Empirical Study of Twitter and Tumblr for Sentiment Analysis using Soft Computing Techniques

Akshi Kumar, *Member, IAENG*, Arunima Jaiswal

***Abstract*-Twitter and Tumblr are two prominent players in the micro-blogging sphere. Data generated from these social micro-blogs is voluminous and varied. People express and voice their emotions and opinions over these social media channels making it a big sentiment-rich corpus from which strategic data can be analyzed. Sentiment Analysis on Twitter has been a research trend with constant and continued studies on it to improve & optimize the accuracy of results. As an alternative to the 'restricting' tweets, tumblogs from Tumblr give the 'scrapbook' micro-blogging space. In this paper, we intend to mine and compare these two social media for sentiments to empirically analyze their performance as data analytic tools based on oft computing techniques. We have collected tweets and tumblogs related to four trending events from both platforms (approximately 3000 tweets and 3000 tumblogs) and analyzed the sentiment polarity using six supervised classification algorithms, namely, Naive Bayesian, Support Vector Machine, Multilayer Perceptron, Decision Tree, k-NN, and Fuzzy logic. The results are evaluated for the classifier performance, based on precision, recall and accuracy.**

.

***Index Terms*— Micro-blogging, Sentiment Analysis, Soft Computing, Twitter, Tumblr**

I. Introduction

The new technology paradigm of SMAC (Social media, Mobile, Analytics & Cloud) [1] has revolutionized the way computing is done and also how users get information and engage themselves. This confluence is dominating the IT practices primarily due to the increased popularity of social media & subsequent development of analytic tools. The buzzing term 'Big data' includes data generated from either Social Networks (human-sourced information or Traditional Business systems (process-mediated data) or Internet of Things (machine-generated data) [2].

Social media generates high volume of varied data at a high velocity, thus leading to the 'bigness' in data. People choose to express and voice their emotions and opinions over major social media channels such as blogs, review websites, posts, comments and micro-blogs. A persistent need to leverage this big social data for analytics has been identified by both researchers & practitioners. Sentiment Analysis is one such research direction which drives this cutting-edge SMAC paradigm by transforming text into a knowledge-base.

Formally, Sentiment Analysis, established as a typical text classification task [3], is defined as the computational study of people's opinions, attitudes and emotions towards an entity [4, 5]. These opinions are expressed as written text on any particular topic, available and discussed over social media sources such as blogs, micro-blogs, social networking sites and product reviews sites etc., with the primary intent to categorize content into negative, positive or otherwise neutral polarities.

Amongst the social media channels, Twitter has emerged as a key player from which sentiment-rich data can be extracted. This is primarily due to the characteristics of the real-time messages shared on it. The post size is limited 140 character-set, the users are diverse (Regular user, Public figure, company representative) & globally distributed and the user involvement can range from simply liking, commenting on or re-tweeting the post, making it a leading sentiment rich corpus. Moreover, the easy availability of several Twitter API's and programming services add to making research on sentiment analysis on Twitter has been a significant trend [3, 5, 6].

The techniques for sentiment analysis on Twitter have been categorized into Lexicon Based, Machine Learning Based, Hybrid (Lexicon+ Machine-Learning) and Concept-based (Ontology or context) across pertinent literature [4, 5, 6, 7].

Tumblr is another micro-blogging portal, which came almost around the same time as Twitter but has gained popularity recently due to some value-added features such as posting images, audios, videos, and other media depending on user's knowledge for customizing, managing and uploading such files to create tumblogs (short-form blogs). 'Tumblogging' has not been used much in research studies whereas 'Tweeting' has been the core of most prominent baseline studies.

Moreover, there has been a constant need to improve sentiment classification accuracy with the increase in practical use of sentiment analysis for various analytic domains. The high-dimensional, uncertain social data-space is complex and researchers are keen on applying & testing novel techniques to improve the generic sentiment classification task.

Soft computing has emerged as a significant approach to solve real-world problems which are pervasively imprecise and uncertain. It covers a variety of techniques, namely Machine Learning (Supervised; Unsupervised; Ensemble);

Manuscript received July 16, 2017

Akshi Kumar is a faculty with the Dept. of Computer Science & Engineering, Delhi Technological University, Delhi, India (e-mail: akshikumar@dce.ac.in).

Arunima Jaiswal is a research scholar with the Dept. of Computer Science & Engineering, Delhi Technological University, Delhi, India (e-mail: arunimajaiswal@gmail.com).

