Meta-Terrorism: Identifying Linguistic Patterns in Public Discourse After an Attack

Panos Kostakos
Center for Ubiquitous Computing
University of Oulu
Oulu, Finland
panos.kostakos@oulu.fi

Markus Nykänen
Center for Ubiquitous Computing
University of Oulu
Oulu, Finland
markus.nykanen@student.oulu.fi

Mikael Martinviita
Center for Ubiquitous Computing
University of Oulu
Oulu, Finland
mikael.martinviita@gmail.com

Abhinay Pandya
Center for Ubiquitous Computing
University of Oulu
Oulu, Finland
abhinay.pandya@oulu.fi

Mourad Oussalah
Center for Ubiquitous Computing
University of Oulu
Oulu, Finland
mourad.oussalah@oulu.fi

Abstract—**When a terror-related event occurs, there is a surge of traffic on social media comprising of informative messages, emotional outbursts, helpful safety tips, and rumors. It is important to understand the behavior manifested on social media sites to gain a better understanding of how to govern and manage in a time of crisis. We undertook a detailed study of Twitter during two recent terror-related events: the Manchester attacks and the Las Vegas shooting. We analyze the tweets during these periods using (a) sentiment analysis, (b) topic analysis, and (c) fake news detection. Our analysis demonstrates the spectrum of emotions evinced in reaction and the way those reactions spread over the event timeline. Also, with respect to topic analysis, we find "echo chambers", groups of people interested in similar aspects of the event. Encouraged by our results on these two event datasets, the paper seeks to enable a holistic analysis of social media messages in a time of crisis.**

Keywords—terrorism, social media, echo chambers, event summarization

I. INTRODUCTION

Terrorism is a subject that divides people and escalates the division of people into so-called "echo chambers". These chambers consist of people who share similar opinions and interests and who discuss similar topics and use same news sources. The short definition of an echo chamber is "[a] metaphorical description of a situation in which information, ideas, or beliefs are amplified or reinforced by communication and repetition inside a defined system" [1]. Identifying echo chambers can give us information about how many chambers there are and on what characteristics they are based. Echo chambers can then be examined in more detail to see how they develop over time, what different "roles" there are for the members of the chamber and how the chambers interact with each other.

To identify how these echo chambers are conducted and how they behave before and after massive events, we used sentiment analysis for people's tweets, topic analysis for shared URL sites, and fake news detection for those shared sites. These analyses give us a general indicator of the positive and negative opinions of these echo chambers based on the discussed topics and insights of shared news.

II. STATE OF THE ART

In this section two different studies are reviewed. The first one is closer to our work since it uses textual content for analysis, and the second one focuses more on actual, uunambiguous connections between entities.

One of the best-known incidents that has been studied for detecting and explaining social behavior patterns of people through time is the murder of Fusilier Lee Rigby in Woolwich, London in 2013. Researchers saw an opportunity for testing an experimental state-of-the-art program called Sentinel that has been supported by the European Commission [2, 3]. Data was collected from social media and filtered by a series of "channels" using relevant search terms. Connecting it to our work, the channels are considered to be similar to "echo chambers". The search terms were tuned over time, and finally there were three main channels: "Woolwich", "Rigby" and "EDL" (English Defense League). The monitoring of the channels was done with an open source tool called FlexiTerm [4] and high-level terms.

Another study that was more tightly tied to the actual definition of "echo chambers" investigated how content is exchanged between political groups [5]. The study used Twitter streaming API for collecting tweets; user mentions of major political entities were traversed to form directed links. With the help of link relationships, party groupings were formed. The Parlgov dataset, mentioned in the study, was employed for calculating ideological distances between groups. For actual presentation of the structures within networks, a Fruchterman-Reingold force-directed graph algorithm was used. The study hypothesize and confirmed that political groupings that are ideologically further apart interact less, and the extremists are more likely to form echo chambers. An interesting finding was that political parties may be successful offline, but less connected online.

