Predicting Social Unrest Using GDELT

Divyanshi Galla[1(✉)] and James Burke[2]

[1] PWC, BG House, Hiranandani Business Park, Powai, Mumbai, India
divyanshi.galla@pwc.com
[2] PWC, 600 13th Street Office, Washington, DC, USA
james.a.burke@pwc.com

Abstract. Social unrest is a negative consequence of certain events and social factors that cause widespread dissatisfaction in society. We wanted to use the power of machine learning (Random Forests, Boosting, and Neural Networks) to try to explain and predict when huge social unrest events (Huge social unrest events are major social unrest events as recognized by Wikipedia page 'List of incidents of civil unrest in the United States') might unfold. We examined and found that the volume of news articles published with a negative sentiment grew after one such event - the death of Sandra Bland - and in other similar incidents where major civil unrest followed. We used news articles captured from Google's GDELT (Global Database of Events, Language, and Tone) table at various timestamps as a medium to study the factors and events in society that lead to large scale unrest at both State and County levels in the United States of America. In being able to identify and predict social unrest at the county level, programs/applications can be deployed to counteract its adverse effects. This paper attempts to address this task of identifying, understanding, and predicting when social unrest might occur.

Keywords: Social unrest · News media · GDELT · Themes · Events
Random forest · Ada boost with random forest · LSTM · County level USA

1 Introduction

Social unrest can be extremely detrimental to society, especially when it escalates into rioting and violent demonstrations. Local communities bear the brunt of the impact, and it can have a lasting effect on their socioeconomic development.

A great deal of research has gone into detecting social unrest using social media data; most of the work centering around Twitter [2], Tumblr [4], Facebook and some research was carried using Google's GDELT event table [5]. Our work focuses on the exploration and analysis of news media data to detect social unrest. News media data contains a wealth of rich insights that can be used to study various social factors and events that take place in our society. By isolating underlying trends of factors that lead to social unrest, we could utilize this power to warn the public of the impending danger of adverse social events. Analysis of this data allows us to identify certain themes and events that are closely associated with a given region. It would be truly remarkable if we could develop a system that will be able to detect when and where a social unrest event is brewing, and alert the appropriate authorities to take necessary action; either in

P. Perner (Ed.): MLDM 2018, LNAI 10935, pp. 103–116, 2018.
https://doi.org/10.1007/978-3-319-96133-0_8

preventing the event all together, or providing adequate safety measures for the people. By developing machine learning models that can adequately predict social unrest using news media data, and by reducing manual effort associated with tracking and monitoring such events, we can move a step closer to achieving this system.

Previous research [5] used GDELT's event table to build a Hidden Markov Model (HMMs) based framework to predict indicators associated with country instability. Identifying these indicators at a more granular geographic level could enable focused and effective action from the government, whether in the form of targeted outreach programs, distribution of additional law enforcement resources, or development of early warning systems. Based on these assumptions, we decided to focus our research on huge social unrest events that occur not only at the state level, but also at the county level (within the contiguous US).

We hypothesized that huge riots that took place in Baltimore in the year 2015 and in Milwaukee and Charlotte in the year 2016 had an ethnic discrimination factor and that the protests associated with the Dakota Access Pipeline had environmental related factors associated with a deteriorating sentiment in news articles indicative of building unrest levels in a region. We also hypothesized that a region that is subject to higher occurrences of events associated with *escalated threatening, coercing, assaults, protests* and other forms of *disapproving* behavior might suffer from a buildup of higher levels of social unrest.

We chose one such event, Sandra Bland's death, to test our hypotheses, using GDELT's Global Knowledge Graph (GKG) and Event tables that monitor print, broadcast, and web news media across world. Sandra's death in prison caused major civil unrest after FBI investigation revealed that the required policies weren't followed. After her death on 14th July, 2015 discussion grew with worsening sentiment till impact subsided. A huge unrest event in the form of a protest occurred on 29th July 2015. This confirmed our hypothesis that news reflects society and can be used to detect building unrest. Figure 1 depicts how sentiment of articles discussing the incident vary with time in the case of Sandra Bland's death. The size of the bubble represents mentions of event in news; the Y axis contains the sentiment value and the X axis is timeline.

