# Predicting the Level of Violence and Participation During Social Unrest with the Use of ELMo and Social Media Language

Alicja Mincewicz
Department of Computing & Information Science
Mercyhurst University
501 East 38th St
Erie, PA 16546
alicja.mincewicz@gmail.com

## Abstract

This research focuses on developing a model that can predict the level of violence and participation during a social event, such as a protest, rally, or demonstration based on Twitter language. The data consists of tweets collected around specific demonstrations throughout the USA, Australia, and the UK. Additional data on the specifics of the events were collected from official news outlets. I utilized Embedded Language Modeling (ELMo) to create a methodology to define the level of violence and participation. In this method, the word representations are functions of the entire input sentence and are computed on top of a deep bidirectional language model (biLM). ELMo considers not only word vectors but also syntax, semantics, and model polysemy.

## Keywords

Protest, Demonstration, Violence, Participation, Twitter, Natural Language Processing, Embedded Language Modeling, Text Mining, Social Media Analysis, Social Unrest

## 1. INTRODUCTION

Many citizens partake in demonstrations to exercise their democratic rights to do so and to express their opinions. In some places, it is a way to initiate change and show approval or disapproval of a socio-economical or geopolitical situation, even if such social events are not necessarily approved by the governments. Approximately 75% of the demonstrations are legal and organized in advance [2]. Additionally, protests with larger attendance are much more likely to be successful and carry a lower risk of arrests. Therefore, participants try to plan and announce demonstrations ahead of time in order to mobilize a larger number of participants and increase their event's success rate [2][10]. In some instances, demonstrations and protests can escalate into violence, causing deaths, injuries, and property damage.

Increasingly, social media serves as one of the main open sources of information and plays an important role in the organization of social events, such as demonstrations, parades, and other gatherings. Twitter alone attracts 126 million active daily users and 321 million active monthly users [3]. Ordinary citizens, government agencies, and journalists use Twitter to disseminate information in real-time, being much faster than traditional news outlets. Social media was widely used during the Baltimore protests in 2015, Arab Springs between 2010 and 2012, and the Charlottesville, VA rally in August 2017 [4][13]. Attendees used Twitter to update others on the developments of the events, to communicate with other participants, and to make further plans. Such events can be costly in many ways, as all three of the above-mentioned protests caused property damage, injuries, and deaths.

Timely and accurate predictions of protests and demonstrations, and the likelihood of violence can help law enforcement, authorities, and public officials to prepare in advance and take necessary precautions. These precautions can not only limit the monetary costs of damage but also protect people's lives. In corporate security, many organizations, such as Amazon, Microsoft, and Facebook have their global security departments and intelligence teams that monitor social and unrest daily. For many organizations, such tasks are done manually by analysts who keep track of many different information outlets using several different types of software. Analysts determine the likelihood of demonstrations escalating into violence with the use of their expertise and other information, such as perceived levels of participation, presence of extremist groups, location, and cause of the demonstration. While some protests are inherently more likely to escalate into violence, leadership bases their actions on the analysts' analysis.

The main objective of this research is to develop a model that estimates the level of violence and the number of participants during a demonstration, protest, or social unrest. Corporate security and their intelligence departments would benefit from the model as it would reduce the amount of manual work the analysts need to perform. Additionally, with the use of Natural Language Processing, the model would pick up on linguistic nuances that a human might not be able to. In this project, I will focus on Embedded Language Modeling with the use of Twitter. The data includes tweets collected around the Unite the Right Rally in August 2017 in Charlottesville, VA, and tweets collected between January and March 2020.

Most of the previous studies have been conducted at the strategic level, such as considering demonstrations and protests as organized social movements and looking at how socioeconomic and political situations impact protest moods [8]. This research focuses more on the tactical level of analysis and focuses on the estimated number of participants and the level of violence.

## 2. RELEVANT WORK

This research is related to the following research areas: prediction of events and social behaviors, social media mining, and research on behaviors and language that help predict the level of violence and estimated attendance during an event. Previous studies have mainly focused on event detection and the socioeconomic and geopolitical aspects, while modeling demonstrations as social movements [12]. Scholars have also examined the role of collective identity and the sense of empowerment that participation in protests bring. Groups that are perceived to be strong are more likely to experience anger and are more determined to take action [12]. Combining social sciences, such as psychology and sociology, with computer science allowed researchers to look at modeling behaviors, predicting actions, and predicting participants' behaviors and attitudes [7][8].

