Mining Twitter Big Data to Predict 2013 Pakistan Election Winner

Dr. Tariq Mahmood, Tasmiyah Iqbal, Farnaz Amin, Wajeeta Lohanna, Atika Mustafa
Department of Computer Science
National University of Computer & Emerging Sciences
Karachi, Pakistan
Email: tariq.mahmood@nu.edu.pk, tasmiyahiqbal@gmail.com, farnaz87@gmail.com, wajeeta@gmail.com, atika.mustafa@nu.edu.pk

***Abstract*—Twitter is a well-known micro-blogging website which allows millions of users to interact over different types of communities, topics, and tweeting trends. The big data being generated on Twitter daily, and its significant impact on social networking, has motivated the application of data mining (analysis) to extract useful information from tweets. In this paper, we analyze the impact of tweets in predicting the winner of the recent 2013 election held in Pakistan. We identify relevant Twitter users, pre-process their tweets, and construct predictive models for three representative political parties which were significantly tweeted, i.e., Pakistan Tehreek-e-Insaaf (PTI), Pakistan Muslim League Nawaz (PMLN), and Muttahida Qaumi Movement (MQM). The predictions for last four days before the elections showed that PTI will emerge as the election winner, which was actually won by PMLN. However, considering that PTI obtained landslide victory in one province and bagged several important seats across the country, we conclude that Twitter can have some type of a positive influence on the election result, although it cannot be considered representative of the overall voting population.**

Keywords—Pakistan, Election 2013, Winner, Twitter, Big Data Analysis, Predictive Analytics

I. INTRODUCTION

Twitter is a well-known micro-blogging website which currently allows millions of users to inter-connect through voiced opinions, in the form of fixed-length textual statements called tweets. Although not strictly defined, Twitter has the concept of user *communities* related to different types of discussion topics, e.g., politics, education, research, entertainment, humor, domestic issues etc. The field of Social Network Analysis has concentrated particularly on the detection of such communities and an analysis of their tweet flow, particularly regarding the opinions and sentiments of users. In fact, considering that around 540 million tweets are being generated daily [1], Twitter becomes a perfect application for big data analysis (data mining) regarding the extraction of the generic public opinion on the diverse discussion topics. These opinions can positively impact different types of crucial objectives related to strategic management, governmental policies and reforms, product brand and retail management, customer relationship management, operational optimization etc [2] [3] [4] [5] [6].

An interesting application of Twitter analysis is focused on politics, e.g., gauging the sentiments of users regarding the current government's policies, or the most favorable candidate (or political party) to win an upcoming election. Usually near to or prior to the election date, the users tweet about their favorite candidate and their opponents. Some tweets highlight the events happening around the elections that could be useful in determining the public opinion of the users, the problems they face, the challenges they have, and the hopes that they attach to the future elected government.

The days around May 11, 2013 were quite eventful for the Pakistani users on Twitter, as Pakistan's General Elections were held on this date for the 2013-2018 tenure [9]. These elections represented the full-term completion of one democratically elected government and transfer of power to another one. We observed a significant tweeting rate on this topic starting from January 2013 uptil the election date.

In this paper, we focus on Twitter to analyze the relevant Pakistani Election-based tweets, from January 2013 till 7th May, 2013. Our goal is to determine the election winner based on Predictive Analytics techniques and we selected Twitter because of its previous use in election opinion analysis and electoral prediction in different countries [7] [8] [12] [13] [16] [17] (see Section II). After monitoring of tweets it was revealed that three political parties were more significantly tweeted as compared to others, i.e., Pakistan Tehreek-e-Insaaf (PTI), Pakistan Muslim League Nawaz (PMLN), and Muttahida Qaumi Movement (MQM). We categorized the tweets as either in favor of (Pro) or against (Anti) a given party. After carrying out the relevant tweet processing steps (removal of unnecessary information), we employed Predictive Analytics to learn decision trees [18] for each political party with the Rapid Miner data mining tool. We then employed these trees to accurately predict the election winner based on tweets of 8th May, 9th May, 10th May and 11th May. Although the results showed that PTI will emerge as the election winner, the actual overall winner was PML-N while PTI won comprehensively in the Khyber Pakhtunkhaa province and in several other constituencies. In other words, Twitter can have some type of a positive influence in the electoral results, but it cannot be considered completely representative of the voting population.

This paper is structured as follows. Section II explores work related to Twitter big data analysis for electoral prediction and other applications. Section III elaborates our tweet gathering and preprocessing stages that form a vital component before any analysis activity. Then, Section IV highlights our experimental methodology and Section V discusses the results including interpretation and accuracy of the learnt decision tree models, and the prediction of the winner. Finally, Section VI concludes with a mention of the limitations of our work and future work directions.

II. RELATED WORK

Twitter has been analyzed in the 2008 US presidential elections. Topsy, which a real time social network analytics provider, in a joint work with Twitter, built the "Twitter Political Index" which presented a daily analysis of user's sentiments for various topics related to the elections. Topsy analyzed tweets about both the presidential candidates (Obama and Romney) and calculated their popularity score based on the sentiment present in the users' tweets [10] [11].

