Spatio-Temporal Signal Recovery from Political Tweets in Indonesia

Anisha Mazumder, Arun Das, Nyunsu Kim, Sedat Gokalp, Arunabha Sen, Hasan Davulcu
School of Computing, Informatics and Decision Systems Engineering
Arizona State University
Tempe, Arizona - 85287
Email: {Anisha.Mazumder, adas22, nkim30, Sedat.Gokalp, asen, hdavulcu}@asu.edu

***Abstract*—Online social network community now provides an enormous volume of data for analyzing human sentiment about people, places, events and political activities. It is increasingly clear that analysis of such data can provide great insights on the social, political and cultural aspect of the participants of these networks. As part of the Minerva project, currently underway at Arizona State University, we have analyzed a large volume of Twitter data to understand radical political activity in the provinces of Indonesia. Based on analysis of radical/counter radical sentiments expressed in tweets by Twitter users, we create a Heat Map of Indonesia which visually demonstrates the degree of radical activities in various provinces of Indonesia. We create the Heat Map of Indonesia by computing (i) the *Radicalization Index* and (ii) the *Location Index* of each Twitter user from Indonesia, who has expressed some radical sentiment in her tweets. The conclusions derived from our analysis matches significantly with the analysis of Wahid Institute, a leading political think tank of Indonesia, thus validating our results.**

***Index Terms*—radical, tweet, Radicalization Index, Location Index, Heat Map**

I. Introduction

The sheer popularity of online social media nowadays is reflected by the immense amount of data being fed every second by people from all over the world. It is becoming increasingly evident that analysis of this huge online dataset can provide great insights on the social, political and cultural aspect of the Twitter users and possibly the non-Twitter users as well. In [2], the authors have developed *Socioscope*, a tool for extracting signal from noisy social media data. Utilizing a Socioscope like mechanism, we have developed a tool for recovering spatio-temporal signals from tweets generated in Indonesia. Our interest in analyzing tweets from Indonesia developed in the context of the Minerva[1] project, currently underway at Arizona State University. The goal of this project is to increase the understanding of movements within Muslim communities towards *radicalism* or *counter radicalism*. Based on the *support* and *opposition* of certain *beliefs* and *practices* of an individual (as expressed in her tweet), we can assign a *Radicalization Index* to that individual. In addition, from the self declared *home location* of a Twitter user and the locations of her tweets, we can compute a distribution of *Location Index* for that user. The map of Indonesia is divided up into a set of *regions* and the *Location Index* of a user provides the probability of the user to be in a specific *region* at a specific time. For this analysis a *region* corresponds to a province of Indonesia. Finally, from the *Radicalization Index* and *Location Index* of individuals, *Heat Index* of a *region* , which is a composite measure of the number of radical tweeters of that *region* and their 'degree of radicalism', is computed.

In our model we have a set of tweeters (or users), $U = \{U_1, U_2, \ldots, U_n\}$. Each user $U_i, 1 \leq i \leq n$ creates a set of tweets $T_i = \{T_{i,1}, T_{i,2}, \ldots, T_{i,t_i}\}$. The set of all tweets by all users is denoted by $T = \bigcup_{i=1}^{n} T_i$. The geographic area from where the tweets originate is divided into a set of *regions* $R = \{R_1, R_2, \ldots, R_m\}$. In our study m is equal to thirty four, the number of provinces and special administrative regions of Indonesia. Each user $U_i, 1 \leq i \leq n$ has a *home location* $HL_i, 1 \leq i \leq n$ associated with her, which may or may not be declared. Each tweet $T_{i,k}, 1 \leq i \leq n, 1 \leq k \leq t_i$ has a *geo-location* $GL_{i,k}, 1 \leq i \leq n, 1 \leq k \leq t_i$ associated with it. However, $GL_{i,k}$ for some tweets $T_{i,k}$ may not be known as the user U_i might turn her GPS off. Accordingly, we can divide the set of users in four different classes:

(i) Class 1: user U_i whose *home location* is declared and *geo-location* of at least one tweet is known,
(ii) Class 2: U_i whose *home location* is not declared and *geo-location* of at least one tweet is known,
(iii) Class 3: U_i whose *home location* is declared and *geo-location* of none of the tweets are known, and
(iv) Class 4 : U_i whose *home location* is not declared and *geo-location* of none of the tweets are known.

From the input data set (U, T, R), we compute, (i) *Location Index*, L_i of each user $U_i, 1 \leq i \leq n$, (ii) *Radicalization Index*, RD_i of each user $U_i, 1 \leq i \leq n$, and finally, combining L_i and RD_i, we compute (iii) *Heat Index*, H_j of each *region* $R_j, 1 \leq j \leq m$. It may be noted that whereas $RD_i, 1 \leq i \leq n$ is a scalar value, L_i is a vector of size m, $(L_{i,1}, \ldots, L_{i,m})$, where $L_{i,j}$ indicates the probability of user U_i being located in *region* R_j i.e. $L_{i,j}$ indicates the probability of the *Actual home location* of U_i being R_j. Finally, the *Heat Index* H_j of region $R_j, 1 \leq j \leq m$ is computed as $H_j = \sum_{i=1}^{n} RD_i \times L_{i,j}, \forall j, 1 \leq j \leq m$. We thus provide a generic technique for generating time-varying political Heat Maps of a geographical region based on the Twitter data analysis. Throughout this paper we have used '*region* ' and 'location' interchangeably

[1]A project sponsored by the U.S. Department of Defense

to mean an 'Indonesian Province'. It is to be noted that for our calculations, we have considered all Indonesian provinces including special administrative regions such as Yogyakarta and special capital region such as Jakarta.

II. Related Work

Computation of Heat Map of Indonesia requires the computation of the following: First, we compute the *Radicalization Index* of a user U_i by analyzing the content of her tweets. Second, *Location Index* of the user U_i is computed from her *geo-location* containing tweets (if any) and also from her *home location* declared as a part of her Twitter profile (if at all provided). It is to be noted that we do not consider users who have neither of these two sources of location information present.

Identification of the location of users using Twitter data has been quite a focus of recent research. Inferring location from tweets have been pursued by [14], [15], [16]. Studies conducted in [4], [5], [6], [7] combine location information and text from social-network data history to infer various questions such as user preferences and provide recommendations. However, we do not rely on any 'checking in' information for our computations and providing recommendations is not our goal.