Many researchers have investigated predicting social unrest on social media with the use of several methods, with clustering being one of the most prevalent. The existence of the event can be deducted from changes in trends and topic popularity on social media. With the use of cluster analysis, scholars divided the data into subtopics based on common keywords and hashtags [8], creating clusters of different sizes. Additionally, clusters are created dynamically and evolve over time, which enables the model to add new points to the clusters [2][8]. The main limitation of this method is that in large social media networks and datasets, the number of nodes may be too large to use efficiently. To combat this issue, Aggarwal and Subbian [2] used a sketch-based retrieval technique in addition to regular cluster analysis. With the use of this method, the researchers maintained node counts in underlying clusters. Additionally, a sketch table was used to maintain frequency counts of the nodes in the incoming data. With the increase of the sketch-table length, the purity of clusters increases as well, improving the overall accuracy [2]. Aggarwal and Subbian used two datasets for their research: a collection of 1,628,779 tweets and an Enron email dataset.

Other scholars focused more on NLP. Mutial et. al. [10] first used linguistic processing with the use of tokenization, lemmatization, and named entity extraction to analyze the language. The documents were then filtered by phrases and parsed with the use of a dependency parser. The researchers used Probabilistic Soft Logic (PSL) in the geocoding of the news and blog sources. PSL is a framework that uses collective, probabilistic reasoning that uses first-order logic rules [6]. Tweets and Facebook posts that did not already include information on locations and geotags were geotagged based on the locality that the text was about. The aforementioned methods, together with additional rules and constraints to the model, were used to determine dates and locations of demonstrations. This system has been used for several years by analysts in Latin America [6].

Won et. al. researched perceived violence during social unrest with the use of image analysis of a UCLA Protest Image Dataset [14]. The scholars focused on visual analysis with the use of a Convolutional Neural Network (CNN) and OpenFace models. The CNN model uses the input of thousands of single pictures, automatically classifying these images, and outputs a series of prediction scores that included visual attributes, binary image categories (e.g. non-protest and protest), perceived violence, and image sentiment [14]. The OpenFace model with the use of a CelebA facial attribute dataset outputs information on race and gender, among other expressions. While this research is one of very few that address the issue of perceived violence, it does so with image and not text analysis.

This paper takes social media analysis and text mining a step further as it focuses on tactical data on predicting estimated attendance and level of violence. Unlike the aforementioned scholars, this research uses Embedded Language Modeling (ELMo) [11]. The main dataset used for this project includes Tweets with #charlottesville spanning a few days in August 2017. The instance of Charlottesville, VA is a great example of how a spontaneous demonstration can drastically escalate into violence. Additional data was collected between January and March 2020. It includes information on numerous demonstrations that occurred within this timeframe.

## 3. PROPOSED SOLUTION

To solve the research problem, I used Embedded Language Modeling (ELMo) [11]. I collected data about specific demonstrations and protests by using geolocations and hashtags related to various topics. The Tweet data was cleaned by removing hyperlinks, emoticons, stop words, non-Latin characters, punctuations, and user mentions. Each of the demonstrations, plus tweets on negative data, were treated separately. Each demonstration's dataset included a collection of tweets about that event. That text was compared against the official data on attendance and levels of violence. The language from these tweets was analyzed using Embedded Language Modeling. This model is often used in Sentiment Analysis or Named Entity Extraction [11] as it looks at sentences as a whole to calculate word embeddings. This approach allows for future prediction of perceived attendance and violence based on language changes.

## 4. METHODOLOGY

### 4.1 Data

I obtained an existing data set containing tweets posted on 15 August 2017 from Charlottesville, VA [1]. The Tweets were collected following violent demonstrations in that city. Additionally, I collected tweets between January and March 2020 from various locations throughout the US, including Seattle, New York City, Austin, Washington D.C., and Portland. I also collected data from the UK and Australia. These tweets were mined around specific demonstrations. The data also included randomly collected tweets from Los Angeles and Seattle to include negative data. Overall, I collected 28,338 tweets, which were cleaned. Afterward, I looked at official news articles and government websites to obtain the number of participants and determine the level of violence. The number of participants and the level of violence was divided into the below-listed subcategories.

Levels of violence are divided into the following categories (police actions include the use of batons, tear gas, etc.):

- *Level 0* ->–Not a social event.
- *Level 1* -> Low level of violence – no violence during the event/peaceful event.
- *Level 2* -> Medium level of violence – possible injuries, police actions, and property damage.
- *Level 3* -> High level of violence – at least 1 death, police actions, and property damage.