III. DATA AND METHODOLOGY

A. Dataset collection

To examine how people behave on social media during terror-related events and to identify echo chambers during those events, we used sentiment analysis for people's tweets, topic analysis for shared URL sites, and experimental fake news classification for those sites. We used two events: the Manchester bombing/attack on 22.5.2017 and the Las Vegas shooting on 2.10.2017. We analyzed tweets containing terror-related keywords (terror, terrorism, terrorist, ISIS) from 18.5.2017 to 5.6.2017 for the Manchester attack event, and from 29.9.17 to 7.10.2017 for the Las Vegas shooting event. The Tweepy library was used for data scraping, and MondoDB for storage.
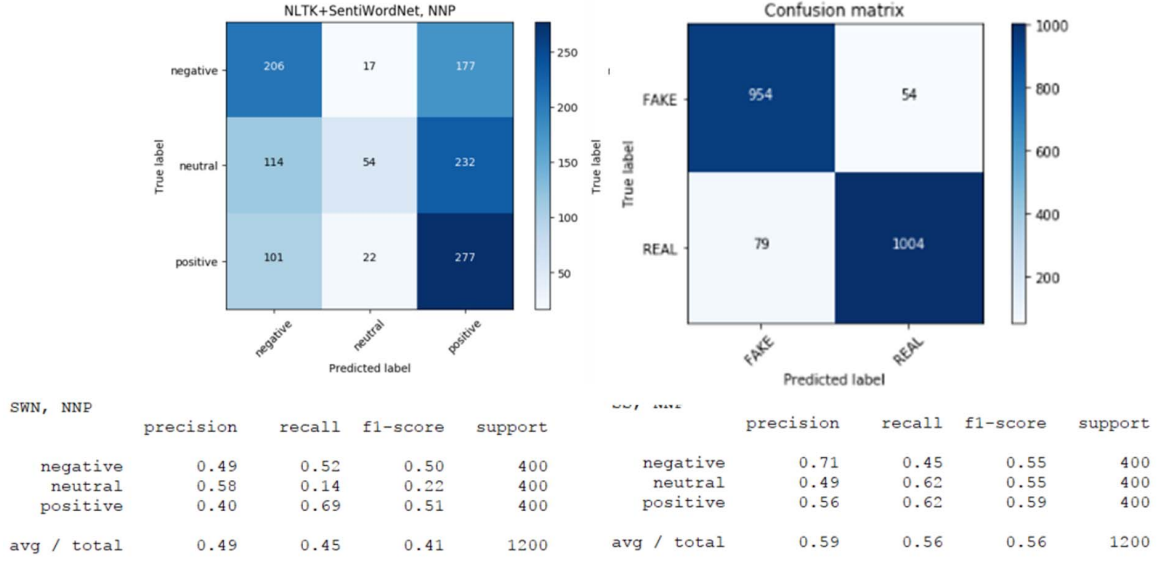
Fig. 1. Confusion matrices of NLTK+SentiWordNet and SentiStrength with negative, neutral and positive sentiments.
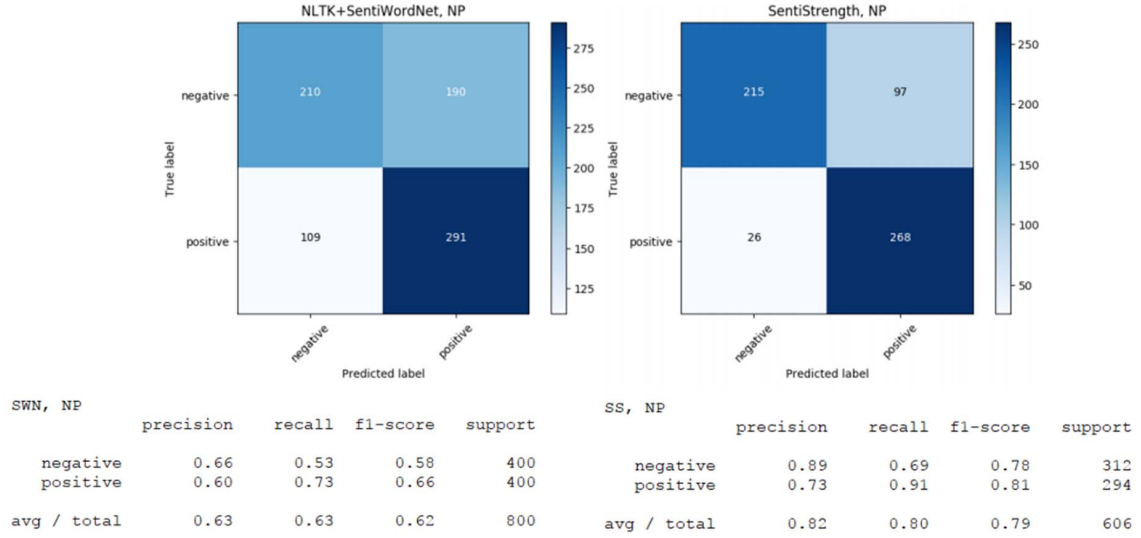
SWN, NNP	precision	recall	f1-score	support
negative	0.49	0.52	0.50	400
neutral	0.58	0.14	0.22	400
positive	0.40	0.69	0.51	400
avg / total	0.49	0.45	0.41	1200

