# Real-Time Prediction of Employee Engagement Using Social Media and Text Mining

**Conference Paper** · December 2018

DOI: 10.1109/ICMLA.2018.00225



**6 authors**, including:

N. Sadat Shami
IBM
**52** PUBLICATIONS **734** CITATIONS

SEE PROFILE

2018 17th IEEE International Conference on Machine Learning and Applications

# Real-Time Prediction of Employee Engagement Using Social Media and Text Mining

Abbas Golestani
*IBM*
Armonk, USA
aghadri@us.ibm.com

Mikhil Masli
*IBM*
Armonk, USA
mnmasli@us.ibm.com

N. Sadat Shami
*IBM*
Armonk, USA
sadat@us.ibm.com

Jennifer Jones
*IBM*
Armonk, USA
Jen.Jones@ibm.com

Abhilash Menon
*IBM*
Armonk, USA
abhilash.menon3@ibm.com

Joydeep Mondal
*IBM*
New Delhi, India
jomondal@in.ibm.com

***Abstract*— Employee engagement is becoming a critical component of an organization's success. For many companies, increases in engagement mean increases in productivity. While organizations have long used traditional questionnaires to measure engagement, these leave room for interpretation; and trying to get clarity makes the questionnaires longer, which decreases participation. On the other hand, user-generated content can often act as a barometer on employee engagement, but does not lend itself easily to engagement prediction. We present a model leveraging machine-learning techniques to analyze and predict employee engagement, using both user-generated social analytics and text content from organization-wide surveys, to calculate a real-time engagement level for employees. Our method helped identify specific words and phrases used by individual employees that play an important role in engagement prediction. This study was able to uncover critical themes driving employee engagement and behavior at the enterprise level. Our model for textual analysis cannot only predict engagement, but also provide highly valuable, and actionable, data for business leaders about the health of their organizations.**

***Keywords*— *Text classification, Naïve Bayes, Social media, Employee engagement.***

## I. Introduction

In today's competitive world, employee engagement plays a major role for companies wanting to deliver a strong customer experience. Employee engagement reflects the relationship between an organization's goal and its employees. An "engaged employee" is fully committed to the company's goal and interest. When employees are enthusiastic about their work, they often take personal initiative to positively influence the organization's reputation [1-3]. It has been shown that there is a strong correlation between employee engagement and organizational productivity metrics such as improved performance, higher levels of customer satisfaction, better employee retention, and lower absenteeism rates [2, 3]. These different studies leave little room for doubt about the importance of employee engagement. The challenge is how to measure it effectively. Organizations usually use surveys to measure engagement. There are two major concerns with surveys. First, it is very time consuming to run surveys. Second, it has been reported that surveys are not usually accurate since they are often related to specific time of the year, and more importantly, may not reveal an employee's true opinions [4-6].

One good alternative could be using social media data to perceive employee's engagement [6] since employees are increasingly using social media, expressing their opinions and feelings about work related issues on internal social platforms [7-9]. Proper analysis that can be used for this purpose requires text mining, machine learning and natural language processing techniques to extract required knowledge information. Despite of many studies in textual analysis field [10-12], there is relatively little work on finding relationship of social media activity (text generated by employees) with their engagement [6]. However, other type of text mining applications like sentiment analysis or emotion detection are very popular and widely used in academia and organizations. Analyzing frequency of words from social media has been used for predictions in many domains. For example, researchers developed models to predict stock price movement using sentiment and words used on social media [13, 14]. Additionally, researchers have also tried to predict results of political events like presidential race analyzing social media data [15, 16]. Usually lexicon-based methods and machine learning are the most popular approaches in text mining. The lexicon-based approach as detailed in [17], [18] and [19] uses word counts or dictionaries of words annotated with their semantic orientations. Machine learning based approaches require developing a model by training a classifier with labeled dataset [17], [20]. Researchers have used these approaches in unison as well [20], [21]. Social media data has been used in many other applications such as prediction of individual's depression, well-being and healthcare [22-26]. Textual analysis of employee social media content allows an organization to infer employee engagement from the feeling and thinking that employees reflect on social media. Motivated by success of different studies on using social media data and text mining for prediction purposes [27-32], we presented a model for analyzing and processing the views and experiences of employees of a company reported in the form of status updates, blogs or comments on social media platforms.