Attendance is divided into the following categories:

- *Level 0* -> 0 participants – not a social event
- *Level 1* -> 1 – 1,000 participants
- *Level 2* -> 1,000 – 10,000 participants
- *Level 3* -> 10,000 – 50, 000 participants
- *Level 4* -> 50,000 participants <

Attendance and violence levels were treated as two separate categories. I first looked at the tweets and levels of violence. The collected data was classified into one of the above categories based on the additional information obtained from news, government data, etc.). Initially, each demonstration/protest was labeled as a whole (e.g. Washington, DC, demonstration – violence level 1) and then divided into individual tweets with the same labels as the event overall. Afterward, I joined all 28,338 tweets and created one Pandas data frame.

| | Tweet | Violence_Level |
|---|---|---|
| 16629 | Charlotte Pence Bond author and the daughter ... | 1 |
| 7082 | In solidarity with everyone who protested ton... | 1 |
| 1051 | Fake things like trend overnight because the ... | 2 |
| 13970 | Had great afternoon and performing in Whitehal... | 1 |
| 1620 | The anti Semitic attacks happening across our ... | 1 |
| ... | ... | ... |
| 13029 | We are fighting for climate justice We are fi... | 1 |
| 342 | ANTIFADomesticTerrorism | 2 |
| 5766 | After that disturbing interview with Morrison ... | 1 |
| 10738 | Can barely capture the whole crowd Huge clima... | 1 |
| 7682 | Epic crowd now heading down Park Street towar... | 1 |

28338 rows × 2 columns

*Figure 1: Data Frame for Violence Levels*

I later divided the data frame into three parts: training data, validation data, and testing data. The testing data consisted of 2,338 tweets. The training data consisted of 20,800 tweets and the validation data was 5,200 tweets (80% train data and 20% validation data). I implemented the same process to the number of participants.

| | Tweet | Participation_Level |
|---|---|---|
| 10684 | Speeches over packed crowds funnel out to beg... | 3 |
| 880 | Fixed | 4 |
| 7972 | Incredible photo taken at the Sydney rally | 3 |
| 1543 | Why should Catholics stand with our Jewish br... | 3 |
| 874 | Los Angeles Dodgers LED Flashlight | 0 |
| ... | ... | ... |
| 4946 | AntifaTerrorists This hashtag created by corp... | 1 |
| 10234 | It heartening to see in some areas the bush i... | 3 |
| 1662 | Awesome crowd global protest against inequali... | 1 |
| 9399 | How First Australians ancient knowledge can he... | 3 |
| 966 | aux militants de la Je me bats pour vous via ... | 4 |

28338 rows × 2 columns

*Figure 2: Data Frame for Participation Levels*

Later, I built two separate, but very similar, ELMo models. The first model included four violence classifications and the second model included five participation classifications.

## 4.2 Embedded Language Modeling

ELMo is a "deep contextualized word representation that models both complex characteristics of word use (e.g., syntax and semantics), and how these uses vary across linguistic contexts (i.e., to model polysemy)"[5]. The word representations in the model are functions of the entire input sentence and are computed on top of a deep biLM with character convolutions, as a linear function of the internal network states[11]. The two-layer construction allows for semi-supervised learning as biLMs are pre-trained on a large text corpus. There are three key features of ELMo [5]:

- *Context* – each word representation depends on the context in which the word is used.
- *Deep* – the representations of words combine all layers of a deep pre-trained neural network.
- *Character-based* – character-based word representations allow the "network to use morphological clues to form robust representations for out-of-vocabulary tokens unseen in training"[5].

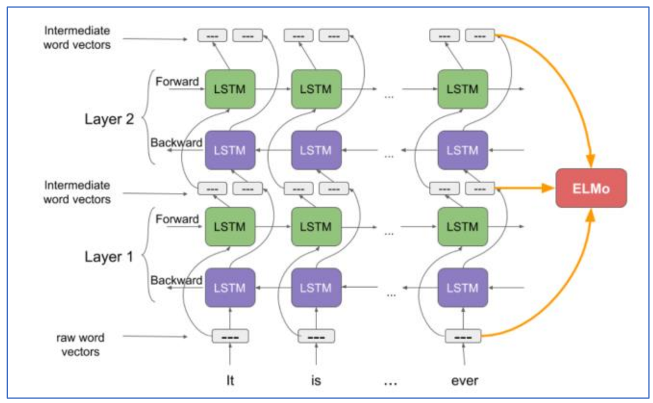


*Figure 3: ELMo word vectors computed on top of a two-layer bidirectional language model (biLM)*[9]