A big data analysis project for a Stanford course adopted a related approach for the 2012 US presidential elections [7]. They found that support vector machines are better predictive models than naïve Bayes, and used them to correctly predict Obama as the election winner. Notwithstanding the result, the authors also highlight that Twitter cannot be considered completely self-sufficient in predicting elections. Also, in [16], the authors employ sentiment analysis on tweets to predict the voting percentage, i.e., the percentage of votes, which each presidential candidate will receive in the Singapore Presidential Election 2011. They create sentiment divisions and employ re-weighting techniques on these divisions to predict voting percentage. The results didn't predict the correct winner, and the authors indicate a change in voting trend and fake Twitter sentiment for these discrepancies.

In [12], the authors analyzed about 100,000 tweets containing the name or mention to a political party or politician. This study was done to understand if there was any linkage or similarity between the online political sentiment and offline sentiment about the German Federal Elections. They used Twitter messages to predict the popularity of political parties, candidates and coalitions in the real world. This work also uses sentiment analysis techniques to measure the popularity of the candidates in question. Twitter opinion mining has also been applied to the Belgian elections of 2010. The experiment was conducted on 7600 tweets about Belgian politicians, and positive and negative sentiments were calculated using sentiment analysis techniques. Different analytics were extracted from the study and are elaborated in [13]. Notwithstanding these applications, it has been significantly highlighted in [17] that the predictive accuracy of Twitter in predicting the election winner is quite exaggerated. The authors prove that Twitter analysis often provides a myopic view of the election results, but it definitely cannot replace the traditional voting method.

In this paper, we are not aiming for this replacement – we are simply gauging the opinions of the users based on what data we have, although we already know that Twitter is not that commonly used across Pakistan as in USA, Germany etc. Our results are hence valid across the computer-savvy population which is mainly resident in important cities like Karachi, Lahore, and Islamabad etc. Other than politics, Twitter data has also been used in domains such as healthcare and disease control. The study in [14], has conducted the surveillance of Twitter data for asthma related tweets using NLP and Machine learning techniques for this study. Their models are able to predict with good certainty whether the tweet was made for the symptom of the disease, its medication or its triggers. Also its sentiment could be determined as positive or negative. Statistical machine-learning classifiers including K-nearest neighbors, Naive Bayes and Support Vector Machines were trained on the unigram and bigram models of the data.

III. DATA PRE-PROCESSING

The first step in big data analysis is data pre-processing (data cleaning), which involves crucial steps for preparing the data for analysis, e.g., data extraction, selection, reduction, transformation etc. [18]. Our data pre-processing steps are described below.

A. Data Gathering

Data was gathered using a website called *Twimemachine* [15], which allows you to fetch the latest tweets of a user. However, it only fetches around 3200 latest tweets of a user. This cap was enough for us as we only needed data from January 2013 - May 2013 till the Election Day, that is, May 11. For most users' timelines, this was enough. As mentioned in Section I, we concentrated on tweets related to three political parties which were significantly tweeted, i.e., Pakistan Tehreek-e-Insaaf (PTI), Pakistan Muslim League Nawaz (PMLN), and Muttahida Qaumi Movement (MQM).

Visual Analysis:

After observing, tweets of around 50 relevant users, for around two to three weeks, we decided to fetch tweets of 24 users, 15 of which were political analysts and 9 were normal everyday users. We selected these users based on the content of their tweets, i.e., we deemed their tweets most relevant for electoral prediction. This activity got us approximately 55000 tweets out of which only 9000 were relevant with respect to the attributes we considered important for a tweet to be political. Some of these attributes are highlighted in Figure 1. Many of them were popular *hash tags* that were trending on Twitter during the election period. Others were frequently used terms or names of political parties and politicians.

NA250	ECP	AltafBhai
PTI	Election2013	Altaf Bhai
PMLN	NayaPakistan	ReElection250
MQM	TeamMQM	Victory4MQM
PPP	WaadaByMQM	WazirEAzamNawazSharif
JUI	IK4NayaPakistan	VoteNawazDa
JI	StampOnBat	SherAya
ANP	TeamMQM2013	Vote4Tarazu
IK	iVoteMQM	TeamJI
IMRANKHAN	MyQuaidAltaf	TeamTarazu2013
TEHREEKEINSAAF	MQMCondemns	JI4Karachi
Nawaz Sharif	YomeTashakurMQM	WeRejectPTI
BilawalBhutto	AltafHussain	
Bilawal	Altaf	

Fig. 1. Identified attributes which make a tweet politically relevant

B. Data Cleaning

We downloaded the tweets as raw text files and cleaned them by performing the following steps:

- Removed the word "RT" (denotes re-tweet)
- Removed usernames, written as "@username"
- Removed all URLs
- Removed the following punctuation marks:, " ' ? ! ; : # $ % & ( ) * + - / < > = [ ] \ ^ _ { } | ~
- Removed all dates and timestamps of tweets from the tweet content but stored it in a separate record against the tweet.
- Replaced double and triple spaces between words with single space.
- Removed stop words, but later when constructing the attribute matrix (detailed in next section), this effort seemed unnecessary.