We do employ the notion of *regions* - the thirty four provinces of Indonesia are the *regions* of interest for our problem. Thus, 'geo-coding' (the use of gazetteers) is applicable to our problem. However, just as in [3], we too argue that location estimates are multi-modal probability distributions, rather than particular points or *regions*. However, it may be noted that in contrast to [3], we are interested only in Indonesia and in Indonesian provinces - thus our estimate of the location of the user must be the probability of each Indonesian province as the *Actual home location* of the user under consideration, rather than the probability of the user being located in each and every point on the surface of the earth. This implies that our world comprises of Indonesia only and individual geo-co-ordinates are bunched into the corresponding province of Indonesia. As a result, we apply the combination of 'geocoding' and the modification of the techniques in [3]. Thus, we use gazetteers for the *Declared Home Location* of the Twitter Users to map those to a specific province of Indonesia (This is explained in further details in Section VII). This combined with the geo-co-ordinate information about the user (obtained from her tweets containing geo-location) gives us the probability distribution of the user across the thirty four provinces of Indonesia. We thus obtain a simple yet effective means of computing the geo-location of the user as compared to other more complex methodologies such as Topic Detection Techniques [20], [21], [22].

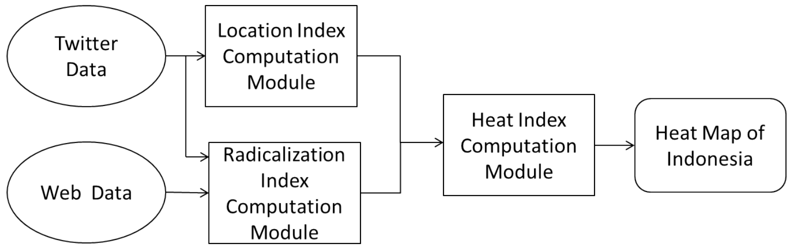
Human mobility is modeled as a stochastic process in [8]. Following the studies of [8], in [1], the authors study the manner in which the movements of human beings are related to time of the day, geography as well as social ties. They intend to predict the exact location of a person based on various factors which the authors have identified, including impact of social network. Similar problems have been studied by [9], [10], [11]. However, in our problem, there is no notion of prediction of location of users involved. Besides, we consider categorical distribution. However, we do use the concept of mixture of distributions in the lines of [1].

Another line of research which focuses on location estimation by content-analysis of the tweets of a user has been studied by [12], [13]. They use the techniques of feature selection from tweets of users, following it up with training and classification. However, we do not apply content based analysis in this current work, but rely on the *geo-location* containing tweets of users in our dataset and also the user declared *home location* to obtain the location distribution of the users.

In [17], the authors analyze tweets generated during the United Kingdom 2010 General Election to measure political sentiments as well. They have identified the specific features of the political parties and their ultimate goal is to infer the political affiliation of a user based on her tweets. We also study a similar problem, however our goal is not to identify the political affiliations of users, rather we compute the 'degree of radicalism' of the user. Besides, our technique is completely different from theirs. Unlike them, we apply a very simple yet effective term-frequency analysis of tweets and leverage heavily on our team of domain experts. We validated our classification of users into radicals and counter radicals by classifying some well-known counter radical leaders of Indonesia (Our validation process is discussed in further details in Section IX).

The work in [35] which is followed by [18] is very relevant to our technique of *Radicalization Index* assignment to users. These works too deal with the recovery of radical signals from the online posts of social media users and thereby identify individuals as potential 'lone wolf terrorists'. They specifically focus on presenting a framework for combining entity matching techniques for detecting extremist behavior on discussion boards. These 'lone wolf terrorists' might leave weak signals of radicalism through their comments or posts on discussion boards, signals made further weaker by the use of aliases. Identification and analysis of such weak signals of radicalism by the use of topic-filtered web harvesting as well as application of natural language processing techniques, thereby fusing aliases for identifying the person form the basis of the works of [35]. Their work is fundamentally different from ours because we deal specifically with the users' publicly available tweets only - this eliminates the availability of the vital background information such as characteristic ('radical internet forum', 'capability internet forum') annotation of particular discussion boards that is leveraged in [35] . However, we also have used the technique of application of keyword analysis and crawling of web-sites of well-known radical/ counter radical organizations of Indonesia as discussed in later sections. Furthermore, [35] and [18] do not deal with location profiling of users which is one of the two major goals of our work.

Fig. 1: The flow diagram of our Heat Map computation technique. The Web data mentioned here refers to the documents generated by crawling the web pages of radical and counter radical organizations of Indonesia.



III. Motivation and Distinguishing Features of our work

The motivation for our work is to provide a visualization of the spatio-temporal distribution of the radical population of Indonesia by recovering political signals from Twitter data. A pictorial description of our methodology is provided in Figure 1. Similar retrieval of signals using Twitter data is the motivation of the work of [2]. In [2], they find the location distribution of tweets mentioning roadkills using human beings as sensors. Similarity of our work with [2] is that we too use human beings as sensors to the extent that we use tweets of people of Indonesia to infer radicalism *Heat Indices* of the provinces of Indonesia. However, our work is significantly different from theirs. First, unlike [2], we intend to find the distribution of (radical) individuals, so we should not factor in any 'human population bias' i.e variation of densities of people across the different provinces of Indonesia. Second, our problem is much more complex because we not only need to know from which location have the radical tweets come in greater number, but also the 'degree of radicalism' of the tweets - so we need to comprehend the sentiment of the tweets. The major difference here is that in our case, we need a finer grained distinction among the radical tweets specifying which tweets are more radical and which tweets are less. So, questions of interest for us are-

(Qs1) the 'degree of radicalism' of tweet *tw*

(Qs2) the originating location of tweet *tw*

Thus, *Heat Index* of a region factors in both the count of the radical tweets from the region as well as the 'degree of radicalism' of the tweets. However, there are certain challenges in answering these questions. As for Qs1, a tweet can at most be 140 characters long. This is indeed too little information to ascertain the 'degree of radicalism' of tweets on individual basis. Thus, we go one level up the hierarchy and consider individual users instead of individual tweets and try to answer the two questions in the context of individual users. We collect all the tweets from individual users and assign the 'degree of radicalism' to the user based on her tweets. Now, Qs2 would have been easy to answer with respect to individual tweets if all the tweets had geo-co-ordinate information because Twitter API[2] provides *geo-location* information of tweets if the user had chosen to reveal her location at the time of tweeting. However, there are certain problems with this approach - first, the percentage of tweets containing *geo-location* information is very scarce (such tweets constitute less than 1% of our dataset). Second, when we consider individual users, it is unjustified to assume that all her tweets containing *geo-location* information will point to a single region, even if all her tweets contained geo-location information. Thus, the best estimate of the location of the user is the probability distribution of the user's location over the Indonesian provinces.