The pre-trained ELMo model can be found in TensorFlow Hub. The output for an ELMo vector is a 3 dimensional tensor of shape, TensorShape([Dimension(a), Dimension(b), Dimension(1024)]. The first dimension (a) of the output represents the number of training samples. The second dimension (b) represents the length of the maximum string. The last dimension is equal to the length of the ELMo vector. Below are model summaries for both models.

```
Model: "model_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_4 (InputLayer)         [(None, 1)]               0
_________________________________________________________________
lambda_3 (Lambda)            (None, 1024)              0
_________________________________________________________________
dense_6 (Dense)              (None, 256)               262400
_________________________________________________________________
dense_7 (Dense)              (None, 4)                 1028
=================================================================
Total params: 263,428
Trainable params: 263,428
Non-trainable params: 0
```

*Figure 4: Model Summary for Violence Levels*

```
Model: "model_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 1)]               0
_________________________________________________________________
lambda_1 (Lambda)            (None, 1024)              0
_________________________________________________________________
dense_2 (Dense)              (None, 256)               262400
_________________________________________________________________
dense_3 (Dense)              (None, 5)                 1285
=================================================================
Total params: 263,685
Trainable params: 263,685
Non-trainable params: 0
```

*Figure 5: Model Summary for Participation Levels*

As violence levels and participation levels do not impact each other, I created two models that were trained, validated, and tested separately. While the model for violence had four classifying categories, the model for participation had five. Both models' features included five epochs and batch sizes of 256. This means that there were five complete passes through each model. Additionally, looking at the training sets of 20,800 tweets, there were 81 batches with 256 samples and one batch with 64 samples, which means each model was updated 82 times within each epoch. This was repeated four more times to constitute five epochs. Each of the models were validated on 5,200 tweets and tested on 2,388 tweets.

# 5. RESULTS AND DISCUSSION

With every epoch, training and validation, accuracies increased whereas training and validation losses decreased. Below are the representations of training and validation accuracies, as well as training and validation losses.

## 5.1 Levels of Violence Model:

| Epoch | Training Acc | Training Loss | Validation Acc | Validation Loss |
|---|---|---|---|---|
| **1** | 0.8728 | 0.6937 | 0.9537 | 0.4394 |
| **2** | 0.9700 | 0.3424 | 0.9783 | 0.2785 |
| **3** | 0.9829 | 0.2319 | 0.9831 | 0.2015 |
| **4** | 0.9874 | 0.1729 | 0.9846 | 0.1583 |
| **5** | 0.9891 | 0.1415 | 0.9850 | 0.1372 |

Between the first and the fifth epoch, the training accuracy increased by approximately 13.32%, and the validation accuracy increased by approximately 3.28%. The final accuracy of the testing set was 0.9859, which can be interpreted as the model accurately predicted the level of violence for 98.59% of tweets.

If the model included only one epoch, the accuracy of the training data would be 0.8670, and 0.9460 for the validation data. The final test accuracy would be 0.9491. The accuracy of the test data increased by approximately 3.88% when using five epochs instead of one.