Neural Networks; Evolutionary Computation: (Evolutionary Algorithms- Genetic Algorithms, Differential Evolution, Swarm Intelligence -Nature-Inspired Algorithms such as Ant Colony Optimization, Particle Swarm Intelligence); Fuzzy Logic and Probabilistic Reasoning (Naïve Bayesian-Bayesian probability) [8]. Studies to understand the theory, research and practice of using soft computing techniques for sentiment analysis exist but are limited and have majorly considered Twitter as the database. This motivated us towards the work presented in this paper, where we implement and analyze few (not all of these) supervised soft computing techniques for sentiment analysis on both Twitter and Tumblr.

In this paper, we considered tumblogs(re-blogs) and tweets (re-tweets)of four most trending topics in last two years, that is, the US presidential elections (2016), Donald Trump's plans to ban Muslims from the US (2017), Rio Olympics (2016) and release of Pokemon Go second generation (2017) for empirical evaluation and comparison. The extracted data was preprocessed for feature selection and was manually labelled to accomplish coarse grain sentiment analysis (positive, negative or neutral). It was then assessed using six supervised soft computing techniques namely, Naive Bayesian, Support Vector Machine, Multilayer Perceptron, Decision Tree, k-NN, and Fuzzy logic in Weka (while other supervised soft computing techniques can also be evaluated but we selected few to define the scope of work presented in this paper, although the rest can be evaluated in future). The results were evaluated based on efficacy measures like precision, recall, accuracy for probing the capabilities and scope of sentiment analysis within the two micro-blogs.

The paper has been organized as follows. Section 2 provides a brief idea on the key concepts of this work, namely the micro-blogs (Twitter and Tumblr), sentiment analysis and soft computing techniques followed by discussion on the related studies. Section 3explicates the system architecture for this empirical analysis and describes the implementation details, whereas section 4 illustrates the results. Section 5 confers the conclusion and future scope of this work.

II. BACKGROUND WORK

Micro-blogging is the practice of posting small pieces of digital content—which could be text, pictures, links, short videos, or other media—on the Internet [9]. A micro blogging portal connects large number of individuals globally for an easy, constant and engaging information flow creating a sense of online community. There are loads of micro-blogging sites such as Twitter, Tumblr, Plurk, Pownce, Hictu, Friendfeed, Plerb etc. with varied intent such as brand management, customer services or simple family/friends messaging & networking. With the growing popularity of Twitter, and Tumblr catching up, micro-blogging is making it big among bloggers, both general public & personalities.

The growing pursuit of Internet users in the micro-blogging portals has further escalated interest in the area of sentiment analysis, both quantitatively and qualitatively. Sentiment Analysis is a multi-step process encompassing various sub-tasks, namely, Sentiment Data collection; Feature Selection; Sentiment Classification and Sentiment Polarity detection [10].

Feature selection directly impacts the classification accuracy but the high- dimensional, un-structured social media content makes this sub-task even more challenging, fostering the need for improved & optimized techniques for feature selection. Studies exploring and evaluating micro-blogging portals, especially Twitter are available in literature [11, 12, 13, 14].

The challenging aspects of using micro-blogs for sentiment analysis stem from the domains of Natural Language Processing (NLP), text analytics and computational linguistics such as issues related with the fixed text length; spelling variation due to use of short forms like 'gr8' for 'great', 'gud' for 'good'; use of colloquial words, multilingual usage of content in the same tweet or posts; use of emoticons; co-reference resolution; negation handling, sarcasm/ irony/ emotion detection and word sense disambiguation, etc. [3, 15].

The following table I illustrates the basic difference between the two micro-blogging portals: Tumblr, a blog based social media website created in February 2007 and Twitter, a free social networking micro-blogging service that allows registered members to broadcast short posts called tweets, released in March 2006.

In this paper, we have focused on six supervised soft computing techniques namely, Naive Bayesian, Support Vector Machine, Multilayer Perceptron, Decision Tree, k-NN, and Fuzzy logic to perform the task of sentiment classification in two micro-blogging portals, Twitter and Tumblr. In a supervised learning model the data (observations, measurements, etc.) are labeled with pre-defined classes and the test data is classified into these classes.

The training and the testing dataset selection procedure has been done in Weka using 5-fold cross validation method. Table II gives brief information about these six supervised soft computing techniques.