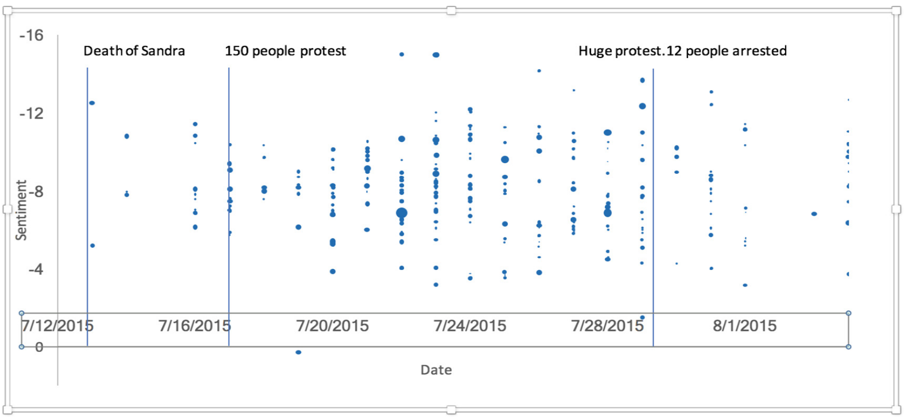

Fig. 1. Figure indicates variation of average tone with respect to time.

2 Related Work

This paper is focused on using news media to predict social unrest at county and state levels for USA. There is a large amount of work done in predicting social unrest using social media. [1] uses social messaging to predict social unrest. [2] works towards automated unrest prediction and filtering the vast volume of tweets to identify tweets relevant to unrest. [3] focuses on detecting emerging civil unrest events by analyzing massive micro-blogging streams of Tumblr. [4] presents an event forecasting model using activity cascades in Twitter to predict the occurrence of protests in three countries of Latin America: Brazil, Mexico, and Venezuela. [5] is using information based on news media from temporal burst patterns in GDELT event streams to uncover the underlying event development mechanics in five countries of South East Asia. Our research aims at measuring social factors prevalent in society using the GKG table of GDELT along with event types gathered from GDELT's event table that might lead to huge social unrest. Our main contribution is to take into account a thoroughgoing picture of society to predict unrest level at the county level of location compared to previous work that looked at larger geographical regions. We are predicting unrest levels in 2918 counties across USA which can be an efficient tool to combat the ill effects of protests.

3 Data

3.1 About GDELT

GDELT is the global database of events, location and tone that is maintained by Google. It contains structured data that is mined from broadcast, print and web news sources in more than 100 languages since 1979. It connects people, organizations, locations, themes, and emotions associated with events happening across the world. It describes societal behavior through eye of the media, making it an ideal data source for measuring social factors and for testing our hypotheses. We have used two GDELT tables, the Global Knowledge Graph (GKG) and the events table.

GDELT's GKG captures what's happening around the world, what its context is, who's involved, and how the world is feeling about it; every single day. Of the 27 fields that it contains, we used the following four: *Location, Date, Themes* with associated *Sentiment*. *Location* field contains all locations found in news article and the average tone of the document is *sentiment* associated with it. The *Date* format was - YYYYMMDD. We divided 472 themes of interest into the following four categories - Crime, Economy, Environment and Health. This was done to observe which category was influencing escalation of unrest. Following are some illustrative examples of themes - examples of Crime themes include *armed conflict*, *arrest, crimes against humanity*; Economy themes include *democracy, constitution* and *alliance*; Environment themes include *deforestation* and *climate change* and Health themes include *disease*.

The GDELT events table captures information on events in CAMEO format (conflict and mediation event observations) capturing two actors and action performed by Actor1 on Actor2. The events table contains 58 fields. We used *SQLDATE*, *EventRootCode*, *NumMentions*, *AvgTone* and *Location* fields. SQLDATE is the date in YYYYMMDD format. *EventRootCode* is the root-level category the event falls under. *NumMentions* is total number of mentions of an event across all source documents which helps to measure importance of an event. *AvgTone* is the average tone of all documents containing one or more mentions of an event (the score ranges from −100 (extremely negative) to +100 (extremely positive)). *Location* is the city, state, country description of all locations mentioned in the article. The cameo event codes that were used include *Disapprove, Threaten, Protest, Coerce* and *Assault* as we hypothesized that these event types are more likely to lead to huge social unrest.