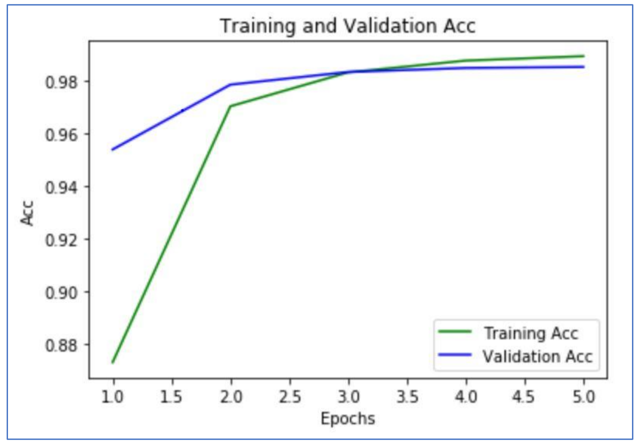


*Figure 5: Training and Validation Accuracies for Levels of Violence*

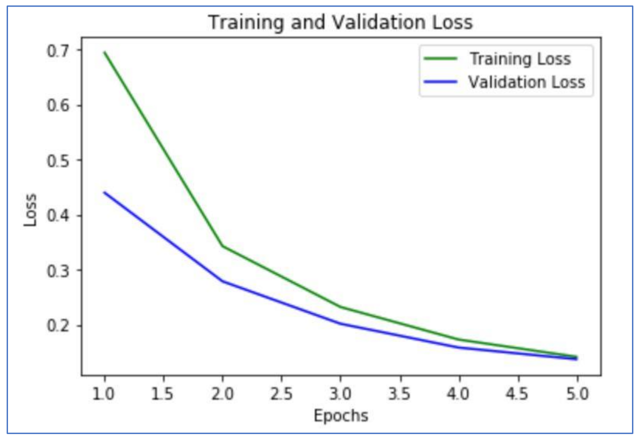


*Figure 6: Training and Validation Losses for Levels of Violence*

## 5.2 Levels of Participation Model

| Epoch | Training Acc | Training Loss | Validation Acc | Validation Loss |
|---|---|---|---|---|
| **1** | 0.8530 | 0.7874 | 0.9571 | 0.4679 |
| **2** | 0.9757 | 0.3662 | 0.9790 | 0.2913 |
| **3** | 0.9878 | 0.2430 | 0.9871 | 0.2076 |
| **4** | 0.9913 | 0.1788 | 0.9881 | 0.1640 |
| **5** | 0.9918 | 0.1461 | 0.9879 | 0.1461 |

Between the first and the fifth epoch, the training accuracy increased by approximately 16.27%, and the validation accuracy increased by approximately 3.22%. The final accuracy of the testing set was 0.9846, which can be interpreted as the model accurately predicted the level of violence for 98.46% of tweets.

If the model included only one epoch, the accuracy of the training data would be 0.8497 and 0.9635 for the validation data. The final test accuracy would be 0.9632. The accuracy of the test data increased by approximately 2.22% when using five epochs instead of one.

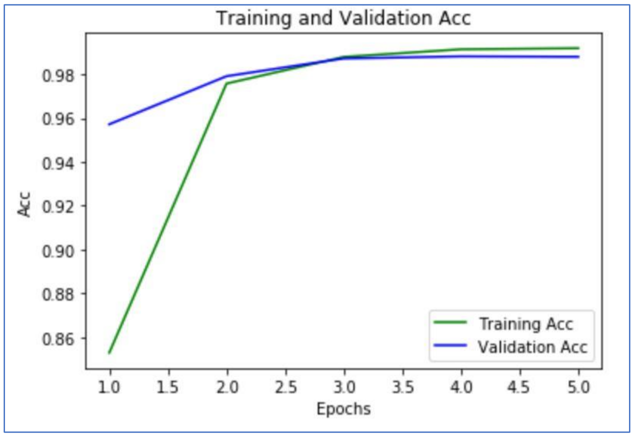


*Figure 7: Training and Validation Accuracies for Levels of Participation*

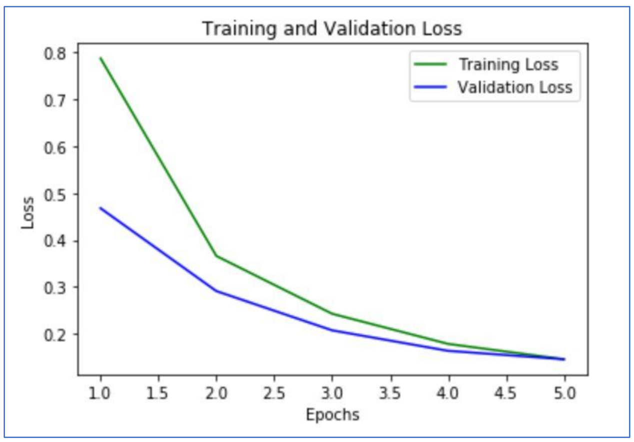


*Figure 8: Training and Validation Losses for Levels of Participation*

### 5.3 Both Models

The testing accuracy was higher in the model used for predicting the level of violence. While the difference is insignificant, it may be caused by the different number of classifiers. There are four different levels of violence the tweets were classified into, while there are five levels of participation. When comparing accuracies and losses between using one epoch versus five, we can conclude that five epochs give more accurate results. The model iterates over the data five times, enforcing the training and validation.

## 6. CONCLUSIONS & FUTURE WORK

ELMo is one of the best performing models in text analysis. It does not only look at traditional word vectors, but also at syntax, semantics, and model polysemy. Linguistic nuances are important to the intelligence community, as the words can have different meaning based on context. The sentence “This movie set was the bomb” has a very different meaning than “I will bomb the movie set.” Traditional word embeddings will come up with the same vector for the word “bomb” in both sentences.