TABLE I
TWITTER VERSUS TUMBLR

Twitter	Tumblr
Twitter is used to publish short bits of text or links.	Tumblr is used to publish content of any kind.
Allows the broadcasting of short messages (including texts, comments, links, quotes, images or video etc.) restricting to single object, single thought	Multimedia and multiple-media support is allowed, that is, users are free to posts texts, images, audios, videos
Tweet text length is restricted to 140 characters	No restriction on text length limit
No such UI customization available	Users can customize themes in Tumblr UI (user interface)
Twitter does keep old tweets, but using the archive is not emphasized.	Tumblogs have attractive views on archives, are searchable and can be organized with tags.
Twitter is less flexible to these features and depends on other	Tumblr is more structured with features like "Notes" which

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

programs which aren't a part of the default offering (it uses HootSuite for queuing) making it a simple micro-blogging only platform	allows a full list of everyone who re-blogged or liked the post and "Queue" which allows to line up posts into your queue and have them posted over time, as you see fit making it social networking-cum- micro-blogging platform
Much simplified user interface	A bit complex user interface with added features and functionalities
Quick adoption by the masses	Users find it quite daunting to use at one go, hence slow adoption by the users
Purely micro-blogging service	Micro-blogging + More comprehensive social networking service
Can be used over variety of devices	Its use for the devices is quite limited

TABLE II
SUPERVISED SOFT COMPUTING TECHNIQUES

Technique	Description
Naive Bayesian (**NB**)	These methods are based on applying Bayes theorem with the "naive" assumption of independence between every pair of features [12, 13]. They require a small amount of training data to estimate the necessary parameters.
Support Vector Machine (**SVM**)	Formally, a SVM constructs a hyper-plane or set of hyper-planes in infinite-dimensional space, which can be used for classification, etc. Intuitively, a good separation is achieved by the hyper-plane that has the largest distance to the nearest training-data point of any class. In general, the larger the margin the lower the generalization error of the classifier [12, 13, 14].
Multilayer Perceptron (**MLP**)	They belong to the class of 'Feed Forward Artificial Neural Network' having input layer, one or more hidden layers and an output layer [13]. The leftmost layer (input) consists of a set of neurons representing the input features. Each neuron in the hidden layer transforms the values from the previous layer with a weighted linear summation followed by a non-linear activation function, for example like the hyperbolic tan function etc. The output layer receives the values from the last hidden layer and transforms them into output values.
Decision Tree (**DT**)	This algorithm breaks down a dataset into smaller and smaller subsets and simultaneously develops an associated DT, having decision nodes and leaf nodes. DT using J48 is the implementation of algorithm ID3 (Iterative Dichotomiser3) developed by the Weka project team which is a top-down, greedy search through the space of possible branches with no backtracking.
k-nearest neighbors algorithm (**kNN**)	It is non-parametric method used for classification where input consists of the k-closest training examples in the feature space. In case of classification, the output is a class membership [14]. An object is classified by majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small.
Fuzzy Logic (**FL**)	The basis of fuzzy logicis to consider the inputs and outputs in the form of fuzzy sets. Each fuzzy set contains of elements that have varying degrees of membership [11]. Fuzzy set enables to transform real number to the membership degrees changing from 0 to 1. Fuzzy rules relate input variables to output variables. These rules represent expert's knowledge in the system. De-fuzzification is the last step over the system which is a mapping process from a fuzzy space defined of output into crisp values. Multi-objective Evolutionary Fuzzy was implemented in Weka.

The term "sentiment analysis" was first observed in the published work in 2003[16] and expansive research efforts in this domain have been observed ever since [5]. Recently the use of various soft computing techniques for sentiment analysis on user- generated textual content on social media predominantly Twitter has been discussed significantly [17]. Fuzzy Logic had been applied for determining sentiment in tweets [11], sentiment analysis on tweets was also carried using machine learning algorithms like Support Vector Machine (SVM), Naïve Bayesian (NB) in [12]. One of the reported work depict the applicability of SVM, NB, Artificial Neural Network (ANN) and ensemble methods like random forest [13] for determining sentiments in tweets. In [14], authors had used k-nearest neighbors (k-NN) and SVM for analyzing the sentiment in tweets. Literature studies reported in sentiment analysis have considered majorly Twitter as a viable dataset due to its intrinsic characteristics. Even the use soft computing techniques for sentiment analysis have been tested on Twitter extensively. The micro-blogging portal Tumblr has been skimpily examined. One of the recent works focusing on the identification of sentiments for multimodal communication on Tumblr for texts, images etc. is reported in [18].