In this paper, a machine learning technique has been applied to social media data, to predict employee engagement. Our objective in this study was to conduct a robust and simple test of the effectiveness of our method for identifying important words and phrases for employee engagement prediction. This paper is organized as follows: In section 2, we explain the details of methodologies used for engagement prediction and then in section 3, we present the obtained

978-1-5386-6805-4/18/$31.00 ©2018 IEEE
DOI 10.1109/ICMLA.2018.00225

results from applying prediction method to the social media data.

## II. METHOD

The general idea is to use machine-learning techniques to capture the level of employee engagement from his/her social activity, as a reflection of his/her thoughts and opinions. The machine-learning model tries to connect the dots between the social media data and employee engagement.

### A. Dataset

The study was conducted for IBM, a large multinational company that operates in more than 170 countries and has approximately 400,000 employees around the globe. The company has an internal social media available to all employees [6]. The platform supports main social media features such as status updates, blog posts, and online community forum posts. For this study, we gathered 'public' records from the IBM internal social media anonymously. The social data is available only for people with social activity, which is around 130,000 posts each month by 40,000 employees.

Each year, an employee engagement survey is conducted in the company for better understanding of the employee experience and identifying possible areas where the company can improve. Based on previous studies [2, 6], employee engagement is measured using three criteria: pride ("I am proud to work for [company name]"), work experience ("For me, [company name] is a great place to work"), and advocacy ("I would recommend [company name] to a good friend as a place to work"). Each question was rated on a scale of 1=strongly disagree, to 5=strongly agree. The average of these three questions was the Employee Engagement Index (EEI) [6]. For the purpose of prediction, we constructed two classes based on the EEI: class 1 "Disengaged", is EEI less than 4 and class 2 "Engaged" is EEI equal to 4 and above. Usually about 55% of employees (220,000) respond to the engagement survey. Therefore, there are two kinds of datasets: enterprise social media data and annual engagement survey data.

### B. Algorithm

Our method for employee engagement prediction is based on the Naïve Bayes Multinomial method and an optimization process. The general idea is to find unique set of words and phrases for prediction of employee engagement. These set of words and phrases (linguistic features) represents the "engaged" or "disengaged" behavior of employees. Subsequently these linguistic features could shed light on engagement drivers and barriers and can help business leaders to react and make appropriate decisions accordingly.

To build a model for predicting employee engagement, we collected anonymized 'internally public' social media data for Nov-2015, Oct-2016 and Oct-2017 months for which we also had employee engagement survey responses. After lowercasing post texts, tokenizing texts to unigrams, and removing stop words, we were left with a vocabulary of 68,271 words. Using TF-IDF of words as features, we trained the Naïve Bayes Multinomial classifier to predict binary engaged/disengaged class corresponding to EEI in the survey. The summary of the proposed method has been shown in Fig. 1. Here are the details of proposed method on ground truth data.

The Naïve Bayes Multinomial model uses the frequency of words, not just their presence or absence [33]. In the Multinomial model, a social media post (Snippet) is a sequence of words, coming from a specific vocabulary [34]. Let $x_i$ be the Multinomial model feature vector for the $i$ th social media post Si. The $t$ th element of $x_i$, written $x_{it}$, is the count of the number of times word $w_t$ occurs in snippet $S_i$. Let $n_i = \sum_t x_{it}$ be the total number of words in snippet $S_i$.