We consider categorical distribution of the users into the thirty four provinces of Indonesia. The motivation behind employing categorical distribution instead of say Gaussian distribution over the entire landscape of Indonesia is that we want to obtain a political Heat Map of Indonesia with the granularity level of a province. Another possibility, that is feasible however not pursued by us in this current work, is dividing Indonesia in the form of grids with varied granularity. Finer granularity poses the problem of insufficient data from every grid, because, as mentioned previously, most tweets do not contain *geo-location* information. So, most grids will have no *geo-location* containing tweet. Our technique of *Location Index* computation is discussed in further details in the following section.

Sentiment Analysis using social media data has been attempted by works such as [31] which tries to exploit patterns in online social media communication and also by [32] which uses background lexical information and refining of the same for specific domains by supervised learning techniques. However, we have computed *Radicalization Indices* using simpler text regression techniques similar to [33] and [34]. Our technique of *Radicalization Index* computation, which is verified to be quite accurate is discussed in further details in Section V.

In summary, individual Twitter users are our chosen level of granularity - we obtain all the necessary information pertaining to each user. Next we characterize the user based on those information, not only on the radicalization scale but we also obtain a location distribution of the user over the *regions* of Indonesia. Hence, there is no prediction of the location of the user involved as in [1]. It is to be noted that we consider only the users classified as radical by our *Radicalization Index* computation method. Hence, our major contribution is the

[2]https://dev.twitter.com/docs/streaming-apis and https://dev.twitter.com/docs/platform-objects/tweets have been used

development of a robust technique to obtain the political Heat Map of any geographic area.

IV. Location Index Computation

As discussed earlier, each user $U_i, 1 \leq i \leq n$ has a *home location* $HL_i, 1 \leq i \leq n$ associated with her, which may or may not be declared. Each tweet $T_{i,k}, 1 \leq i \leq n, 1 \leq k \leq t_i$ has a *geo-location* $GL_{i,k}, 1 \leq i \leq n, 1 \leq k \leq t_i$ associated with it. However, $GL_{i,k}$ for some tweets $T_{i,k}$ may not be known as the user U_i might turn her GPS off. Even when user U_i has a *Declared Home Location* DHL_i, it may not be accurate. User U_i might intentionally or inadvertently misstate her location. Accordingly, we do not accept the DHL_i at its face value as the *Actual home location* of U_i. Instead, we compute a matrix, which we term as the *general Computed Home Location* matrix $gCHL$, from the entire dataset barring the timespan (month in our case) for which the Heat Map is being generated. The created matrix $gCHL$ is an $m \times m$ matrix where $gCHL_{a,b}, 1 \leq a \leq m, 1 \leq b \leq m$, is the *conditional probability* of the *Actual home location* of a user being *region* R_b, when her *Declared Home Location* is *region* R_a, as learnt from the dataset. The $gCHL$ matrix is computed using the following three steps provided in Algorithm 1. Thus, $gCHL_{a,b}$ is given by:

$$gCHL_{a,b} = \frac{X}{Y}$$

where,

X = The number of tweets in T such that the author of the tweet has *Declared Home Location* as R_a and *geo-location* of the tweet is R_b

Y = The number of tweets in T such that the author of the tweet has *Declared Home Location* as R_a

Let the a^{th} row of the $gCHL$ matrix be denoted by $gCHL_a$. Now *Computed Home Location* vector for the user U_i denoted by CHL_i is assigned the value of $gCHL_a$ if the *Declared Home Location* of U_i is *region* R_a. It is to be noted that the $gCHL$ matrix is *general* (and not user specific) and is computed using the entire Twitter data set comprising all users. From those tweets $T_{i,k}, 1 \leq k \leq t_i$ of user U_i, that contain the *geo-location* information $GL_{i,k}$ (i.e., when the GPS is not turned off at the time of the tweet), we compute the *Computed Geo Location* vector CGL_i of length m, where $CGL_{i,j}, 1 \leq j \leq m$, is the *probability* of the *Actual home location* of user U_i being *region* R_j, as learnt from the tweets of U_i. The $CGL_{i,j}$ is computed in the following way:

$$CGL_{i,j} = \frac{A}{B}$$

where,

A = The number of tweets in T_i whose *geo-location* is R_j and

B = The number of tweets in T_i whose *geo-location* is known

We thus obtain two pieces of information about the *Actual home location* of the user U_i in the form of two distributions: CHL_i and CGL_i, where CGL_i is completely user-specific. However, CHL_i is partially user-specific - it does depend on the user because CHL_i is based on her *Declared Home Location*, but it also depends on the general distribution which depends on the entire population mass. It is evident that both CHL_i and CGL_i are categorical distribution over the thirty four Indonesian provinces. Now, we know that a mixture of discrete distributions over any finite number of categories is just another distribution over those categories. In order to combine CGL_i and CHL_i we obtain a convex combination of the two to obtain $L_{i,j}$ in the following way:

$$L_{i,j} = (1 - \omega_i) * CHL_{i,j} + \omega_i * CGL_{i,j} \quad (1)$$

Now, the *mixture weights* ω_i for the user U_i is learnt from the data itself and is calculated as $\omega_i = |T_i'|/|T_i|$.