Often, it is difficult to go through the data manually. Intelligence analysts do not always have the time or resources to try to pick up on detailed language differences that ELMo can. This research showed that with the use of ELMo, the model can predict the level of violence 98.59% of the time and participation 98.46% of the time. Such information is useful not only to analysts but to entire corporate security departments. Daily, leadership decides whether to keep assets open, suspend work trips, or even evacuate buildings if necessary. If an analyst can determine how violent a demonstration will get based on language changes, the leadership can make appropriate decisions on how to proceed, often impacting the health and safety of the personnel.

There are several improvements to the model that I would like to make in the future. Firstly, I would like to start with k-means clustering to detect trends in tweets in real-time. This would allow for appropriate data collection as the events are happening and not afterward. The clusters that include information on social events such as demonstrations, protests, and rallies would then be used to determine the level of violence and participation. This way, intelligence analysts would be able to witness the escalation of violence and changes in participation in real-time, while observing the events. Additionally, I would like to get more granular as it comes to participation and level of violence and establishes more categories. Currently, there are four levels of violence plus negative data, and five participation categories plus negative data. The smaller the brackets for violence and participation, the more actionable intelligence the analyst has.

Additionally, time constraints and the availability of data with the use of regular Twitter API did not allow for the collection of a larger dataset. Similar language is often used within a social event, therefore, limiting the diversity of words, context, and syntax. The more information collected from different protests and demonstrations, whether based on the reason for the demonstration or country of origin, the broader language spectrum, thereby making the model more comprehensive.

## 7. REFERENCES


[1] #Charlottesville on Twitter | Kaggle: *https://www.kaggle.com/vincela9/charlottesville-on-twitter#aug15_sample.csv*. Accessed: 2020-04-13.

[2] Aggarwal, C.C. and Subbian, K. 2012. Event detection in social streams. *Proceedings of the 12th SIAM International Conference on Data Mining, SDM 2012*. (2012), 624–635. DOI:https://doi.org/10.1137/1.9781611972825.54.

[3] Bahrami, M. et al. 2018. Twitter Reveals: Using Twitter Analytics to Predict Public Protests. (2018), 1–24.

[4] Baltimore Riots: Social Media and the Crisis on My Doorstep - WSJ: *https://www.wsj.com/articles/baltimore-riots-social-media-and-the-crisis-on-my-doorstep-1430243047*. Accessed: 2019-10-21.

[5] ELMo: Deep contextualized word representations: *https://allennlp.org/elmo*. Accessed: 2020-03-23.

[6] Kimmig, A. et al. 2012. A Short Introduction to Probabilistic Soft Logic. *Proceedings of the NIPS Workshop on Probabilistic Programming: Foundations and Applications*. 1 (2012), 1–4.

[7] Korolov, R. et al. 2015. Actions are louder than words in social media. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015*. (2015), 292–297. DOI:https://doi.org/10.1145/2808797.2809376.

[8] Korolov, R. et al. 2016. On predicting social unrest using social media. *Proceedings of the 2016 IEEE/ACM*

*International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*. (2016), 89–95. DOI:https://doi.org/10.1109/ASONAM.2016.7752218.

[9] Learn ELMo for Extracting Features from Text (using Python): *https://www.analyticsvidhya.com/blog/2019/03/learn-to-use-elmo-to-extract-features-from-text/*. Accessed: 2020-03-23.

[10] Muthiah, S. et al. 2015. Planned protest modeling in news and social media. *Proceedings of the National Conference on Artificial Intelligence*. 5, (2015), 3920–3927.

[11] Peters, M.E. et al. *Deep contextualized word representations*.

[12] van Stekelenburg, J. and Klandermans, B. 2013. The social psychology of protest. *Current Sociology*. 61, 5–6 (2013), 886–905. DOI:https://doi.org/10.1177/0011392113479314.

[13] The Role of Social Media in the Arab Uprisings | Pew Research Center: *https://www.journalism.org/2012/11/28/role-social-media-arab-uprisings/*. Accessed: 2019-10-21.

[14] Won, D. et al. 2017. Protest activity detection and perceived violence estimation from social media images. *MM 2017 - Proceedings of the 2017 ACM Multimedia Conference*. (2017), 786–794. DOI:https://doi.org/10.1145/3123266.3123282.