Let $P(w_t \mid C)$ be the probability of word $w_t$ occurring in class $C$ (engaged or disengaged), estimated using the word frequency information from the snippet feature vectors. We make the Naïve Bayes assumption, that the probability of each word occurring in the snippet is independent of the occurrences of the other words. We can then write the snippet likelihood $P(S_i \mid C)$ as a multinomial distribution, where the number of draws corresponds to the length of the snippet, and the proportion of drawing item $t$ is the probability of word type $t$ occurring in a snippet of class $C$, $P(w_t \mid C)$.

$$P(S_i|C) \sim P(x_i|C) = \frac{n_i!}{\prod_{t=1}^{|V|} x_{it}!} \prod_{t=1}^{|V|} P(w_t|C)^{x_{it}} \propto \prod_{t=1}^{|V|} P(w_t|C)^{x_{it}} \quad (1)$$

We won't need the normalization coefficient $n_i!/\prod_t x_{it}!$, because it does not depend on the class $C$. The numerator of the right-hand side of this expression can be interpreted as the product of word likelihoods for each word in the snippet, with repeated words taking part for each repetition [33].

The parameters of the likelihood are the probabilities of each word given the snippet class $P(w_t \mid C)$, and the model parameters also include the prior probabilities $P(C)$. To estimate these parameters from a training set of snippets labelled with class $C = k$, let $z_{ik}$ be an indicator variable which equals *1* when Si has class $C=k$, and equals *0* otherwise. If $N$ is again the total number of snippets, then we have:

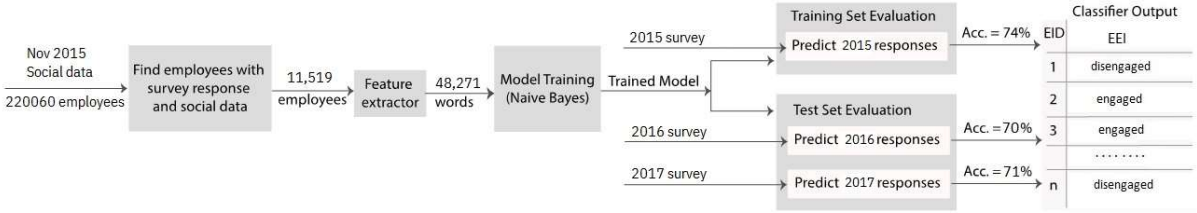


Fig. 1. Summary of model for engagement prediction using social activity and engagement survey data.

$$P(w_t|C=k) = \frac{\sum_{i=1}^{N} x_{it} z_{ik}}{\sum_{s=1}^{|V|} \sum_{i=1}^{N} x_{is} z_{ik}} \qquad (2)$$

an estimate of the probability $P(w_t \mid C=k)$ as the relative frequency of $w_t$ in snippets of class $C=k$ with respect to the total number of words in snippets of that class.

If there are $N$ snippets in total in the training set, then the prior probability of class $C=k$ may be estimated as the relative frequency of snippets of class $C=k$ (let $N_k$ be the total number of snippets of class $C=k$):

$$P(C=k) = \frac{N_k}{N} \qquad (3)$$

Thus, given a training set of snippets (each labelled with a class) and a set of K classes, we can estimate a multinomial text classification model as follows:

1. Define the vocabulary $V$; the number of words in the vocabulary defines the dimension of the feature vectors.
2. Count the following in the training set:
    a. $N$ the total number of snippets,
    b. $N_k$ the number of snippets labelled with class $C=k$, for each class $k=1$ *(disengaged), 2 (engaged).*
    c. $x_{it}$ the frequency of word $w_t$ in snippet $S_i$, computed for every word $w_t$ in $V$.
3. Estimate the likelihoods $P(w_t \mid C=k)$ using (2).
4. Estimate the priors $P(C=k)$ using (3).