$L_{i,j}$ essentially is given by

$$L_{i,j} = |T_i^{''}| * CHL_{i,j} + |T_i^{'}| * CGL_{i,j} \quad (2)$$

which gives equation (1) when normalized by $|T_i| = |T_i^{'}| + |T_i^{''}|$ i.e the total number of tweets posted by the user U_i

where,

- T_i = set of tweets produced by user U_i
- T_i' = subset of T_i and represents the set of tweets by U_i that contains *geo-location* information
- T_i'' = subset of T_i and represents the set of tweets by U_i that do not contain *geo-location* information

The motivation behind this definition of the mixture weight is that for the $T_i^{'}$ tweets which contain *geo-location* information, we consider the user-specific location distribution information inferred from the particular user's *geo-location* containing tweets. However, for the tweets of $T_i^{''}$, we have no location information except for the general information that given a *Declared Home Location* for *any user* U_v in our dataset as R_a, the location distribution for U_v is $CHL_v = gCHL_a$. Thus, if the *Declared Home Location* of U_i is given to be R_a, we consider $CHL_i = gCHL_a$. Evidently, we depend on this semi-user-specific location distribution information for the tweets $T_i^{''}$ of U_i. This simple formulation of $L_{i,j}$ also captures the fact that we rely more on CGL_i than on CHL_i when the number of tweets with *geo-location* information, generated by U_i is high, however if that count is low (or even absent), instead of discarding the particular user's information, we obtain the location distribution of U_i from her *Declared Home Location*. We experimented by using only *geo-location* containing tweets and we saw that the results are far more accurate if we included users of Type 3 - This is intuitively correct because the *geo-location* containing tweets form less than 1% of the entire dataset.

As noted earlier, the set of users can be divided into four different classes. We do not try to compute $L_{i,j}$ values for the users belonging to Class 4. For users belonging to the other three classes, we compute $L_{i,j}$ using equation (1). For the users belonging to Class 3, we obtain ω_i to be zero, as we do not have any *geo-location* data from the tweets to compute ω_i.

Algorithm 1 Counting Algorithm for computation of the *general Computed Home Location* $gCHL$

- Step 1: Initialize $gCHL_{a,b} = 0$, $1 \leq a \leq m$, $1 \leq b \leq m$
- Step 2: For each tweet *tw* in T, increment $gCHL_{a,b}$ if *Declared home location* of the author of *tw* and the *geo-location* of *tw* are R_a and R_b respectively
- Step 3: Make each row $gCHL_a$ of $gCHL$ matrix row stochastic, $1 \leq a \leq m$

V. Radicalization Index Computation

We intend to assign a *Radicalization Index* RD_i to a Twitter user U_i based on the *content* of her tweets. Each tweet can contain up to 140 characters. Thus the *content* of a single tweet does not provide adequate information regarding the user's ideology. We collect tweets from users over a period of time (in our case a month) and for each user U_i we create a document D_i that contains all the tweets T_i of that user, during that period of time. As there exists a one-to-one correspondence between U_i and D_i, by assigning a *Radicalization Index* to D_i, we essentially assign a *Radicalization Index* RD_i to U_i. Classical predictive model Multiple Linear regression [23], [24], [25] fits our application, since it is a dichotomous classification problem with multiple predictor variables, where the predictor variables are the terms of our "vocabulary". Classical classification methods such as Logistic Regression which has applications in a wide variety of domains can also be used for document classification [26]. Thus, Logistic Regression can also be applied for our problem. However, Linear Regression was selected instead of Logistic Regression because it out-performed the Logistic one through 10-fold cross validation. This well-known technique divides the given dataset into 10 segments and then uses 90% of the data (i.e nine segments) as training data and 10% of the data (i.e. one segment) as the test data. Linear Regression showed around 98% of accuracy, but Logistic Regression showed 83-85% of accuracy. The implementation of our approach proceeds in the following way: First, we identify a set of Indonesian political organizations. Next, social scientists in our Minerva team, who are domain experts for Indonesia, hypothesize a classification to label each organization as radical or counter radical based on these organizations beliefs and practices. Using web crawling tools, we download a large number of documents from the web sites of these organizations. We use the term "vocabulary" to mean the set of all unique terms that appear in all documents from all organizations. All the documents of an organization are assigned the same *Radicalization Index* as the *Radicalization Index* assigned to the organization by the domain experts in our team. This set of documents together with their *Radicalization Indices* form the training dataset for our model. After that we use the model to assign a *Radicalization Index* to the document D_i created from the tweets of user U_i. This *Radicalization Index* of document D_i is taken to be the *Radicalization Index* of user U_i.

A. Problem Formulation:

We formulate the problem in a general sparse learning framework and solve the following optimization problem (3) using the techniques from [27] . This is indeed a sparse learning problem because the vocabulary is very large compared to the number of words used in a document.

$$\min_x \frac{1}{2} \|Ax - y\|_2^2 + \frac{\rho}{2} \|x\|_2^2 + \lambda \|x\|_1 \quad (3)$$

where $A \in \mathbb{R}^{s \times p}$, $y \in \mathbb{R}^{s \times 1}$, and $x \in \mathbb{R}^{p \times 1}$

In our application, we have

- A is Document $\times$ Term matrix which is constructed as follows: The *set of terms* $(t_1, \ldots t_p)$ includes all the terms from all the documents by all the organizations, barring the stop words. The size of the vocabulary in this case is p.
  If data is collected by crawling web sites of different organization $(O_1, \ldots, O_q)$ and documents $(d_{i,1}, \ldots, d_{1,r_i})$ are collected from the web site of organization $O_i, 1 \leq i \leq q$, the total number of rows of the matrix A is
  $$s = \Sigma_{i=1}^{q} r_i \ ,$$
  and it has the following structure.

$Document/Term$	t_1	t_2	....	t_p
d_1	....	....	....	....
d_2	....	....	....	....
....	....	....	....	....
d_s	....	....	....	....

- $A_{ij} = term\ frequency$ of the j^{th} term in the i^{th} document such that $A_{ij} \geq 0, 1 \leq i \leq s, 1 \leq j \leq p$.
- $y_i \in \{+1, -1\}$ is the class of each document $D_i, 1 \leq i \leq s$. As indicated earlier, the *Radicalization Index* of a document is the same the *Radicalization Index* of the organization that created that document. Thus, when an organization is labeled as radical (or counter radical) by the domain experts, all the documents pertaining to that organization is marked as $+1$ (or -1). Thus $y_i = +1$ if $D_i, 1 \leq i \leq s$ belongs to an organization marked as radical by the experts, or $y_i = -1$ if $D_i, 1 \leq i \leq s$ belongs to an organization marked as counter radical by the experts.
- x_j is the weight for each term $t_j, 1 \leq j \leq p$. This is the parameter estimated by optimizing the objective function (3). The x_j's thus form the predictor variables of the model.