3.2 Social Unrest Events Source

Wikipedia was used to identify major social unrest events in the US from 2015–2017 These events are referred to as ‘huge social unrest events’ throughout the paper. These events were verified using various news sources.

3.3 Creation of Mapping

A mapping here is a time and location representation of USA. The mapping points have two attributes- location and time. Different locations at various points of time have distinct behavior. Hence the sample points were taken every day to train and test the model. A sample window of one month was considered and a performance window of one month was used. Observation windows of values one, three and six months were also used (Fig. 2). 2918 counties on all dates in period 2015–2017 were taken as sample points. The data contained 1,049,840 points. If an unrest event occurred within the performance window of the data point, it has been marked as an event point. Other data points are marked as nonevent points. Unrest events were identified from the list of major unrest events in the USA in the period from 2015 to 2017 extracted with location and time attributes from Wikipedia. Final Data on which model was trained contained 1265 event points. Nonevent points were under sampled to 8735 randomly to create a balanced dataset.

Fig. 2. Positioning of windows with respect to sample point

3.4 Extraction of Data

As GDELT adds news articles every fifteen minutes, each article is concisely represented as a row in the GDELT table. A number of processing steps were carried out to represent the GDELT data with location and time attributes of required format. The GKG table contained 7.4 TB of data and events table had 127 GB of data. We used Google BigQuery to subset this data to obtain data for the USA and carried out cleaning like activities to collect entries of themes and event codes of interest. Later we stored Theme and Event tables in Google cloud storage and performed the cleaning steps outlined below using PySpark. GDELT contained all themes discussed in an article as an entry in a single row. We separated these themes into separate entries. The themes that we targeted were displayed in a number of ways, they were either present isolated from other themes, or in conjunction with other themes. For instance, *agriculture* is one of our themes of interest; there are entries that call out *agriculture* specifically, then there are entries where *agriculture* is combined with another them such as *agriculture and food*. We needed to collect all instances where agriculture is present, so we created a mapping with all possible combinations of each theme of interest to make sure we capture all possibilities and not just the exact matches of theme of interest.

The *location* field within the GDELT tables has a format of city/landmark, state and country level. City to county mapping was used to extract data at the county level. Improving this mapping will also enable us to collect more accurate information at County level. Themes with their respective sentiments were cast at county and state levels of location. For every data point of created mapping, we captured the following features for time periods of one, three, and six months prior to the sample date: Total number of times the themes of interest occurred (GKG table), the overall sentiment associated with themes (GKG table), the total number of times the cameo event codes of interest were mentioned (event table) and the sum of Tone associated with them (Event table). A Total of 1446 features were extracted from the GDELT tables.

4 Methods

4.1 Framework

Huge social unrest events can be preceded with the occurrence of a number of different categories of events and a wide range of the social factors/themes. Using the social unrest events that were captured from Wikipedia (and cross referenced with various news sources), we created a mapping as described previously in Sect. 3.3 for all the unrest event points and non-unrest event points for the 2015–2017 period. The location and time where major social unrest occurred was marked event point, while remaining data points were marked as non-event points. Using this mapping, we collected data from cleaned tables of GDELT.

We trained machine learning models on the data using the 1446 features that were captured for the period 2015–2016 as training data. In order to capture how the variation within the features impacted social unrest over time, we used an LSTM model with the 33 features (out of the 1446) that were identified as most significant factors to explain social unrest (based on the results of the Random Forest analysis). We then tested the models on the 2017 data for out of date validation. Results and metrics have been analyzed and discussed in the sections that follow.

In addition, we also predicted locations that would be subject to higher levels of social unrest for the next month as of 12th November 2017. To achieve this, we collected data from January 2015 to 11th November 2017 and trained machine learning models to predict the levels of social unrest. We then identified counties that had the highest unrest levels and after completion of the performance period, we validated the results of our prediction with the events that actually occurred in the county during that one month performance period by reading news articles. Figure 3 summarizes framework of our research as a diagram.