To classify an unlabeled snippet $S_j$, we estimate the posterior probability for each class using (1):

$$\begin{aligned} P(C|S_j) &= P(C|x_j) \\ &\propto P(x_j|C)P(C) \\ &\propto P(C)\prod_{t=1}^{|V|} P(w_t|C)^{x_{Jt}} \end{aligned} \qquad (4)$$

The words that do not occur in the snippet (i.e., for which $x_{it}=0$), do not affect the probability (since $p^0=1$). Thus, we can write the posterior probability in terms of words u, which occur in the snippet:

$$P(C|S_j) \propto P(C) \prod_{h=1}^{len(D_i)} P(u_h|C) \qquad (5)$$

Where $u_h$ is the $h$ th word in snippet $S_i$. The snippets will be classified to the class with highest probability. The final step is to aggregate the results at user level to determine if the employee is engaged or disengaged by voting process. For example, a given employee with 5 disengaged snippets (posts) and 4 engaged snippets, will be classified as a disengaged employee. In the case of a tie, the employee will be classified as engaged.

## III. Results

Our objective in this study was to conduct a robust test of the effectiveness of our framework for prediction of employee's engagement and introducing a set of events inferred by algorithm that could help business leaders to react accordingly.

The survey responses provide us with ground truth data for training and testing an engagement prediction model. We built a classifier by training a Naïve Bayes multinomial model under 10-fold cross validation. The classifier was trained on the social media data from Nov-2015 and corresponding EEI class values from the November 2015 survey; this model was tested on two unseen datasets of similar structure pertaining to Oct 2016 and Oct 2017. In preliminary testing, employees' social data during Nov 2015, Oct 2016 and Oct 2017 were monitored since the company engagement survey ran on those times and we were able to evaluate the performance of the engagement prediction model. For comparing the quality of prediction, four measures of accuracy, true positive rate (sensitivity), true negative rate (specificity), precision, and ROC area have been used. The global accuracy shows the percentage of correctly classified samples. The true positive (negative) rate presents the percentage of true classified positive (negative) samples. Precision is also referred to as positive predictive value. Finally, ROC area reveals sensitivity by measuring the fraction of true positives out of the positives and specificity by measuring the fraction of true negatives out of the negatives.

We evaluated the accuracy of our classifier by comparing predicted responses against survey responses from each year's engagement survey. Upon comparing with Nov-2015 survey response data, the training accuracy was 74%. To evaluate the robustness of our classifier, we tested our model accuracy on Oct-2016 and Nov-2017 survey response data. We obtained a predictive accuracy of 70% and 71% respectively for these test datasets. At this stage, we had built a classifier that can predict binary labels of engaged and disengaged for EEI (calculated from the survey questions) with reasonable accuracy.

As a first step in building the classifier, we filtered the set of employees who had both participated in the Nov 2015 survey and had contributed to enterprise social media in the same month – a total of 11,727 employees. We created a table with two columns: text (comment, blog post…) and EEI (engaged or disengaged). We trained the model to predict binary engaged/disengaged class using words and phrases of employee's social activity. For the training dataset (November 2015 social data), using 10-fold cross-validation, our model has a total accuracy of 74%, the two classes being predicted with almost the same accuracy. The accuracy of the prediction on training datasets with 10-fold cross-validation has been shown in Table I.

TABLE I. Results of Engagement Prediction by Naïve Bayes Multinomial Model on Training Dataset.

| Class | TP Rate (Recall) | Precision | ROC Area |
|---|---|---|---|
| **Disengaged** | 0.71 | 0.74 | 0.79 |
| **Engaged** | 0.76 | 0.72 | 0.79 |
| **Total** | 0.74 | 0.73 | 0.79 |

To evaluate the robustness of our model, we tested the model accuracy on Oct-2016 and Oct-2017 survey response data, which are completely separate dataset from the dataset used at training stage but retained similar structure. We obtained a predictive accuracy of 70% and 71% respectively for these test datasets (Table II and Table III), which clearly demonstrated that textual social media data can be used to predict employee engagement with reasonable accuracy on unseen data.