Let us further clarify the three terms involved in the convex optimization problem:

- $\frac{1}{2} \|Ax - y\|_2^2$ - this first term is related to the sum of the squared errors to fit a straight line to a set of data

points. The objective function (3) thus is the optimization problem of minimizing this sum of squared-errors.
- $\frac{\rho}{2}\|x\|_2^2$ - this term deals with the ridge regression, which is an extra level of shrinkage. We set $\rho = 0$ as we were mainly driven by sparsity.
- $\lambda\|x\|_1$ - this term involving the $L1$ norm deals with the sparsity of the solution vector x. For different values of λ we obtain a solution vector x which represents the weights associated with each term $t_j, 1 \leq j \leq p$ (the same terms which are considered in the A matrix). Some of these weights are positive, some negative (values can be very close to 0). The terms with positive (or negative) weights are the radical (or counter radical) words. The top (ones with weights having high magnitude) radical and counter radical words are presented to the experts for validation. We experiment with several λ values resulting in x vectors of various sparsity until the list of top radical and counter radical words are approved by the field experts.

We use the Matlab implementation of the SLEP package [28] that utilizes gradient descent approach to solve the optimization problem (3). This package can handle matrices of 20M entries within a couple of seconds on a machine with standard configuration. The input to the SLEP package are the values of A, λ, and y. The SLEP model outputs the weight vector x.

B. Assignment of Radicalization Index:

For each time period (in our case one month), each user U_i will be assigned *Radicalization Index* RD_i based on their tweets within that period. This is done as follows:

- As mentioned earlier, each tweet which can only contain a maximum of 140 characters is too insufficient for inferring the radicalism of the user. Hence, from the tweets of each user U_i we form a User Document D_i which is the conglomeration of all her tweets over a period of one month. It is to be noted here that many users choose to tweet quite infrequently, hence even if we collect tweets for one month, a user might have tweeted only once or twice during the entire one month which defeats the purpose of collecting tweets for a month. Hence, we further apply the constraint that we consider only those users who have tweeted at least seven times in a month. The value of this threshold has been arrived at empirically after experimentation with various values of the threshold.
- With the help of the model that has been fitted using the organization documents, we classify the User Documents. Let each User Document D_i which is a $termfrequency$ row be denoted by the row vector t_c of count of terms from our "vocabulary".
- Each user U_i receives a 'score' which we refer to as *Radicalization Index* RD_i of user U_i. RD_i is given by

$$RD_i = t_c.x = \sum_{j=1}^{p} t_{c_j} x_j$$

where p is the size of our vocabulary.

This provides us a time-series of RD_i values for the users, which will make it possible to analyze the transition dynamics for each user. It is evident that a high positive RD_i indicates that U_i is highly radical whereas a high negative RD_i indicates that U_i is highly counter radical.

VI. Heat Index Computation

Once we have obtained the *Location Indices* $L_i, 1 \leq i \leq n$ and *Radicalization Indices* $RD_i, 1 \leq i \leq n$, for all the users U_i , $1 \leq i \leq n$, the *Heat Index* H_j of region $R_j, 1 \leq j \leq m$ is computed as

$$H_j = \sum_{i=1}^{n} RD_i \times L_{i,j}, \forall j, 1 \leq j \leq m.$$

The *Heat Index* H_j for a *region* R_j indicates the degree of prevalence of radical ideologies among the people of R_j by taking into account both the number of radical tweeters living in R_j and also their 'degree of radicalism'. In Table I we present a time-varying Heat Map of Indonesia by computing the map in three different time intervals of October 10 - November 10, November 11 - December 10 and December 11 - January 10. We found a drastic change in the *heat indices* during the interval of November 10 – December 10. But we could not discern any particular event which could have triggered the same.

VII. Data Collection

Since our model requires the computation of both the *Radicalization Index* RD_i as well as the *Location Index* L_i for each user U_i, we followed a two step data collection procedure described as follows:

- For the purpose collecting the training data set for computing the *Radicalization Index*, we crawled the websites of 36 well-known Indonesian organizations which are classified as radical or counter radical by our field experts. A few of the organizations are mentioned in Table II. We crawled the websites of all these different organizations and collected a total of 78,135 documents which after pre-processing and filtering resulted into 49,250 documents. The reason for the reduction from the number of crawled documents to the number of useful documents is that many of the crawled documents did not have any relevant information (for example documents having only advertisements) and hence were discarded during pre-processing. Each of the documents on a average contained 280 words i.e on an average 2880 characters. All documents pertaining to an organization were labeled as radical or counter radical depending on the outlook professed by the organization itself. These were then used for fitting our *Radicalization Index* computation model.
- For our study on recovery of political signals pertaining to trend of radical activities in Indonesia, we chose Twitter as the data collection platform as Indonesia features as one of the top five global market segments of Twitter by reach, and accounts for 19.0% to 20.8% of Twitter's

TABLE I: The table provides the top 5 province or special region names based on their computed Heat Index values (also mentioned alongwith) for October 10 - November 10, November 11 - December 10, December 11- January 10

Province Name	Heat Index	Province Name	Heat Index	Province Name	Heat Index
Jakarta	5.48	Jakarta	16.16	Jakarta	4.71
East Java	2.95	East Java	12.33	Yogyakarta	1.82
West Java	2.68	Yogyakarta	4.53	West Java	1.25
Yogyakarta	1.74	Central Java	3.7	East Java	1.20
Central Java	1.68	West Java	3.39	Central Java	0.69

TABLE II: Table showing some of the well-known radical and counter radical organizations of Indonesia

Radical Organizations	Counter radical Organizations
AdianHusaini	DaarulUluum
PKS	Interfidei
Arrahmah	IslamLiberal
AbuJibriel	NU
EraMuslim	PPIM
HizbutTahrir	Paramadina
MillahIbrahim	LKIS

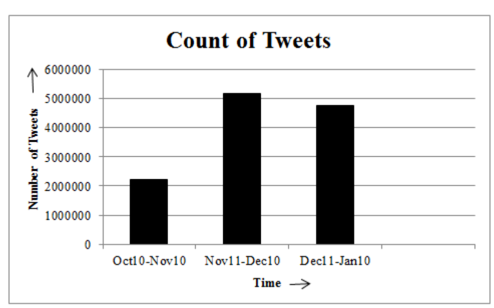
TABLE III: Keyword markers used for filtering Twitter Stream API

Keyword	Interpretation
"penegakan syariah"	enforcement of Sharia
"jihad majelis"	jihad assemblies
"mati syahid"	martyrdom
"ajaran islam"	the teaching of Islam
"pendidikan agama di sekolah"	religious education in schools
"asasi manusia"	human rights
"demokrasi yang"	democracy
"kebebasan beragama"	religious freedom
"sekularisme"	secularism
"di negeri negeri islam"	Islamic state in the country

total reach by country (Dec 2010)[3]. No other publicly available portal offers access to opinions posted online by the Indonesian populace on a similar scale as does Twitter. For gathering tweets, we use Twitter's Stream API to access Twitter's global stream of publicly available tweet data. Since our goal is to recover "political signals", we setup a keyword filter on the Stream API to gather tweets that relate to radical and counter radical ideologies. The keywords used for this filtration have been identified by the social scientists in our Minerva project team and are considered to be significant markers of radical and counter radical ideologies in the Indonesian context. The keyword list includes radical markers and a few such markers are listed in Table III.