SS, NNP	precision	recall	f1-score	support
negative	0.71	0.45	0.55	400
neutral	0.49	0.62	0.55	400
positive	0.56	0.62	0.59	400
avg / total	0.59	0.56	0.56	1200

Fig. 2. Confusion matrices of NLTK+SentiWordNet and SentiStrength with negative and positive sentiments.

SWN, NP	precision	recall	f1-score	support
negative	0.66	0.53	0.58	400
positive	0.60	0.73	0.66	400
avg / total	0.63	0.63	0.62	800

SS, NP	precision	recall	f1-score	support
negative	0.89	0.69	0.78	312
positive	0.73	0.91	0.81	294
avg / total	0.82	0.80	0.79	606

B. Methodology

Our implementation consisted of three main analytical parts. First, the sentiment analysis gives the overall indication of positive and negative opinions. Second, the topic analysis identifies the echo chambers. Third, the experimental fake news classification of shared URL sites was analyzed.

1) Sentiment analysis

In our work we computed positive and negative scores of each tweet using two lexicon-based methods. Moreover, if the scores were equal, we considered a tweet to be neutral. Furthermore, only English tweets were selected for the analysis, and they were preprocessed by removing URLs, hashtags, mentions, retweet identifiers, and unnecessary whitespace characters.

In the first method we used NLTK [6] and SentiWordNet 3.0 [7]. Scores were calculated for each word by searching positive and negative scores from a lexicon by a word string and its part-of-speech tag. The final sentiment (polarity in the figures) score was computed by summing the positive and negative scores and then averaged by the total number of words in a tweet. As the second method we used SentiStrength [8]. The main difference from the first method is that the SentiStrength does not perform grammatical parsing, i.e., part-of-speech tagging [9]. So the order of the words did not matter in that sense. However, SentiStrength is bundled with a more diverse set of tools for determining sentiment scores, such as handling of negation words, emoticons, idioms and slang words [9].

The two methods were compared with each other against the Sanders Twitter Sentiment Corpus [10]. The corpus contains hand-classified tweets in positive, neutral, negative and irrelevant classes. We selected only positive, neutral, and negative classes for evaluation. According to the corpus documentation, positive and negative tweets may contain only

positive and negative sentiments, respectively. In addition to lack of emotional charge, neutral tweets may contain also mixed positive and negative indicators according to the documentation.

In Figure 1 all three sentiments are involved. The first method seemed to be less accurate, and only a few tweets were classified as neutral. SentiStrength, in turn, was more likely to predict tweets as neutral, but did so much more accurately. The overall accuracy improved in both methods when the neutral sentiment was ignored as seen in Figure 2. Note that on the right side in Figure 2, support values were a bit smaller, since there were not enough positive and negative tweets left in the Sanders corpus due to ignored true and predicted neutral tweets. Overall, the SentiStrength seemed to perform better, so we determined to use SentiStrength for the actual analysis.
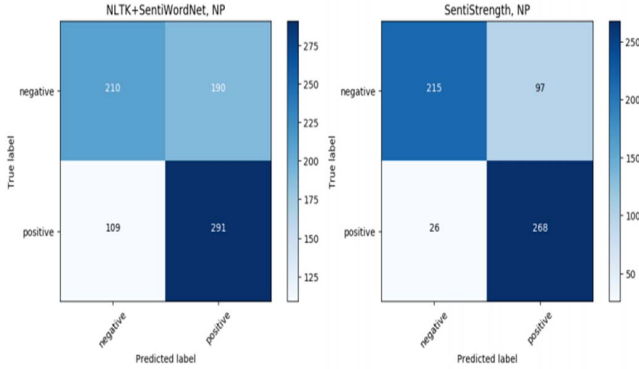

Fig. 3. Confusion matrix of fake news classifier.

