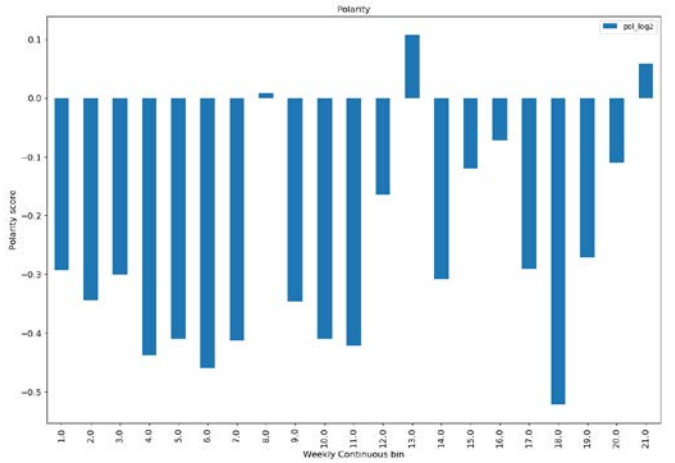

Fig. 3. Polarity Values in continuous bi-weekly bins

B. Subjectivity

The labelled tweets are split into bins containing biweekly data. The bins are made continuous by using dates from previous week and current week to determine the gradual changed in subjectivity.

Using the formula defined in "Predicting the Future with Social Media" by Sitaram Asur and Bernardo A. Huberman, The Barnaghi, P., Ghaffari, P., & Breslin, J. G. (2016).

We have captured as follows.

$$\text{Subjectivity} = \frac{|\text{Positive and Negative Tweets}|}{|\text{Neutral Tweets}|} \quad (1)$$

Fig. 2 shows the subjectivity spike in October 15 when #MeToo was launched. The next spike starts during the week of November 5th and continues to week of November 12th. During these weeks Brett Ratner and Kevin Spacey came under fire for sexual harassment allegations. November 12th week was also the week when a march was conducted for this movement in LA. Twitter was again abuzz when numerous sexual harassment revelations took place against SNL staff and Donald Trump during late December that continued into January. March have had some additional offenders thrown into limelight.

C. Polarity

The labelled tweets are split into bins containing biweekly data. The bins are made continuous by using dates from previous week and current week to determine the gradual changed in polarity.

Using the formula defined in "Predicting the Future with Social Media" by Sitaram Asur and Bernardo A. Huberman, The Barnaghi, P., Ghaffari, P., & Breslin, J. G. (2016).

We have captured polarity as follows,

$$\text{PNratio} = \frac{|\text{Tweets with Positive Sentiment}|}{|\text{Tweets with Negative Sentiment}|} \quad (2)$$

Fig. 3 shows the polarity to be generally negative. From various revelations to acceptance and penalizing are reflected in tweets.

The data included has some noise in it. The tweets extracted have multiple languages in them. These tweets are categorized as Neutral by all the lexicons. To further understand the score assignment by Vader, we plot the density distribution along with box plot of Vader compound scores for each bin.

Looking at the Fig. 4 and 5, we find that most of the tweets classified as positive and negative have scores between 0.25 and -0.25. When we further analyzed the text, we found tweets such as "I felt the same when one of the true life monsters