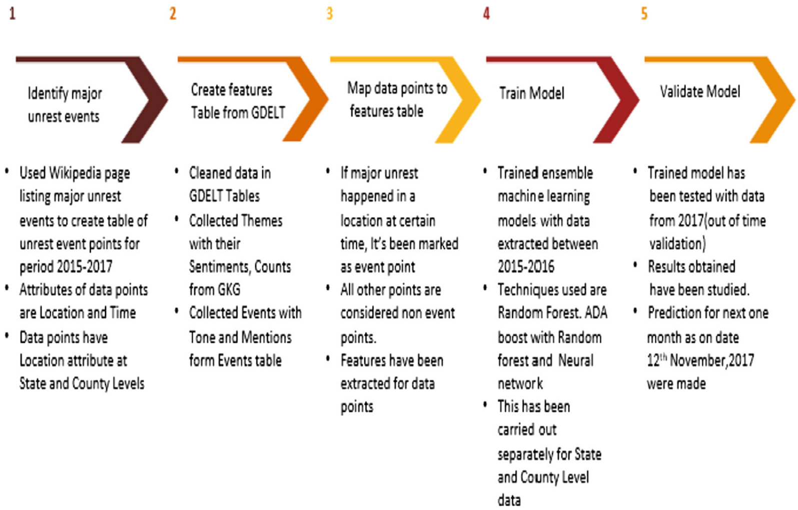

Fig. 3. Diagram explaining framework of analysis

4.2 Models Summary

Random Forest. Random forests are a form of ensemble learning techniques that can be used for classification and regression problems. They operate by constructing a multitude of decision trees at training time and output a predicted class based on a voting mechanism that takes into consideration the most commonly occurring class at each split in the decision tree process. Random forests perform implicit feature selection and provide feature importance which will help us understand the features that have the most significant impact in building of huge unrest event. Having a large data set with many feature variables, we applied this technique to build our model.

ADA Boost with Random Forest. ADA boost is a boosting algorithm developed for classification problems that is less prone to the problem of overfitting. Hence, we explored this technique.

Neural Network. Artificial neural networks are computing systems inspired by the brain's biological neural networks. Neural networks perform particularly well at finding nonlinear relationships in data. We used neural networks to analyze if such patterns exist in our data.

Long Short-Term Memory Networks. Long short-term memory networks are a special kind of Neural Network known as Recurrent Neural Networks (RNNs). They are capable of learning long-term dependencies within a data sequence. In order to capture time variation of the features and their impact on social unrest events, we ran a preliminary LSTM model using 33 most significant features obtained from random forest feature importance. This can be further developed in the future.

5 Experimental Results

5.1 Feature Description and Importance

1446 features were extracted from the GKG and Event tables. Feature importance obtained from the random forest model was used to find top features list. Some of these are *armed conflict, arrest, conflic*t and *violence, corruption* in the Crime category and *alliance, constitution, democracy* in Economy category. To visualize how features vary over time, the Dakota access pipeline unrest event was analyzed at the state level, it is discussed in following paragraphs.

Consider the value of a feature from sample date t to one month prior $t-1$ as F_{t-1}, value from $t-2$ to $t-3$ as F_{t-3} and value from $t-5$ to $t-6$ as F_{t-6}. Let the percentage change in feature from F_{t-3} to F_{t-1} be represented as P2 and percentage change in feature F_{t-6} to F_{t-3} be represented as P1

$$P_2 = \frac{F_{t-1} - F_{t-3}}{F_{t-3}} \times 100 \tag{1}$$

$$P_1 = \frac{F_{t-3} - F_{t-6}}{F_{t-6}} \times 100 \tag{2}$$

Figure 4 is a comparison of how much features vary in case of event points and nonevent points with time, which is a comparison of the value change from P1 to P2. To illustrate this change the Crime armed conflict feature has been used and it can be seen that the feature varies a lot in case of event points compared to nonevent points. This is the case with many top features. Figure 5 depicts how different important features vary with respect to a specific unrest event. The percentage change in the values of features is higher as a major unrest event approaches (P2 > P1). It can also be observed that the feature environment oil count has a high variation from −69.2% to 19.75% indicating news articles have increased coverage of articles talking about oil before a huge unrest event occurred (The Dakota access pipeline unrest incident).