We collected tweet data for a three-month interval and gathered a total of 12,152,874 tweets from October 10, 2012 to January 10, 2013 (Figure 2) that matched the keyword filtration criteria (Table III). We used the three months of data to calculate the *Radicalization Indices* of the users. In this research, we are interested in the probability distribution *Location Index* L_i of user U_i over the thirty four provinces of Indonesia, thus we focus only on users from Indonesia. The keywords used are in Indonesian language and narrows down the tweets we obtained from the Twitter API. Thus, the geo-code in majority of cases indicated a location in Indonesia. However, not all geo-codes are from Indonesia. We ignore those tweets in the current work. Thus, out of these 12 million tweets, 110,063 tweets contained *geo-locations* that mapped to *regions* within Indonesia. To apply this *reverse geo-coding*, we used the OpenStreetMap API[4]. A user repository was constructed by including only those users whose *Declared Home Locations* matched with an identifiable Indonesian city or province. Now the user declared *home location* which the user mentions as a part of her profile could consist of any text according to the user's whim. We found texts such as "Dark side of the moon" or "somewhere in this big world" or "Here" or "infront of my laptop" and hence, there is a need for pre-processing of the text. Also, the users provided location information to varied degrees of granularity ranging from continents to towns, however we are interested in the fixed granularity level of Indonesian provinces and the special regions such as Jakarta and Yogyakarta. Hence we manually created a database of towns and cities of all of the Indonesian provinces. Each of the provinces were annotated with 42 cities/ towns on an average with Papua being the highest which was annotated with 70 cities/towns. Using this database we then assigned a legitimate *Declared Home Location* to as many users as possible. The final user repository consisted of 959,911 unique users.

Fig. 2: Figure showing the number of tweets collected over our observation period



[3]http://www.billhartzer.com/pages/comscore-twitter-latin-america-usage/ http://www.comscoredatamine.com/2011/02/the-netherlands-leads-global-markets-in-twitter-reach/

[4]The relevant information about the API could be found at http://wiki.openstreetmap.org/wiki/Nominatim

VIII. Experimental Results

We created Heat Maps of Indonesia on a monthly basis. We computed the RD_i of each user U_i for each month from October 10 to January 10, as long as U_i sent at least 7 tweets in that month Again, for each user U_i we computed the *Location Index* L_i by considering all her tweets over the period of the month.. For that we computed the *general Computed Home Location* $gCHL$ matrix. The$(i,j)^{th}$ entry gives the probability of a user with a *Declared Home Location* of R_i being located in $R_j, 1 \leq i \leq m, 1 \leq j \leq m$.

- $gCHL$ matrix : The $gCHL$ matrix provides interesting insights on the Indonesian population. We computed the $gCHL$ - matrix on all possible doublets among the three months of observation period. i.e for each month for calculating the *Location Indices* L_i of users, we have generated the $gCHL$ matrix using the other two months of data. Thus, in each case, we had training data of two months and test data of one month. Among the three $gCHL$ - matrices generated, we saw that the *Declared Home Location* of users give us a good insight on the *Actual home location* of the user. Thus, instead of merely depending on the geo-co-ordinates of users, we should consider the *home location* from the user's profile and *home location* declarations are much more abundant than *geo-location* containing tweets. However, depending solely on *Declared Home Locations* can be deceptive. We also observed that people with *Declared Home Locations* in various different provinces from all around Indonesia such as Bangka Belitung, Banten, Maluku, West Nusa Tenggara, East Nusa Tenggara and Papua have a very strong tendency to have high probability of having *Actual home location* in Jakarta (as observed from our results over three months). This is very intuitive because Jakarta being the Capital Region must have attracted people from different parts of Indonesia for prospective settlement. We further made an interesting observation that people with *Declared Home Location* of East Kalimantan have considerable *geo-location* containing tweets from Central Kalimantan.

The *Heat Indices* values for the thirty four Indonesian provinces are computed using our approach for three months of our observation period - namely October 10 - November 10, November 11 - December 10, December 11 -January 10. Among all Indonesian provinces the top five provinces and special regions along with their *Heat Index* values are presented in Table I for the three months. Color maps of Indonesia with *Heat Indices* is shown in Figure 3, where darker colors indicate a higher level of radical tweeting, and lighter colors indicate a lower level of radical tweeting. It may be seen from Figure 3 that the area around Jakarta and the Java provinces are highly active in radical tweet creation. According to our Twitter data analysis, the provinces Jakarta, East Java, Yogyakarta and Central Java, along with West Java are the top provinces that generate a high level of radical activities.