To the best of our knowledge, no previous studies have been done to compare these micro-blogging portals for sentiment analysis. Moreover, use of soft computing techniques for sentiment analysis have been researched and practiced taking Twitter as the dataset, making Tumblr open to application and testing. Thus, this work is a non-trivial effort to empirically contrast Twitter with Tumblr based on employing supervised soft computing techniques for analyzing sentiment from text.

III. SYSTEM ARCHITECTURE

The initial step is to retrieve the desired data about the four top-most trending events from Twitter and Tumblr using their respective APIs. We consider the four most trending topics (over the same duration, to maintain uniformity) in last two years, that is, the US presidential elections (2016), Donald Trump's plans to ban Muslims from the US (2017), Rio Olympics (2016) and release of Pokemon Go second generation (2017) and extract nearly 3000 tweets and 3000 tumblogs. The next step is to pre-process the dataset by removing any noise, incompleteness, inconsistency within it. The data is firstly cleaned by removing #tags, @symbols, URLs, Email ids, punctuations, symbols, numbers, digits, alphanumeric, non-English posts and then transformed removing stop-words, expanding

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

short-forms & replacing emoticons, tokenization and stemming [19]. After pre-processing, we had obtained 2,272 relevant tweets and 1,983 relevant tumblogs (removing redundant, junk tweets & tumblogs).The pre-processed tweets & tumblogs are manually labeled and trained over the classifier (Sentiment Classifier using the selected supervised soft computing techniques). Weka (Waikato Environment for Knowledge Analysis) tool, version 3.8.1 has been used for this empirical evaluation. The following figure 1depicts the overall process flow of the system:

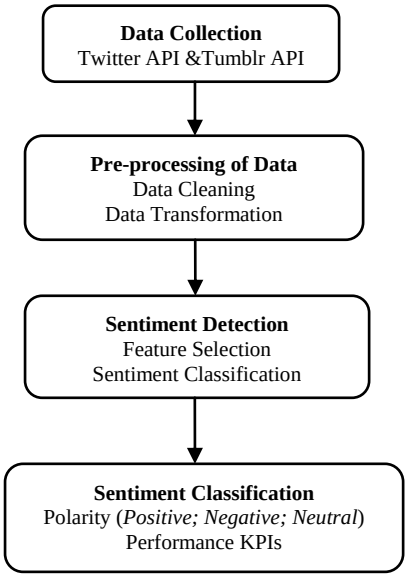

Fig.1. System Architecture

IV. RESULT & ANALYSIS

Key performance indicators (KPIs) have been used to measure and compare the performances of the techniques. We have considered three key performance indicators for this assessment, namely Accuracy (Ac), Precision (Pr), Recall (Re).Accuracy is defined as proximity of a measurement to its true value. It is calculated as a proportion of true positives and true negatives among total inspected cases. Precision defines the exactness of any classifier. A higher precision value indicates fewer 'false positives' and vice versa. It is given as the ratio of true positives to all the predicted positives. Recall defines the sensitivity or the completeness of any classifier. A higher recall value indicates less 'false positives' and vice versa. Recall and precision are bounded by inverse relation with each other [20]. The results of applying the six techniques on the data from both Twitter and Tumblr on the four trending topics are presented in the following tables III, IV, V, VI using KPIs (expressed in percentages).

TABLE III
SUMMARY OF RESULTS FOR 'RIO OLYMPICS'

Techniques	Twitter			Tumblr		
	Pr	Re	Ac	Pr	Re	Ac
NB	74.5	73.5	73.9	60.0	59.3	59.3
SVM	70.0	70.5	70.5	71.7	65.8	65.8
MLP	67.4	68.2	68.2	69.9	64.9	64.9
DT	60.7	61.4	61.4	72.3	58.5	58.45
kNN	62.8	62.9	62.9	67.4	61.2	61.24
FL	26.1	51.1	51.1	26.9	51.9	51.9

TABLE IV
SUMMARY OF RESULTS FOR 'RELEASE OF POKEMON GO GEN'

Techniques	Twitter			Tumblr		
	Pr	Re	Ac	Pr	Re	Ac
NB	59.5	58.6	58.6	68.2	70.2	70.17
SVM	62.0	60.4	60.4	76.6	72.5	72.5
MLP	59.1	57.2	57.2	74.2	72.5	72.47
DT	52.6	53	52.9	44.7	66.9	66.9
kNN	54.1	54.5	54.5	71.4	71.3	71.35
FL	4.7	21.6	21.6	64.1	68.0	67.9