Figure 6 represents the same phenomenon in absolute values where it can be observed that as a social unrest event is building up, the amount of discussion around factors increases (A2 > A1).

$$A_2 = F_{t-1} - F_{t-3} \tag{3}$$

$$A_1 = F_{t-3} - F_{t-6} \tag{4}$$

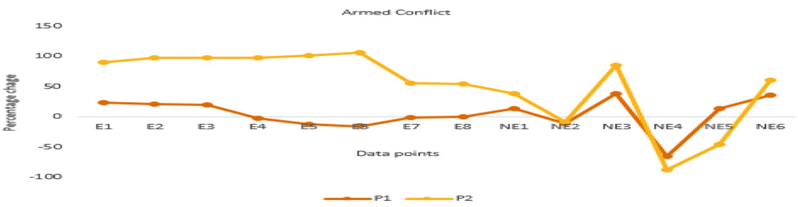

Fig. 4. Comparison of feature variation between event and nonevent points. E is event point and NE is nonevent point

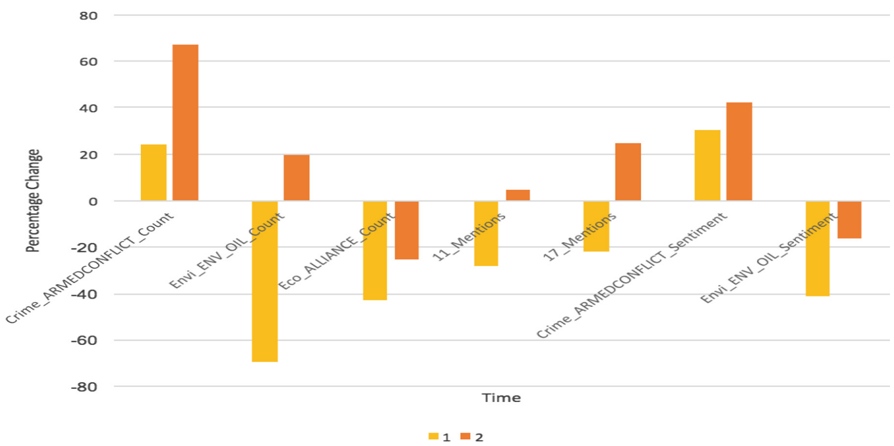

Fig. 5. Comparison amongst features for Dakota access pipeline unrest event.

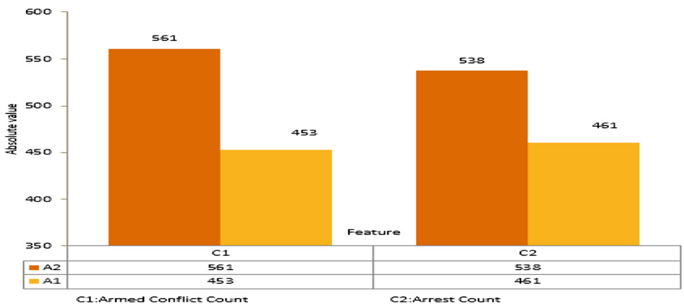

Fig. 6. Comparison of variation of counts of feature with time

To capture this time variation of features, a preliminary model using an LSTM (Long short-term memory network) was run using 33 features that were listed as important by random forest to explain social unrest. Used features are *disapprove* event mentions and tone, *threaten* event mentions and tone, *protest* event mentions and tone, Counts and sentiments associated with themes *protest, armed conflict* and *conflict and* *violence*. Results obtained are below (Table 1). This performance can be enhanced by increasing features considered.

Table 1. LSTM results

Recall	Precision	F1 score
0.78	0.38	0.51

5.2 Metrics

If a nonevent point is being marked as an event point by our model, it is a false positive. If an actual event point is not being detected by our model it is a false negative. The percentage of misclassification is dependent on the probability threshold that is used during the modeling process. If a smaller cut off is chosen, nonevent points might also cross threshold and lead to false positives and if bigger cutoff is chosen some of event points might fail to meet the mark. An optimum cut off is to be selected. We selected a threshold that maximizes the number of event points that are found, while trying to reduce the number of false alerts. The results obtained at the State (Fig. 7) and County (Fig. 8) levels are discussed below. Metrics discussed below are estimated on only out of date validation set.

State Level Results

Random Forest. 90% of Nonevent points were correctly marked as nonevents. 10% of nonevent points were wrongly marked as event points which fall under false positives. 82% of event points were correctly marked as unrest events. 18% of event points were wrongly marked as nonevent points which fall under false negatives. F1 score is the harmonic mean of precision and recall. For the random forest model - Precision is 0.6, Recall is 0.82 and F1 score is 0.69.