IX. Validation

For the purpose of validation of the *Radicalization Index*, we computed the *Radicalization Indices* of some well-known counter radical leaders of Indonesia for the months that they had tweeted for more than 7 times which we consider as our threshold. Our classifier gave perfect accuracy. By accuracy of the classification we mean the percentage of time the leaders who are thus known to be counter radical were classified as counter radical by our classifier. We did not validate the *Location Index* computation technique because of the lack of the ground truth of the *Actual home location* of users. However, our results of Heat Index are validated by the findings of the Indonesia-based Wahid Institute[5] (named after Abdurrahman Wahid, an Indonesian Muslim religious and political leader who served as the President of Indonesia from 1999 to 2001). Wahid Institute promotes a moderate version of Islam through dialogue events, publications, and public advocacies. The institute also releases an annual religious freedom report on religious life in Indonesia. According to the Wahid Institute's Annual Report of 2012[6], the top four provinces of Indonesia where radical activities are most observable are West Java, Aceh, East Java, and Central Java. It may be noted here, that three out of the four most radical provinces identified by the Wahid Institute, also appear at the very top of our list. Also, our field experts have confirmed Jakarta to be a center of radical activities. It may be mentioned here that field studies[7] in January 2012 by Setara Institute[8], a well-known NGO based in Indonesia, showed that the strong radicalism of the young muslim population in Yogyakarta and Central Java are making them hot targets to be recruited as Jihadists. In May 2012, a mob attack by Indonesian Mujahidin Council on a book launch of a well-known Canadian author, an advocate of LGBT, took place in Yogyakarta. In September 2012, there has been arrests of potential terrorists from Yogyakarta[9]. Because, Wahid Institute has mentioned about Indonesian provinces only, it might be expected that Jakarta and Yogyakarta, being special administrative regions, are missing from their list - however, we do not have access to their full report. The high radicalism of the Java provinces are also corroborated by reports of the Setara Institute. The only radically active province that shows up in the Wahid Institute report but does not appear at the top of our list is Aceh, located at the north west corner of Indonesia. It is worth mentioning here that Aceh was completely devastated by the 2004 Indian Ocean Tsunami and is still recovering from its effects. Aceh is also one of the least economically developed provinces of Indonesia. We believe that due to the lack of economic advancement in Aceh, the level of Internet penetration in Aceh is fairly small and not many people from Aceh are active tweeters. This may explain

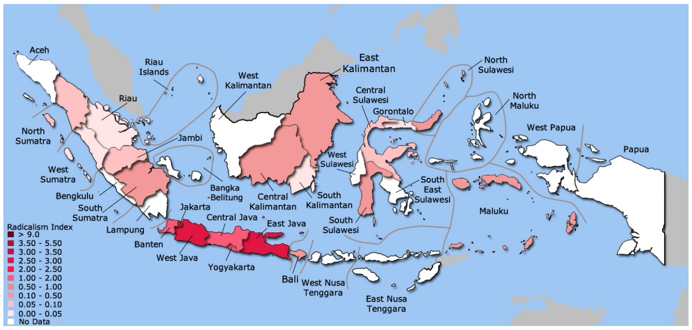
[5]http://berkleycenter.georgetown.edu/resources/organizations/wahid-institute

[6]Released on December 28, 2012

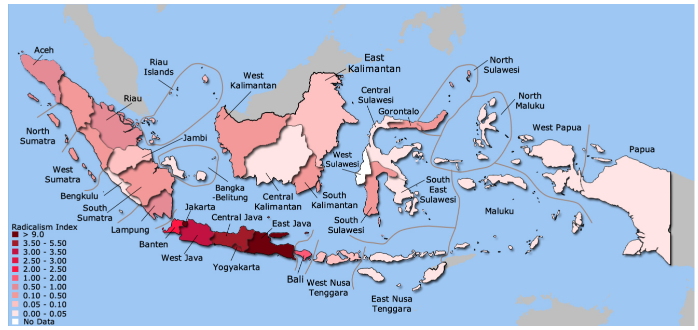
[7]http://www.setara-institute.org/en/content/study-shows-how-young-radical-indonesian-muslims-become-terrorists

[8]http://www.setara-institute.org/

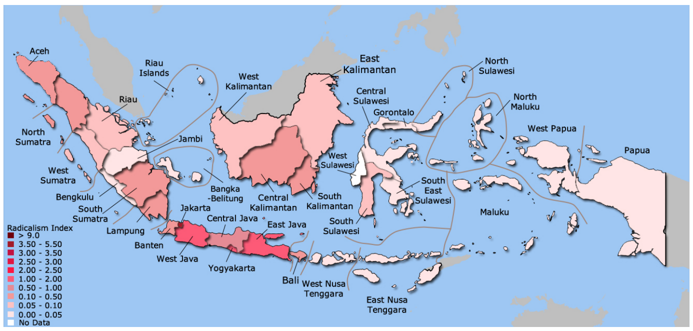
[9]http://www.washingtontimes.com/multimedia/image/indonesia-terrorjpg/



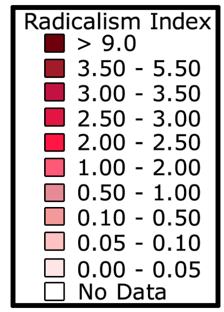
(a) Heat Map for October 10 to November 10



(b) Heat Map for November 11 to December 10



(c) Heat Map for December 11 to January 10



(d) Radicalism Index Scale

Fig. 3: Heat Maps of Indonesia

the reason for Aceh not showing up among our list of top radically active provinces.

X. Conclusion

We have developed a generic robust technique for recovering signals pertaining to a geographical area such as a country using Twitter Data. We have applied our technique to our Indonesian dataset and have observed high accuracy. The goal of our work is the generation of a political Heat Map of Indonesia which will clearly indicate the provinces of Indonesia where radical narrative is prominent. The granularity of the *Radicalization Index* Assignment used is a Twitter user, while the granularity of *regions* used is an Indonesian Province. Thus, we have analyzed tweets made by a user U_i in a month to assign a *Radicalization Index* RD_i to U_i - Thus RD_i indicates how radical is U_i in her political outlook . Also, by mining the tweets in our database we assign a *Location Index* L_i to U_i . For computation of the *Location Index*, the sources of location information used are not only the *geo-location* tagging as provided by Twitter API for the tweets which contain that information (such tweets constitute less than 1% of our dataset), but we also use the user declared *home location* information from her profile (after considerable amount of pre-processing and cleansing). We have combined these two sources of information for inferring the probability distribution of the location of the users among the provinces of Indonesia. Thus, by considering the RD_i's for all the users over a period of one month, in conjunction with the *Location Indices* L_i's we generate the political Heat Maps of the likes of Figure 3. We have got a time series of Heat Maps, which can help us in mapping trends by *regions* in radical discourse. Such Heat Maps can prove to be very useful in studying the spatio-temporal dynamics of the people of Indonesia so far as their political outlook is concerned

References

[1] Cho, E., Myers, S. A., & Leskovec, J. (2011, August). Friendship and mobility: user movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1082-1090). ACM.

[2] Xu, J. M., Bhargava, A., Nowak, R., & Zhu, X. (2012). Socioscope: spatio-temporal signal recovery from social media. In Machine Learning and Knowledge Discovery in Databases (pp. 644-659). Springer Berlin Heidelberg.

[3] Priedhorsky, R., Culotta, A., & Del Valle, S. Y. (2013). Inferring the Origin Locations of Tweets with Quantitative Confidence. arXiv preprint arXiv:1305.3932.