TABLE V
SUMMARY OF RESULTS FOR 'US PRESIDENTIAL ELECTIONS'

Techniques	Twitter			Tumblr		
	Pr	Re	Ac	Pr	Re	Ac
NB	59.2	60.2	60.3	62.6	64.8	64.75
SVM	67.1	64.7	64.7	73.8	70.8	70.75
MLP	59.1	58.4	58.4	69.0	68.9	68.9
DT	57.4	58.0	57.9	49.4	58.0	57.9
kNN	57.56	57.2	57.2	64.9	64.5	64.49
FL	16.1	40.1	40.1	34.2	58.5	58.5

TABLE VI
SUMMARY OF RESULTS FOR 'DONALD TRUMP'S CLAIMS OF MUSLIM BAN'

Techniques	Twitter			Tumblr		
	Pr	Re	Ac	Pr	Re	Ac
NB	72.9	72.5	72.5	68.8	69.5	69.53
SVM	76.2	74.5	74.5	78.2	69.5	69.5
MLP	65.7	67.8	67.8	68.6	68.8	68.75
DT	66.1	65.8	65.8	67.8	62.5	62.5
kNN	60.8	61.1	61.1	62.9	64.8	64.8
FL	20.3	45.1	45.1	34.3	58.6	58.6

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

From these tables, the following observations were made:

The results of our study suggests that the best sentiment accuracy and precision is achieved using Support Vector Machine (SVM) for both Twitter and Tumblr. SVM outperformed all other supervised classification algorithms in terms of Pr and Ac, followed by Naïve Bayesian (NB) and k- nearest neighbor (k-NN) techniques, with the highest Pr for SVM, (approximately 76% for Twitter and approximately 78% for Tumblr). Multi-layer Perceptron Neural Network (NN) also showed encouraging results, quite akin to NB for both the micro-blogs. Next to NN, Decision Trees (DT) had comparable accuracy. Fuzzy logic (FL) demonstrated lower Pr for both Twitter and Tumblr (approximately lower than 30%).

The observed variation in the results is merely due to the fact that large chunks of tweets and tumblogs were skewed towards negative sentiments, predominantly for the topics like Rio Olympics, US presidential elections and Donald Trump's claim for Muslim ban. From the results it is deduced that improved and optimized results were observed for Tumblr in contrast to Twitter. The following graph shown in figure 2 gives the comparison of performance accuracy (Ac) on Y-axis for the six soft computing techniques implemented on X-axis.

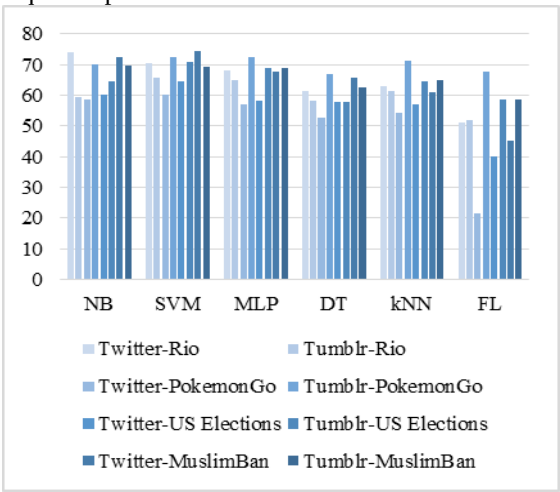

Fig. 2: Comparative Performance Accuracy (Ac)

V. CONCLUSION

This paper empirically contrasted the two micro-blogging portals, Twitter and Tumblr determining the sentiment polarity using six supervised classification algorithms, namely, Naive Bayesian, Support Vector Machine, Multilayer Perceptron, Decision Tree, kNN, and Fuzzy logic. 3000 tweets and 3000 tumblogs related to four trending events from both platforms were analyzed and the results were evaluated for the classifier performance, based on precision, recall and accuracy. The best sentiment accuracy and precision is achieved using Support Vector Machine (SVM) for both Twitter and Tumblr, followed by Naïve Bayesian, k-nearest neighbor, neural network and fuzzy logic which demonstrated the lowest accuracy and precision.