TABLE II. RESULTS OF ENGAGEMENT PREDICTION BY NAÏVE BAYES MULTINOMIAL MODEL ON TEST DATASET (OCT-2016).

| Class | TP Rate (Recall) | Precision | ROC Area |
|---|---|---|---|
| **Disengaged** | 0.68 | 0.71 | 0.75 |
| **Engaged** | 0.72 | 0.69 | 0.75 |
| **Total** | 0.70 | 0.69 | 0.75 |

TABLE III. RESULTS OF ENGAGEMENT PREDICTION BY NAÏVE BAYES MULTINOMIAL MODEL ON TEST DATASET (OCT-2017).

| Class | TP Rate (Recall) | Precision | ROC Area |
|---|---|---|---|
| **Disengaged** | 0.71 | 0.72 | 0.76 |
| **Engaged** | 0.70 | 0.69 | 0.76 |
| **Total** | 0.70 | 0.71 | 0.76 |

Since this model heavily relied on social activity in particular, the text used by individuals and the size of vocabulary used by the model plays an important role. Fig. 2 shows the performance of engagement prediction by the model in relation to vocabulary size. The model reaches a maximum overall accuracy of 74% and it has its best performance at a larger vocabulary size (number of words used).

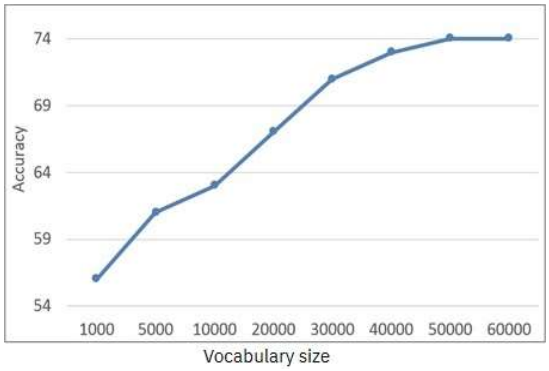


Fig. 2. Performance of engagement prediction in relation to vocabulary size (number of words) on training dataset.

Engagement data usually is reviewed at aggregate level such as business unit or geo since that will allow leaders to be mindful of addressing challenges specific to each segment in a way that creates meaning and purpose. To have an understanding of engagement at aggregate level (like country or business unit), we used a metric called 'Percent favorable' which is percentage of the responses with a positive attitude in that particular segment. In other words, percent favorable is the percentage of engaged label in the target dataset. For this study, 36 segments have been selected to evaluate the performance of proposed method at aggregate level. 18 segments are business units such as Watson, Systems, Cloud… and 18 segments are based on geos such as East Europe, Nordic, Middle East, etc. Our model predicted the engagement scores of 29 segments (out of 36) correctly in 2015. In this case, the total accuracy is about 81%, which shows high level of accuracy at aggregate level. The accuracy of engagement prediction at segment level for 2015, 2016 and 2017 have been shown in Table IV.

TABLE IV. RESULTS OF ENGAGEMENT PREDICTION BY NAÏVE BAYES MULTINOMIAL MODEL ON DIFFERENT SEGMENTS.

| Year | TP Rate (Recall) | Precision | ROC Area |
|---|---|---|---|
| **2015** | 0.81 | 0.8 | 0.84 |
| **2016** | 0.81 | 0.8 | 0.84 |
| **2017** | 0.83 | 0.81 | 0.86 |

We used the same model to predict employee engagement in next three years (2015, 2016 and 2017) as well. As demonstrated, our model achieved high level of accuracy. Since our model had no assumption about the history of individuals' engagement scores and did not use that for prediction, the model has flexibility for prediction and the only feature that matters is current social activity of individuals, which reflects the current thinking and mindset of individuals. We opted for Naïve Bayes classifier because of its robustness to overfitting [33] and its ability to deliver high prediction accuracy in comparison to several other classifiers that we tried, including SVM, Random Forest, Decision Tree and Adaptive Boosting. To control for any spurious overfitting, we re-ran our Naïve Bayes multinomial model with L2-regularization in addition to the 10 fold cross-validation we already had, but did not observe any significant improvement in the prediction accuracy. A key characteristic of Naïve Bayes Multinomial learning methods is a built-in way to control the bias-variance trade-off by managing special parameters introduced to the model. We limit the parameters that can prevent learning algorithms from generalizing beyond their training set. Naïve Bayes Multinomial model performs better particularly on segments with more frequent social activity. The important thing about Naïve Bayes Multinomial is its reliance on social activity of individuals.