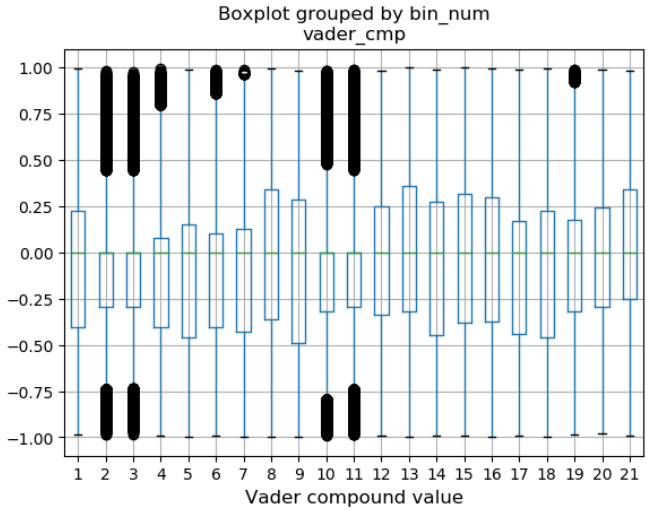

Fig. 4. Boxplot of Vader compound score

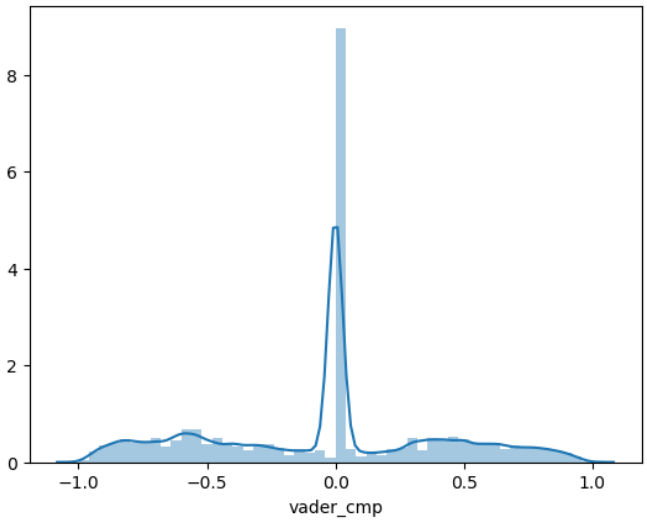

Fig. 5. Density plot of Vader compound score

from my childhood was killed. The only relief from his death I had was that kids were finally safe from his evil touch. Not wishing ill on your monster seems to mean that he didn't turn you into the like. # MeToo" was classified as "Negative" by Vader. As most of the words and sentences have negative connotation, the lexicon classifies it so. But we know that the user is happy in this context and it should be categorized as “Positive”. These discrepancies in labelling can cause the polarity to swing to extremes even though they are representing an opposite sentiment. Thus, we need a manual labelling for training a classifier or a lexicon built for harassment domain.

V. Model Evaluation Using SemEval Dataset

To evaluate the model performance, we used the SemEval-2017 Task 4 Dataset and see how the model performs on this manually labelled dataset. We used the data subset for Subtask A because the data for other subtasks were conditioned on topic. The raw dataset on the official website only consisted of tweet IDs and sentiment label. So, we used the Twitter API to extract these tweets. While extracting out of 50334 records, tweets were available for only 37911 records. After performing the text preprocessing step 37266 records had distinct text and hence we considered only these tweets for classification. Following is the true distribution of sentiment classes in the dataset:

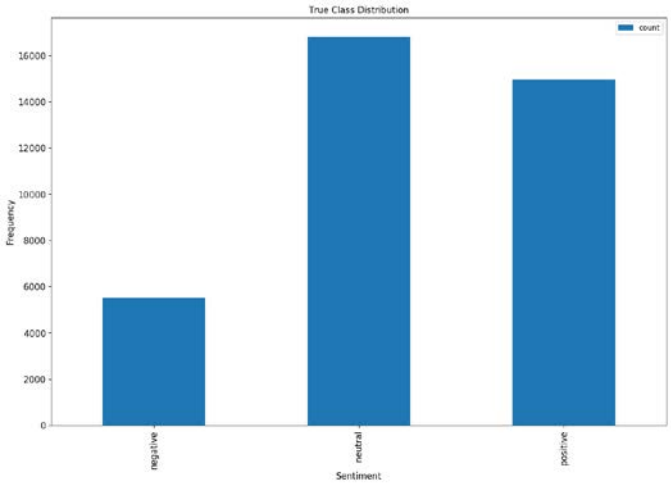

Fig. 3. True Distribution of Sentiments Classes in SemEval-2017 Task 4 Subtask A Dataset

We evaluated the results by AFinn, Text Blob and Vader with 3 metrics namely Accuracy, F1-Score with Weighted average and Confusion Matrix. The distributions for the class labels after each classification of each method are as follows:

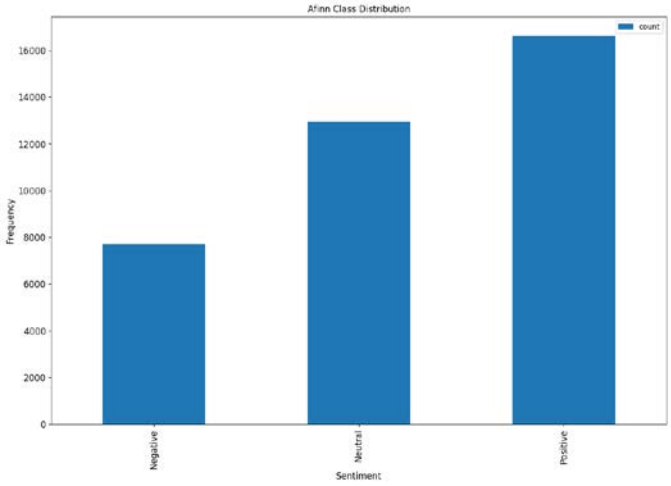

Fig. 3. Distribution of Sentiments Classes Using Afinn

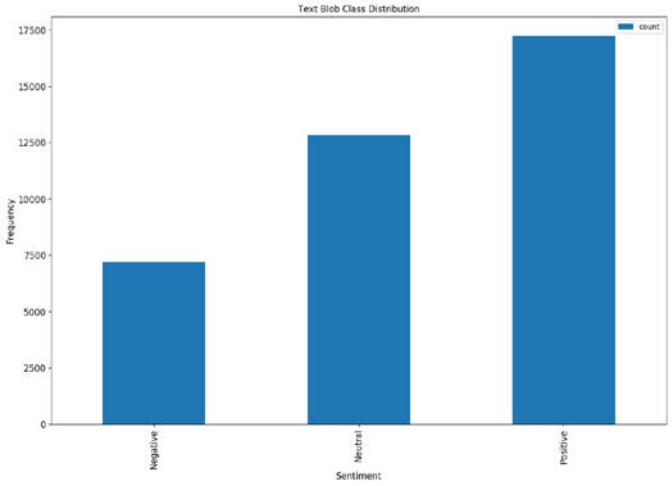

Fig. 3. True Distribution of Sentiments Using Text Blob

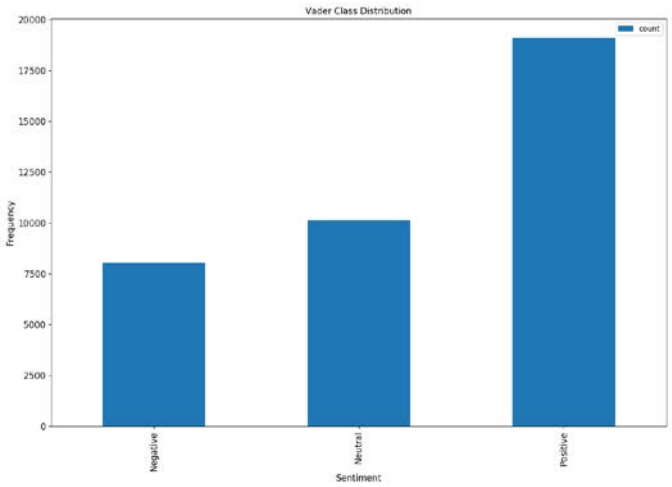

Fig. 3. True Distribution of Sentiments Classes Using Vader

We see that the distribution of classes is similar for Afinn and Text Blob. But according to the Accuracy and F1-Scores, AFinn performs the best and Text Blob performs the worst of the three methods as shown below.

CLASSIFIER	ACCURACY	F1-SCORE
AFINN	0.5505	0.551
TEXT BLOB	0.5051	0.5057
VADER	0.539	0.5319

TABLE II. Performance Metrics for the classifiers

	POSITIVE	NEUTRAL	NEGATIVE
POSITIVE	9683	3800	1466
NEUTRAL	5647	7866	3283
NEGATIVE	1282	1273	2966

TABLE III. Confusion Matrix for AFnn

	POSITIVE	NEUTRAL	NEGATIVE
POSITIVE	9404	3796	1749
NEUTRAL	6205	7279	3312
NEGATIVE	1621	1761	2139

TABLE IV. Confusion Matrix for Text Blob

	POSITIVE	NEUTRAL	NEGATIVE
POSITIVE	10559	2843	1547
NEUTRAL	6947	6438	3411
NEGATIVE	1591	840	3090

TABLE V. Confusion Matrix for Vader

As we can see these methods perform very well on the Positive Tweets but perform extremely poor on the Neutral or Negative Tweets. Most of the misclassification of the Neutral Tweets is towards the Positive side and hence it suggests that these methods are biased towards Positive sentiment.