2) Topic analysis

For topic analysis we used Gensim's Latent Dirichlet allocation (LDA) model [11]. The visualisation of the topics was effected with pyLDAvis [12]. The number of topics were selected to be ten based on the testing of different number of topics. When the number of topics was increased, the distribution began to be too distributed, and identifying the echo chambers was difficult. Topics also started to multiply by sharing many of the same most used terms. For a corpus we used the titles of the shared URL sites. The text corpus was then transformed to matrix market format which was then passed to the LDA model. LDA was run for 20 passes to maximize the coverage of the documents. After the model was trained, the corpus was run through the LDA model and the most suitable topic for each document was collected.

3) Fake news classification

We used one of the pre-trained models for fake news classification [13]. Classifier is a passive aggressive classifier which is targeted for streaming data and binary classification. Based on the evaluation results of this model (accuracy 93% on test data), this model seemed good for this purpose. The passive aggressive classifier is an online linear classifier; this means that the classifier has one weight vector that is evaluated, and its parameters are redefined during the training. Being an online classifier all the training data and previous input is forgotten, and the classification is performed based only on the weight vector. The confusion matrix of the model is presented on Figure 3.

IV. RESULTS

A. Sentiment analysis

The sentiment analysis of the two events is presented in Figures 4-5 by negative, neutral and positive echo chambers. Throughout the time series, negative tweets are more frequent than positive tweets, and the polarity score also indicates negativity. During the events negative tweets can be observed about three to ten times more frequently than positive tweets. In Figure 4, neutral tweets are more frequent during the event. In Figure 5, the polarity score seems to subtly decrease for a rather long time before the event. Similar trends can also be observed in Figure 4.
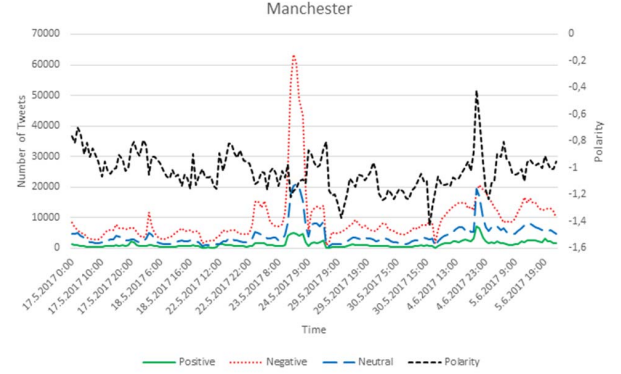

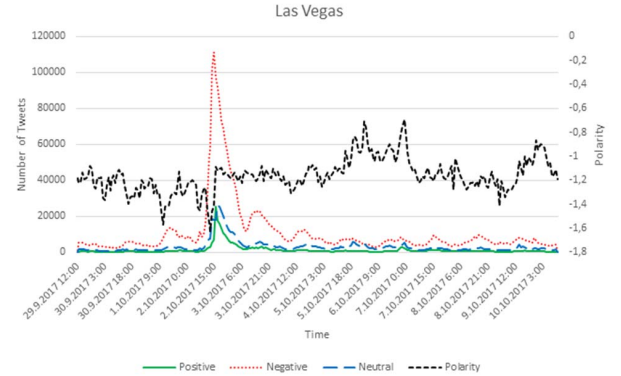Fig. 4. Sentiment analysis on Manchester Arena bombing.

Fig. 5. Sentiment analysis on Las Vegas shooting.

B. Topic analysis

The distribution of topics of shared URL sites are represented on Figures 6-7. The difference of topics is calculated with Jensen–Shannon divergence metrics, and the distance is scaled with classical multidimensional scaling.

From Figure 6, five different echo chambers can be identified for the Manchester attack. Topics 1, 4, 5 and 9 create an echo chamber for themes like ISIS and the Manchester attack. Topics 3 and 5 create an echo chamber for topics like the Philippines, threat and ISIS. Topic 2 creates an echo chamber for corruption news. Topic 10 creates an echo chamber for corruption in India. Topics 8 and 7 create an echo chamber for the mafia.