ADA Boost with Random Forest. 90% of Nonevent points were correctly marked as nonevents. 10% of nonevent points were wrongly marked as event points which fall under false positives. 69% of event points were correctly marked as unrest events. 31% of event points were wrongly marked as nonevent points which fall under false negatives. Using ADA boost with Random Forest, Precision is 0.55, Recall is 0.69 and F1 score is 0.61.

Neural Networks. 72% of Nonevent points were correctly marked as nonevents. 28% of nonevent points were wrongly marked as event points which fall under false positives. 72% of event points were correctly marked as unrest events. 28% of event points were wrongly marked as nonevent points which fall under false negatives. Using the neural network model, Precision is 0.31, Recall is 0.72 and F1 score is 0.44.

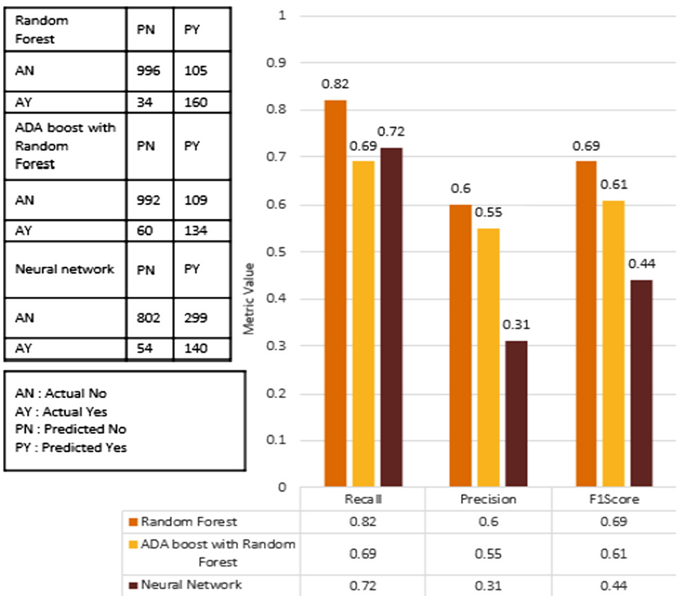

Fig. 7. Model results at state level

County Level Results

Random Forest. 95% of Nonevent points were correctly marked as nonevents. 5% of nonevent points were wrongly marked as event points which fall under false positives. 82% of event points were correctly marked as unrest events. 18% of event points were wrongly marked as nonevent points which fall under false negatives. Using Random Forest, Precision is 0.69, Recall is 0.82 and F1 score is 0.75.

ADA boost with Random Forest. 96% of Nonevent points were correctly marked as nonevents. 4% of nonevent points were wrongly marked as event points which fall under false positives. 79% of event points were correctly marked as unrest events. 21% of event points were wrongly marked as nonevent points which fall under false negatives. Using ADA boost with Random Forest, Precision is 0.76, Recall is 0.79 and F1 score is 0.77.

Neural Networks. 84% of Nonevent points were correctly marked as nonevents. 16% of nonevent points were wrongly marked as event points which fall under false positives. 39% of event points were correctly marked as unrest events. 61% of event points were wrongly marked as nonevent points which fall under false negatives. Using the neural network model, Precision is 0.25, Recall is 0.39 and F1 score is 0.31. It is evident from results that Random forest is performing well with good F1 scores at both State and County levels.

Random Forest	PN	PY
AN	2680	134
AY	68	310
ADA boost with Random Forest	PN	PY
AN	2722	92
AY	78	300
Neural network	PN	PY
AN	2387	427
AY	229	149