[4] Yang, D., Zhang, D., Yu, Z., & Wang, Z. (2013, May). A sentiment-enhanced personalized location recommendation system. In Proceedings of the 24th ACM Conference on Hypertext and Social Media (pp. 119-128). ACM.

[5] Rahimi, S. M., & Wang, X. (2013). Location Recommendation Based on Periodicity of Human Activities and Location Categories. In Advances in Knowledge Discovery and Data Mining (pp. 377-389). Springer Berlin Heidelberg.

[6] Li, Y., Steiner, M., Wang, L., Zhang, Z. L., & Bao, J. Exploring Venue Popularity in Foursquare.

[7] Li, Y., Steiner, M., Wang, L., Zhang, Z. L., & Bao, J. (2012, December). Dissecting foursquare venue popularity via random region sampling. In Proceedings of the 2012 ACM conference on CoNEXT student workshop (pp. 21-22). ACM.

[8] Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. Nature, 453(7196), 779-782.

[9] Sadilek, A., Kautz, H., & Bigham, J. P. (2012, February). Finding your friends and following them to where you are. In Proceedings of the fifth ACM international conference on Web search and data mining (pp. 723-732). ACM.

[10] Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A tale of many cities: universal patterns in human urban mobility. PloS one, 7(5), e37027.

[11] Lu, X., Bengtsson, L., & Holme, P. (2012). Predictability of population displacement after the 2010 Haiti earthquake. Proceedings of the National Academy of Sciences, 109(29), 11576-11581.

[12] Chang, H. W., Lee, D., Eltaher, M., & Lee, J. (2012, August). @ Phillies Tweeting from Philly? Predicting Twitter User Locations with Spatial Word Usage. In Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on (pp. 111-118). IEEE.

[13] Mahmud, J., Nichols, J., & Drews, C. (2012). Where is this tweet from? inferring home locations of twitter users. Proc AAAI ICWSM, 12.

[14] Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2010, October). A latent variable model for geographic lexical variation. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (pp. 1277-1287). Association for Computational Linguistics.

[15] Cheng, Z., Caverlee, J., & Lee, K. (2010, October). You are where you tweet: a content-based approach to geo-locating twitter users. In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 759-768). ACM.

[16] Hecht, B., Hong, L., Suh, B., & Chi, E. H. (2011, May). Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In Proceedings of the 2011 annual conference on Human factors in computing systems (pp. 237-246). ACM.

[17] Boutet, A., Kim, H., & Yoneki, E. (2012, August). What's in Twitter: I Know What Parties are Popular and Who You are Supporting Now!. In Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on (pp. 132-139). IEEE.

[18] Dahlin, J., Johansson, F., Kaati, L., Martenson, C., & Svenson, P. (2012, August). Combining Entity Matching Techniques for Detecting Extremist Behavior on Discussion Boards. In Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on (pp. 850-857). IEEE.

[19] Eisenstein, J., Ahmed, A., & Xing, E. P. (2011, June). Sparse additive generative models of text. In International Conference on Machine Learning (ICML).

[20] Hong, L., Ahmed, A., Gurumurthy, S., Smola, A. J., & Tsioutsiouliklis, K. (2012, April). Discovering geographical topics in the twitter stream. In Proceedings of the 21st international conference on World Wide Web (pp. 769-778). ACM.

[21] O'Connor, B., Eisenstein, J., Xing, E. P., & Smith, N. A. (2010). Discovering demographic language variation. In Workshop on Machine Learning for Social Computing at NIPS.

[22] Yin, Z., Cao, L., Han, J., Zhai, C., & Huang, T. (2011, March). Geographical topic discovery and comparison. In Proceedings of the 20th international conference on World wide web (pp. 247-256). ACM.

[23] Zhang, T. (2009). Some sharp performance bounds for least squares regression with L1 regularization. The Annals of Statistics, 37(5A), 2109-2144.

[24] S.Kim, K.Koh, M.Lustig, S.Boyd and D. Gorinevsky, An Interior-Point Method for Large-Scale l1 - Regularized Least Squares Journal of seleccted topics in Signal Processing, Vol. 1, No. 4, pages 606 - 617, Dec 2007

[25] Kolter, J. Z., & Ng, A. Y. (2009, June). Regularization and feature selection in least-squares temporal difference learning. In Proceedings of the 26th Annual International Conference on Machine Learning (pp. 521-528). ACM.

[26] Brzezinski, J. R., & Knafl, G. J. (1999). Logistic regression modeling for context-based classification. In Database and Expert Systems Applications, 1999. Proceedings. Tenth International Workshop on (pp. 755-759). IEEE.

[27] Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1), 1.

[28] J. Liu, S. Ji and Jieping Ye. SLEP: Sparse Learning with Efficient Projections, Arizona State University (2009)

[29] Lee, S. I., Lee, H., Abbeel, P., & Ng, A. Y. (2006, July). Efficient L˜ 1 Regularized Logistic Regression. In Proceedings of the National Conference on Artificial Intelligence (Vol. 21, No. 1, p. 401). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

[30] T. P. Minka, A comparison of numerical optimizaers for logistic regression, Technical report, 2007

[31] Thelwall, M., Buckley, K., Paltoglou, G., Skowron, M., Garcia, D., Gobron, S., ... & Holyst, J. A. (2013). Damping Sentiment Analysis in Online Communication: Discussions, Monologs and Dialogs. In Computational Linguistics and Intelligent Text Processing (pp. 1-12). Springer Berlin Heidelberg.

[32] Melville, P., Gryc, W., & Lawrence, R. D. (2009, June). Sentiment analysis of blogs by combining lexical knowledge with text classification. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1275-1284). ACM.

[33] Joshi, M., Das, D., Gimpel, K., & Smith, N. A. (2010, June). Movie reviews and revenues: An experiment in text regression. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 293-296). Association for Computational Linguistics.

[34] Kogan, S., Levin, D., Routledge, B. R., Sagi, J. S., & Smith, N. A. (2009, May). Predicting risk from financial reports with regression. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 272-280). Association for Computational Linguistics

[35] Brynielsson, J., Horndahl, A., Johansson, F., Kaati, L., Martenson, C., & Svenson, P. (2012, August). Analysis of weak signals for detecting lone wolf terrorists. In Intelligence and Security Informatics Conference (EISIC), 2012 European (pp. 197-204). IEEE.