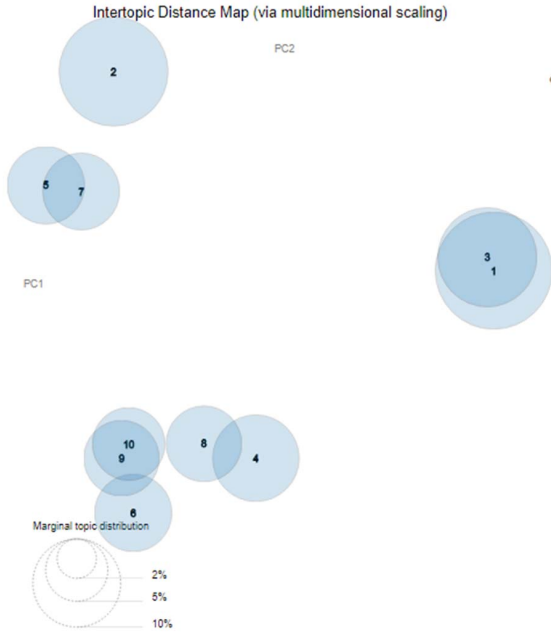Fig. 6.   Topic distribution of the Manchester attack.

Fig. 7.   Topic distribution of the Las Vegas shooting.

From Figure 7, four different echo chambers can be identified. Topics 1 and 3 creates an echo chamber for topics like ISIS, Las Vegas, and Syria. Topic 2 creates an echo chamber for corruption news. Topics 5 and 7 create an echo chamber for Trump corruption. The rest of the topics are unrelated, general noise that don't represent any specific topic.

C. Fake news classification

Our data for fake news classification consist of 63,701 documents for the Manchester attack and of 52,341 documents for the Las Vegas shooting. The results of the classification are presented in tables 1-2. Classifier classified 80% of documents as fake news and 20% as real news for the

Manchester attack. For the Las Vegas shooting, the division was 75% fake news and 25% real news.

TABLE I.   RESULTS OF FAKE NEWS CLASSIFICATION FOR MANCHESTER ATTACK

Type	Number of documents	Percentage of documents
Real	12 196	20%
Fake	51 505	80%

TABLE II.   RESULTS OF FAKE NEWS CLASSIFICATION FOR LAS VEGAS SHOOTING

Type	Number of documents	Percentage of document
Real	13 048	25%
Fake	39 293	75%

How the sharing of news and the number of real and fake news developed over time are presented in Figures 8 and 9. The Y-axis represents the number of documents, and the X-axis represents the date. The blue line is fake news, and the red line the real news. The Las Vegas shooting happened at five AM on the second of October in UTC time, which explains the spike on the same day as the event happened. In both events there is an increase of shared news after the event. A couple of days after the event, the sharing of news decreases.
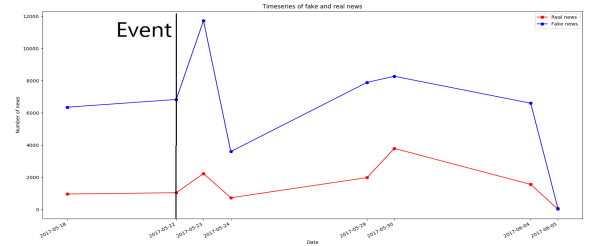

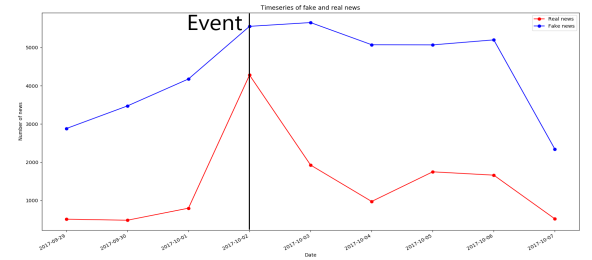Fig. 8.   Time series of news on Manchester attack

Fig. 9.   Time series of news on Las Vegas shooting

V.   DISCUSSION AND CONCLUSION

According to the sentiment analysis, there seems to be a lot of negativity involved in the tweets. One reason for that might be the terrorism-related vocabulary due to the initial filtering of the data set. Intuitively, the fact that content about terrorism and conflicts is negatively charged could explain the negativity. However, after the events, there can be emerging behavior that might be considered positive, e.g., praying. During the selected events, the mood of the tweets seem to rapidly change toward a more negative direction and then

normalize rather quickly. That could be explained by people's first reactions. There is also clear fluctuation in the sentiment curves. Interestingly, negativity seems to strengthen even before the selected incidents. There is certainly some noise that could affect the result. For example, Hurricane Maria could partly explain why there seems to be some long-term change before the Las Vegas shooting incident. There are also small spikes in the figures that could be considered as sub-events as in the presented Woolwich case. Interestingly, tweets are clearly more positive after than before the Las Vegas shooting incident, which could be explained by supportive messages and positive affirmations or events such as concerts, for example.