AN : Actual No
AY : Actual Yes
PN : Predicted No
PY : Predicted Yes

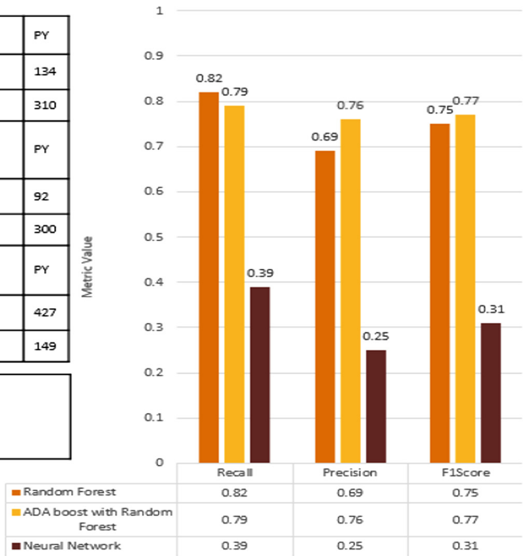

Fig. 8. Model results at county level

5.3 Analysis of Misclassified Cases

To perform out of date validation, Wikipedia data of huge social unrest events in 2017 was considered to test a model that was trained on 2015–2016 data. This contained unrest events related to President Trump, protests against the government on May Day, protest against Milos Yiannopoulos' speech and protests due to improper behavior of police. The majority of these events are due to growing levels of dissatisfaction that the public feel as a result of activities of leaders and government. The values of these themes have been analyzed in the paragraph below for four categories, Event (E), Nonevent (N), False Positive (FP) and False Negative (FN) categories at county level.

A high volume of negative conversations regarding leader and government themes was observed in the news prior to actual unrest event occurrence in associated regions. A high volume of conversation was also observed in locations which are false positives, which could be useful in identifying regions where unrest is boiling up. Not much of a discussion is observed in the case of nonevent points. The average number of times the theme *leader* was discussed is 17174 for event group, 373 for nonevent group, 4554 for false positive group and 123 for false negative group. The average number of times the theme *general government* was discussed was 6614 for event group, 162 for nonevent group, 1684 for false positive group and 51 for false negative group. The average sentiment associated with *local government* is −1.6 for event group, −0.9 for non event group, −1.8 for false positive group and 0.3 for false negative group. false negative cases have lower values like nonevent points and hence are not being detected by model.

Observations made with respect to a selected list of the decision trees is discussed below to understand what the potential root causes of misclassified cases are.

Conditions based on sentiments associated with *armed conflict, protest* and *price*, counts associated with *disapprove* and *coerce* played a prominent role in separation of event and nonevent points. If a particular location in time of interest has conversation about *government* being carried more than 681 times and a greater negative sentiment than -1.82 persisting in society about *armed conflict* and events related to *coercing* and *disapproval* are occurring more than 3116 and 23078 times respectively, it is being marked suspect to an unrest event.

False Positive Analysis. The following observations are also obtained by reviewing how decision action is being performed in a decision tree. 14% of false positive cases have occurred after a major unrest event's occurrence. Hence the model might be taking into account large discussions happening after unrest event as a measure of discontent in society. 85% of false positive cases have negative sentiments associated with *armed conflict* and *protest*. 48% of false positive cases have large count values of *coerce* and *democracy* which is case with event points. If a certain location has negative sentiment associated with particular factors, administrators of that location can act on those areas.

False Negative Analysis. 61% of false negatives have lesser counts associated with features like *disapprove* and *coerce*. 40% of false negatives don't have necessary negative sentiments associated with features like *price*.

5.4 Prediction of Unrest for One Month into Future

We used GDELT data from January 2015 to November 11th 2017 to train models and predicted top suspect states and counties with the high unrest levels for the following month (Nov 12th 2017–Dec 12th 2017). After completion of the performance period, we went through news articles to find out if unrest events actually happened and the findings are listed below the respective tables (Tables 2 and 3).

Table 2. State level predictions.

State name	Probability score
Florida	0.964
California	0.961
Arizona	0.950
Washington	0.945

Hundreds of people protested against the Haitian ruling on 21st November, 2017 in Florida. Anti-Trump protests also happened in Florida on 26th November 2017. Oakland city in California witnessed a strike by 3000 workers from December 5th to December 11th. A number of protests were observed in Washington against Net Neutrality, the GOP Tax bill, LNG plant. No major protests have been verified in Arizona till December 12th.

Table 3. County level predictions.

County name	State name	Probability score
Philadelphia	Pennsylvania	0.167
San Francisco	California	0.166
Dallas	Texas	0.166
Maricopa	Arizona	0.166
Los Angeles	California	0.163
Cook	Illinois	0.162
Bexar	Texas	0.162
Miami-Dade	Florida	0.160