The predictive modeling provides a list of words and phrases that have appeared in engaged employee's profile along with the probability associated to each word (or phrase). Our model also provides a list of words and phrases that have appeared in disengaged employee's profile along with the probability associated to each word (or phrase). With this information, we will be able to investigate engagement drivers and barriers since we have access to the topics and phrases that each engaged and disengaged groups have used. Less participation can increase the possibility of error. In addition, the self-censorship can put the prediction in jeopardy. This model can work best in platforms with high social activity and highly transparent environments.

## IV. CONCLUSION

In this study, machine-learning techniques have been used to predict employee engagement from social activity of employees. We demonstrated that employee engagement

could be captured with high level of accuracy using text mining techniques and social media data. Based on our findings, a specific set of words used by individuals could be a strong signal for certain type of engagement behavior. In other words, engaged and disengaged employees can be identified through the way people express themselves on social media. Because of the role of social media on our daily life and consequently rapid growth of social media data, this method allows business leaders to not only have a temperature check of their employees' engagement but also have access to the nature of events and reasons driving the engagement in their company. Finding such a relationship provides a huge opportunity for business leaders to learn more about and interact with their employees to initiate new programs or stop the programs that aren't beneficial according to employee feedback. Results suggest that the present method may achieve a high degree of accuracy in providing organizations with cautionary information regarding unusual engagement behavior.

## REFERENCES


[1] D. MacLeod, and N. Clarke, "Engaging for success: enhancing performance through employee engagement", London: Department for Business Innovation and Skills. Crown copyright, 2009.

[2] J. Wiley, A. Herman, and B. Kowske, "Developing and Validating a Global Model of Employee Engagement", in Handbook of Employee Engagement: Perspectives, Issues, Research and Practice, ed. S. L. Albrecht, Cheltenham: Edward Elgar Publishing Limited, 2012.

[3] O.C. Andrew, and S. Sofian, "Individual factors and work outcomes of employee engagement", Procedia-Social and Behavioral Sciences, Vol. 40, pp. 498-508, 2012.

[4] R. M. Groves, Survey Errors and Survey Costs (Wiley, New York), 1989.

[5] G. F. Bishop, A. J. Tuchfarber, and R. W. Oldendick, "Opinions on Fictitious Issues: The Pressure to Answer Survey Questions", Public Opinion Quarterly 50, 240–250, 1986.

[6] N. S. Shami, M. Muller, A. Aditya Pal, M. Masli, and W. Geyer, "Inferring Employee Engagement from Social Media", Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015.

[7] A. Archambault, and J. Grudin, "A longitudinal study of facebook, linkedin, and twitter use". In Proc. CHI 2012, ACM Press, 2012.

[8] M. Chui, J. Manyika, J. Bughin, R. Dobbs, C. Roxburgh, H. Sarrazin, G. Sands, and M. Westergren, "The social economy: Unlocking value and productivity through social technologies". McKinsey Global Institute, 2012.

[9] H. Zhang, M. D. Choudhury, and J. Grudin, "Creepy but inevitable?: The evolution of social networking". In Proc. CSCW 2014, ACM Press, pp. 368-378, 2014.

[10] D. Vo, and C. Ock, "Learning to classify short text from scientific documents using topic models with various types of knowledge", Expert Systems with Application , Vol. 42 No. 3, pp. 1684-1698, 2015.

[11] H. Feng, Z. Jiang, and J. Shi, "Unsupervised texture segmentation based on latent topic assignment", J. Electron. Imaging. 22, (1 ), 013026. 1017-9909, 2013.

[12] L. Luo, and L. Li, "Defining and evaluating classification algorithm for high-dimensional data based on latent topics", PLoS One 9(1): e82119. doi: 10.1371/journal.pone.0082119, 2014.