Topic analysis shows the general echo chambers. These chambers are more global chambers and don't take a stand on that particular event with differing opinions. To get more detailed echo chambers, we need to examine these identified echo chambers and topics as their own area of research. Topic analysis also suffers a little bit from filtered data gathering. Because the tweets were filtered during the collection to be about terrorism, ISIS, and corruption, those topics pop out in the topic analysis.

Echo chamber identification could also be performed by identifying the linked users who interact with each other. This approach still has the problem of users who debate and have differentiating opinions, but still message each other frequently. There are two echo chambers, but that approach would identify only one. If sentiment analysis would be applied to this, the problem would be solved. The third possible way to identify echo chambers would be to examine the shared URLs and how those URLs are linked. People who share the same URLs or URLs from the same domain would be identified as one echo chamber. This approach has risk for the reasons mentioned earlier, and similarly, it could be solved with use of sentiment analysis. Of course sentiment analysis on these approaches would be complex because the differentiating sentiment should be identified and the overall sentiment should be decided based on the sentiment of the main tweet.

The results of fake news detection are quite interesting and abnormal. 80% of shared news should not be classified as fake news. It is good to notice that this kind of classification is still experimental and might not be suitable for this purpose. Fake news sites are very complex, and identifying them by linguistic patterns can be extremely difficult; doing so would require some comparison to other news. Based only on the language of that one document, the classification is vulnerable to noise in the document and in the classifier. For example, the most weighted terms for wake news were Clinton, October, and share, while the most weighted terms for real news were said, Tuesday, and GOP. Based on the occurrence of these terms, we can't say whether the document is real or fake news, but the classifier was strongly affected by these terms.

VI. REFERENCES

[1] What is an Echo Chamber? - Definition from Techopedia. (n.d.). Retrieved November 24, 2017, from https://www.techopedia.com/definition/23423/echo-chamber

[2] Roberts, C., Innes, M., Preece, A., & Rogers, D. (2017). After Woolwich: Analyzing Open Source Communications to Understand the Interactive and Multi-Polar Dynamics of The Arc of Conflict. The British Journal of Criminology.

[3] Innes, M., Roberts, C., Preece, A., & Rogers, D. (2016). Ten "Rs" of social reaction: Using social media to analyse the "post-event" impacts of the murder of Lee Rigby. Terrorism and Political Violence, 1-21.

[4] Spasić, I. (n.d.). FlexiTerm. http://users.cs.cf.ac.uk/I.Spasic/flexiterm/

[5] Bright, J. (2016). Explaining the Emergence of Echo Chambers on Social Media: The Role of Ideology and Extremism. SSRN Electronic Journal.

[6] NLTK. (n.d.). NLTK. Retrieved November 25, 2017, from http://www.nltk.org/

[7] SentiWordNet. (n.d.). SentiWordNet. http://sentiwordnet.isti.cnr.it/

[8] Thelwall, M. (n.d.). SentiStrength. Retrieved November 25, 2017, from http://sentistrength.wlv.ac.uk/

[9] Thelwall, M. (2016). The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength. Understanding Complex Systems Cyberemotions, 119-134.

[10] Sanders, N. (n.d.). Twitter Sentiment Corpus. Retrieved November 25, 2017, from http://www.sananalytics.com/lab/twitter-sentiment/

[11] Gensim: topic modelling for humans. (n.d.). Retrieved November 25, 2017, from https://radimrehurek.com/gensim/models/ldamodel.html

[12] Mabey, B. (2017). Bmabey/pyLDAvis. Retrieved November 25, 2017, from https://github.com/bmabey/pyLDAvis

[13] Jarmul, K. (2017). Detecting Fake News with Scikit-Learn. https://www.datacamp.com/community/tutorials/scikit-learn-fake-news