When we compare the results of the Vader classifier for SemEval data with the manual evaluation for MeToo tweets, it suggests that a model trained on SemEval dataset may not be generalized for all domains as there is a stark huge difference between the accuracies obtained. But for the accuracy calculation on the MeToo data we have only considered 50 random tweets and hence cannot state this with a high confidence.

VI. Popularity of Tweet

Now we use the 10K dataset to analyze the features that make a tweet popular. The initial hypothesis was that more than the content of the features, it would be the number of followers a user has that would make the tweet popular.

We have used the methods described by "Predicting Popular Messages in Twitter" by Liangjie Hong Ovidiu Dan Brian D. Davison and "Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network" by Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H. Chi for our analysis.

For this analysis, we have taken the text of the tweet and grouped them by user. We then bin the retweet into 4 classes: 0: not retweeted, 1: retweet count between 1 to 100, 2: retweet count between 100 to 1000 and 3: retweet count greater than 1000. We then use multiple features to determine the retweet bin for tweet posted.

First, we compute the term frequency-inverse document frequency (TF-IDF) for the users' tweet. Then we utilize latent dirichlet allocation (LDA) to obtain the topic distribution in the tweets. We also store the users' metadata information like status, favorites, followers, friends and the days since account created to predict the retweet bin.

We fit logistic regression model with 5-fold cross validation separately on each feature, TF-IDF, LDA and metadata, to determine the retweet rate. The TF-IDF had the highest accuracy of 60% followed by LDA with an accuracy of 51%. The metadata information has an accuracy of 48% same as using only followers count to predict retweet. The results are shown in Table II. We find that the tweets containing certain words and topics are more useful in predicting the retweet rate. The Fig. 10 shows that the number of messages with a minimum of a hundred retweets are most frequent than messages with 1000 retweets. Messages with retweets greater than 1000 are pretty low.

TABLE II
ACCURACY OF CLASSIFICATION TASK

Methods	Accuracy
TF-IDF	0.59
LDA	0.51
Metadata Information	0.48
Followers_Count	0.48

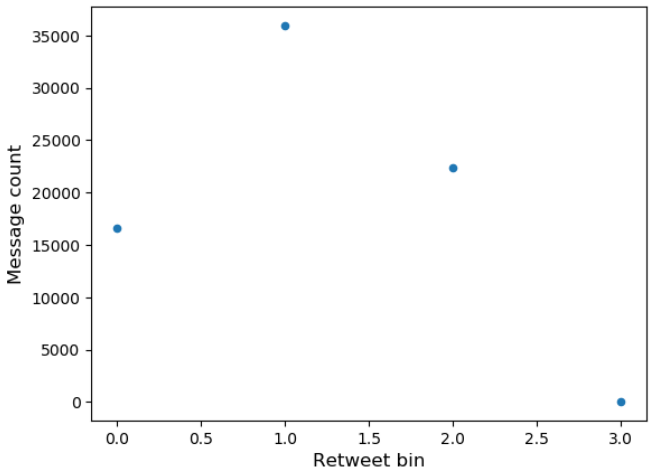

Fig. 10. Scatter plot of retweet bin against number of messages

We also need to note that the dataset has only 75K records from a particular day. While the content associated with #MeToo is of importance for predicting retweet rate, getting a bigger sample with random data may be more accurate prediction of the retweet rate.

VII. CONCLUSION

We have analyzed the #MeToo tweets for sentiment classification and predicting retweets. We first used Vader lexicon to classify the tweets as positive, negative and neutral. Then we visualized the subjectivity and polarity. We found that initially more of the tweets were negative as people were coming out with their experiences, hardships. Twitter was stormed with #MeToo as and when a revelation was made about a celebrity. The sentiments were more negative during these times. There are shifts to neutral but rarely there are positive polarity. The last six months have been the most controversial for sexual harassment. It has been shocking to see how many have used their power and position to extract favors from people taking advantage of the other persons helplessness. As in the papers, "Predicting Popular Messages in Twitter" by Liangjie Hong Ovidiu Dan Brian D. Davison and "Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network" by Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H. Chi, we find that #MeToo, the TF-IDF and LDA feature has been more prominent in predicting the retweet rate rather than followers count.