[13] V. Pagolu, K. Challa, G. Panda, and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements". International Conference on Signal Processing, Communication, Power and Embedded Systems. SCOPES, pp. 3–5, 2016.

[14] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič, "The effects of Twitter sentiment on stock price returns". PLoS One.10(9):e0138441. doi: 10.1371/journal.pone.0138441, 2015.

[15] L. Kaczmirek, P. Mayr, R. Vatrapu, et al. "Social Media Monitoring of the Campaigns for the 2013 German Bundestag Elections on Facebook and Twitter". DOI: http://arxiv.org/abs/1312.4476, 2014.

[16] M. Cha, H. Haddadi, B. Benevenuto, and K. P. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy". In Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM), 2010.

[17] B. Pang, and L. Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, 1-135, 2008.

[18] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining", WSDM '08 Proceedings of the 2008 International Conference on Web Search and Data Mining, New York, pp. 231-240, 2008.

[19] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis", Journal of Computational Linguistics, vol. 37, no. 2, pp. 267-307, 2011.

[20] S. Bhuta, A. Doshi, U. Doshi, and M. Narvekar, "A review of techniques for sentiment analysis Of Twitter data", International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), pp. 583-591, 2014.

[21] F. Balage, P. Pedro, and T. A. S. Pardo, "A Hybrid System for Sentiment Analysis in Twitter Messages", Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval), pp. 568-572, 2013.

[22] M. D. Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media", In Proc. CHI 2013, ACM Press, 2013.

[23] M. D. Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, "Characterizing and predicting postpartum depression from shared facebook data". In Proc. CSCW 2014, ACM Press, pp. 626-638, 2014.

[24] J. Chen, G. Hsieh, J. U. Mahmud, and J. Nichols, "Understanding individuals' personal values from social media word use". In Proc. CSCW 2014, ACM Press, pp. 405-414, 2014.

[25] M. Dredze, "How Social Media Will Change Public Health". Intelligent Systems, IEEE, 27, 4, pp. 81-84, 2012.

[26] A. Sadilek, H. Kautz, and V. Silenzio, "Predicting Disease Transmission from Geo Tagged Micro-Blog Data". In Proc. AAAI Conference on Artificial Intelligence, 2012.

[27] M. D. Sykora, T. W. Jackson, A. O'Brien, and S. Elayan, "National Security and Social Media Monitoring: A Presentation of the EMOTIVE and Related Systems", European Intelligence and Security Informatics Conference (EISIC), pp. 172-175, 2013.

[28] M. Marc, and C. L. Vincent, "A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter", Information Systems Frontiers, vol. 13, no. 1, pp. 45-59, 2011.

[29] K. Glass, and R. Colbaugh, "Estimating the sentiment of social media content for security informatics applications", IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, pp. 65-70, 2011.

[30] E. De Quincey, and P. Kostkova, "Early Warning and Outbreak Detection Using Social Networking Websites: The Potential of Twitter, Lecture Notes of the Institute for Computer Sciences", Social Informatics and Telecommunications Engineering. Istanbul, Turkey: Springer Berlin Heidelberg, ch. 3, pp. 21-24, 2010.

[31] V. D. Nguyen, B. Varghese, and A. Barker, "The royal birth of 2013: Analysing and visualising public sentiment in the UK using Twitter", IEEE International Conference on Big Data, California, pp. 46-54, 2013.

[32] H. Isah, P. Trundle, and D. Neagu, "Social media analysis for product safety using text mining and sentiment analysis", InComputational Intelligence (UKCI), 14th UK Workshop, pp. 1–7, 2014.

[33] H. Shimodaira, "Text classification using naive Bayes", Learning and Data Note 7, 2014.

[34] A. McCallum, and K. Nigam, A comparison of event models for Naive Bayes text classification. In Proc. of the AAAI-98 Workshop on Learning for Text Categorization, 41-48, 1998.