As a future direction, a further improvement in the prediction accuracy of sentiment classification can be tested by employing & evaluating other soft computing techniques like Evolutionary Algorithms (such as genetic algorithms) or hybrid classifiers like Neuro Fuzzy, Genetic optimized Neural Network etc. Also, most of the sentiment analysis research has been done on textual data available across various portals, whereas portals like Tumblr which generate multimedia data can be explored for analyzing the sentiment from audio, video, images, and emoticons. Implementing and empirically contrasting other social platforms like Facebook & Google+ for sentiment analysis using soft computing techniques is another potential direction of research.

REFERENCES

[1] H. I. Alfouzan, "Introduction to SMAC-Social Mobile Analytics and Cloud", *International Journal of Scientific & Engineering Research*, vol. 6, pp. 128-130, Sept. 2015.

[2] S.Vijayarani and S.Sharmila, "Research in Big Data – An Overview", *IEIJ Informatics Engineering International Journal*, vol. 4, no.3, pp. 1-20, Sept. 2016.

[3] A. Kumar and T.M. Sebastian, "Sentiment analysis on twitter", *IJCSI International Journal of Computer Science*, vol. 9, no. 4, pp. 372-378, Jul. 2012.

[4] B. Pang and L. Lee, "Opinion mining and sentiment analysis", *Foundations and Trends in Information Retrieval Journal*, vol. 2, no. 2, pp. 1-135, Jan. 2008.

[5] A. Kumar and T. Sebastian, "Sentiment analysis: A perspective on its past, present and future", *International Journal of Intelligent Systems and Applications*, vol. 10, pp. 1-14, Sept. 2012.

[6] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", in *Proc. 7th Int. Conf. Language Resource and Evaluation*, Malta, 2010, pp. 1320-1326.

[7] B. Liu, *Sentiment Analysis Mining Opinions, Sentiments, and Emotions*. Chicago, Cambridge University Press, 2015.

[8] N.K. Sinha, M.M. Gupta and L.A. Zadeh, *Soft Computing and Intelligent Systems, Theory and Applications*. USA, Academic Press, 2000.

[9] Ulf-Danie Ehlers, *A Guide to Quality, Evaluation, and Assessment for Future Learning*. Germany, Springer, 2013.

[10] A. Kumar, R. Khorwal and S. Chaudhary, "A Survey on Sentiment Analysis using Swarm Intelligence", *Indian Journal of Science & Technology*, vol. 9, no. 39, pp. 1–7, Oct. 2016.

[11] L. Bing and K.C.C. Chan, "Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data", in *Proc. 7th Int. IEEE Conf. Utility and Cloud Computing*, UK, 2014, pp. 652-657.

[12] A.K. Dash, J.K. Rout, and S.K. Jena, "Harnessing Twitter for Automatic Sentiment Identification Using Machine Learning Techniques", in *Proc. 3rd Int. Springer Conf. Advanced Computing, Networking and Informatics*, India, 2016, pp. 507-514.

[13] N. Wang, B. Varghese and P.D. Donnelly, "A Machine Learning Analysis of Twitter Sentiment to the Sandy Hook Shootings", in *Proc. 12th Int. IEEE Conf. e-Science*, USA, 2016, pp. 3-312.

[14] M.R.Huq, A. Ali and A. Rahman, "Sentiment Analysis on Twitter Data using KNN and SVM", *IJACSA International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, pp. 19-25, Jan. 2017.

[15] M.E. Larsen, T.W. Boonstra, P.J. Batterham, Bridianne O'Dea, C. Paris, and H. Christensen, "We Feel: Mapping emotion on Twitter, *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 2168-2194, Feb. 2015.

[16] K. Dave, S. Lawrence and D.M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews",in*Proc. 12th Int.ACMConf. World Wide Web*, Hungary, 2003, pp. 519–528.

[17] B. Sluban, J. Smailović, S. Battiston and I. Mozetič, "Sentiment leaning of influential communities in social networks", *Journal Computational Social Networks*, vol. 2, no. 9, pp. 1–21, Jul. 2015.

[18] E. Bourlai and S.C. Herring, "Multimodal Communication on Tumblr: "I Have So Many Feels!",in *Proc. ACM Conf. WebScience*, USA, 2014, pp. 171-175.

[19] A. Kumar, V. Dabas, "A Social Media Complaint Workflow Automation Tool using Sentiment Intelligence". Lecture Notes in Engineering and Computer Science: *Proc. of The World Congress on Engineering* 2016, pp.176-181

[20] MPS. Bhatia, A. Kumar, "Information Retrieval & Machine Learning: Supporting Technologies for Web Mining Research & Practice", *Webology*, Vol. 5, No. 2, June 2008.

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)