Hundreds of supporters of the imprisoned rapper Meek Mill rallied outside the Criminal Justice Center, Philadelphia on November 13th which was followed after arrest being discussed more than 36,447 times in Philadelphia during observation period of six months with negative sentiment of −4.1 and 2,191,698 mentions of events related to disapproval being registered. Although no major unrest event happened during the performance window in Maricopa, thousands of people protested against Trump in Phoenix on August 23rd 2017. The model might have sensed tension in public about administration. Hundred people in Cook County, Chicago protested on Black Friday on November 24th. Thousands of protestors participated in a Women's rally in Chicago on January 20th, 2018. These events happened after news articles speaking about human right abuses with a horrific sentiment of -5.2 in Cook County. Several groups like SEIU, FANM on January 12, 2018 in Miami Dade county's little Haiti protested against policies and disrespectful talk of government towards immigrant communities like Haitians. Prior to this negative discussion happened in news with sentiment of −3.5 on Hate speech and −2.1 on government in Miami Dade. No major unrest events occurred in Los Angeles, Bexar and Dallas as per our knowledge.

6 Conclusion

As GDELT collects news every 15 minutes from various news sources, events are recorded continuously across the world. This can essentially be used as a tracker in studying events of interest; especially huge social unrest events. Measuring the emotional change of society as it relates to a specific event, and the events that unfold in their wake, is a great instrument to measure the levels of social unrest in various regions. In our paper, themes and events associated with social unrest have been studied using machine learning techniques in an attempt to identify regions at state and county levels where social unrest might occur in near future. It is difficult to manually track all of the factors that influence social unrest and to understand how they vary in different regions, but this task is easier to achieve with machine learning and large volumes of social recordings (news media). As discussed in the sections above, prior to social unrest events occurring, regions suffering from said events generally see a trend of large volumes of negative news articles being published with respect to certain factors leading up to the time that the event unfolds. In being able to predict the factors that indicate when a social unrest event might occur, government officials could roll out proactive safety measures that could reduce the magnitude of the impact that these events have on local communities. We have also recognized some key social factors and event types that appear to play a more significant role in the buildup of social unrest events. This analysis can be applied to a more granular city level, or to the lesser granularity level of country or continent.

Some of the news articles that were included in the analysis reported different associated locations for the same story, and the analysis of themes discussed in those articles were applied to every location that was identified. This is a less than optimal approach and may not necessarily be the most accurate representation of facts but for this analysis we assumed that both locations were related to the theme in question. Missing data was considered to be zero as it was assumed no data implies no event count or sentiment associated with it. Although an attempt has been made to capture time variation of features using LSTM models in our paper, more time series techniques can be applied to improve performance of model. Also in GDELT's event table we have considered events at root level, we can dig deeper to a more described sub-event level. This will be helpful while analyzing which exact events lead to social unrest. The data that was used from the GDELT database contains themes related to economy, environment, crime and health, and as such can be used to find events related to any of these domains. Alternatively, new features can be extracted for any number of new themes that an analyst would want to evaluate. This research can be combined with existing research on social media using Twitter, Facebook and various other social media sites to capture more information in an attempt to enhance the reliability of social unrest prediction. As news articles reflect what is happening in the world, the analysis of this data aided by the power of machine learning and other advanced techniques of artificial intelligence, can help us to understand trends in society that lead to such detrimental events; our research is a step toward achieving this goal.

References

1. Korolov, R., Lu, D., Wang, J., Zhou, G., Bonial, C., Voss, C., Kaplan, L., Wallace, W., Han, J., Ji, H.: On predicting social unrest using social media. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (2016)
2. Mishler, A., Wonus, K., Chambers, W., Bloodgood, M.: Filtering tweets for social unrest. In : IEEE 11th International Conference on Semantic Computing (ICSC) (2017)
3. Xu, J., Lu, T.-C., Compton, R., Allen, D.: Civil unrest prediction: a tumblr-based exploration. In: Kennedy, W.G., Agarwal, N., Yang, S.J. (eds.) SBP 2014. LNCS, vol. 8393, pp. 403–411. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-05579-4_49
4. Korkmaz, G., Kuhlman, C.J., Marathe, A., Ramakrishnan, N., Vullikanti, A.: Forecasting social unrest using activity cascade. PLoS ONE **10**(6), e0128879 (2015). https://doi.org/10.1371/journal.pone.0128879
5. Qiao, F., Li, P., Zhang, X., Ding, Z., Cheng, J., Wang, H.: Predicting social unrest events with hidden Markov models using GDELT. Discrete Dyn. Nat. Soc. **2017** (2015). https://www.hindawi.com/journals/ddns/2017/8180272/