REFERENCES

[1] Gerbaudo, P., “Tweets and the streets: social media and contemporary activism.” London: Pluto Press. (2012)

[2] Lim, M. “Clicks, Cabs, and Coffee Houses: Social Media and Oppositional Movements in Egypt, 2004-2011”. Journal of Communication,62 (2), 231-248. (2012) doi:10.1111/j.1460-2466.2012.01628.x

[3] Harlow, S. “Social media and social movements: Facebook and an online Guatemalan justice movement that moved offline.” New Media & Society,14 (2), 225-243. (2012) doi:10.1177/1461444811410408

[4] Youmans, W. L., & York, J. C. “Social Media and the Activist Toolkit: User Agreements, Corporate Interests, and the Information Infrastructure of Modern Social Movements.” Journal of Communication,62 (2), 315-329. (2012) doi:10.1111/j.1460-2466.2012.01636.x

[5] Barnaghi, P., Ghaffari, P., & Breslin, J. G. “Opinion Mining and Sentiment Polarity on Twitter and Correlation between Events and Sentiment.” 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService). doi:10.1109/bigdataservice.2016.36

[6] Bifet, A., & Frank, E. “Sentiment Knowledge Discovery in Twitter Streaming Data.” Discovery Science Lecture Notes in Computer Science, 1-15. (2016) doi:10.1007/978-3-642-16184-1_1

[7] Sixto, J., Almeida, A., & López-De-Ipiña, D. “An Approach to Subjectivity Detection on Twitter Using the Structured Information.” Computational Collective Intelligence Lecture Notes in Computer Science, 121-130. (2016). doi:10.1007/978-3-319-45243-2_11

[8] Peetz, M., Rijke, M. D., & Kaptein, R. “Estimating reputation polarity on microblog posts.” Retrieved February 10, 2018, from http://dare.uva.nl/search?metis.record.id=476754

[9] Asur, S., & Huberman, B. A. “Predicting the Future with Social Media.” 2010. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology . doi:10.1109/wi-iat.2010.63

[10] Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H. Chi “Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network” IEEE International Conference on Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust 978-0-7695-4211-9/10 $26.00 © 2010 IEEEDOI 10.1109/SocialCom.2010.33 177

[11] Michael D. Conover, Bruno Gonc¸alves, Jacob Ratkiewicz, Alessandro Flammini and Filippo Menczer “Predicting the Political Alignment of Twitter Users”

[12] Liangjie Hong Ovidiu Dan Brian D. Davison "Predicting Popular Messages in Twitter". WWW 2011 – Poster

[13] Zongyang Ma, Aixin Sun, and Gao Cong "On Predicting the Popularity of Newly Emerging Hashtags in Twitter"

[14] Shirky, C. “The Political Power of Social Media.” (2016, January 21). Retrieved February 07, 2018, from https://www.foreignaffairs.com/articles/2010-12-20/political-power-social-media

[15] Sichynsky, T. “These 10 Twitter hashtags changed the way we talk about social issues.” (2016, March 21). Retrieved February 07, 2018, from https://www.washingtonpost.com/news/the-switch/wp/2016/03/21/these-are-the-10-most-influential-hashtags-in-honor-of-twitters-birthday/?utm_term=.0aaaa2fb9c3c

[16] Zaman, H. U. (n.d.). “#MeToo and the worldwide reckoning it brought in 2017”. Retrieved February 07, 2018, from https://www.geo.tv/latest/174883-how-2017-toppled-sexual-predators

[17] Finn Årup Nielsen, "A new ANEW: evaluation of a word list for sentiment analysis in microblogs", Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages. Volume 718 in CEUR Workshop Proceedings: 93-98. 2011 May. Matthew Rowe, Milan Stankovic, Aba-Sah Dadzie, Mariann Hardey (editors)

[18] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[19] Textblob Copyright 2013-2017 Steven Loria