After executing the aforementioned steps, we got a normalized version of most tweets. For instance, the tweet "*RT @NadeemfParacha: A defiant press conference in Karachi by ANP, MQM and PPP. Tue Apr 30 19:39:16 +0000 2013*" was transformed to "*A defiant press conference in Karachi by ANP MQM and PPP*" after cleaning.

IV. EXPERIMENTAL METHODOLOGY

As mentioned in Section I, our goal is to predict the winning political party of the Pakistan 2013 elections. For this, we train predictive models for each party and test them to predict the winning party. Our prediction (classification) labels are Pro and Anti; Pro represents a positive sentiment favoring the party and Anti represents a negative one. We constructed three separate attribute matrices for each political party, and use them to construct predictive models using tweets from 1st January, 2013 till 7th May, 2013. Specifically, we manually visualized and extracted those tweets for each party which contained those attributes that represent either a Pro or Anti opinion. We rejected tweets with a neutral opinion. For example, the PTI matrix contains attributes like "IMRANKHAN" (sentiment for PTI candidate Imran Khan), "TEHREEKEINSAAF" (sentiment for PTI), "IK4NayaPakistan" (support Imran Khan for a New Pakistan), "StampOnBat" (vote for bat which is the electoral image for PTI) etc. Having identified the attributes, we then updated the respective column for an attribute with a '0' if the attribute is not found in the tweet and '1' if the attribute exists. Finally, we labeled each tweet with either Pro or Anti. Table I shows a snapshot of the PTI matrix with both attributes and labels, and Table II shows a snapshot from the MQM matrix with only the labels.

Table I Attribute Matrix and Labeling for PTI

Tweet	PTI	Imran Khan	Naya Pakistan	Stamp OnBat	Label
Clear majority of PTI in Lahores urban areas StampOnBat	1	0	0	1	PRO
PTI Candidate Naeem Shehzad from PS90 arrested in fraud NayaPakistan	1	0	1	0	ANTI
ImranKhan not using Bullet Proof Glass BraveImran PTI	1	1	0	0	PRO

Table II Labeling for MQM

Tweet	Label
when a terrorist org MQM claims to have mandate in Karachi This is what happens killers are roaming around freely	ANTI
going voting ballot boxes havent arrived dad cursing mqm sigh	ANTI
honest young educated dedicated representatives ivotemqm Pakistan	PRO
koi darr nahi bum ka vote sirf mqm ka mqmwontsurrender	PRO

We identified 10 attributes for PTI and PMLN and 13 attributes for MQM. Also, the PTI, PMLN and MQM training matrices contained 318, 362 and 858 tweets respectively.

For training, we used the Rapid Miner tool (http://rapid-i.com/) to experiment with three different standard predictive models, i.e., CHAID decision tree, Naïve Bayes and Support Vector Machine (SVM) [18]. CHAID (Chi-squared Automatic Interaction Detector) is well suited for the analysis of larger datasets. It builds non-binary trees where each non-terminal node identifies a split condition on attributes using the Chi-squared test, to yield optimum prediction of the label. The naive Bayes is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions, i.e., the presence or absence of an attribute is unrelated to the presence or absence of any other feature. Finally, SVM is a non-probabilistic, linear, and a binary predictor which maps a prediction problem into a high-dimensional space to determine the support vectors, i.e., the subset of training data which is most representative to distinguish between the two classes (Pro and Anti). Both naïve Bayes and SVM have shown high predictive performance on many datasets (for more details on these models, refer to [18]). We train our 3 models by using the standard 10-fold cross-validation [18] and compare their predictive accuracy, i.e., the percentage of correct predictions.

After training, we constructed test attribute matrices for each political party. The test matrix for a given party contained the same attributes as listed in its training matrix. The tweets from 8th May 2013 – 11th May 2013 were used in the test matrices. These tweets were pre-processed and the matrices populated with 1 and 0 as for training counterparts. The PTI, PMLN and MQM test matrices contained 121, 123 and 192 tweets respectively. We left the "Label" column blank, as this is predicted by Rapid Miner by using the party-specific predictive model. The party which gives the largest percentage of correct predictions is the predicted winner. Finally, for the sake of increased validation, we manually inserted the labels in each test matrix to compare with the predictions of Rapid Miner.

V. RESULTS AND DISCUSSION

In this section, we discuss our results. Initially, we compare the predictive accuracies of CHAID, naïve Bayes and SVM, which are shown in Table III[1].

Table III Accuracies for CHAID, Naïve Bayes, and SVM

Algorithm	Accuracy	
CHAID	PTI	77%
	PMLN	66%
	MQM	60%
Naïve Bayes	PTI	78%
	PMLN	64%
	MQM	59%
Support Vector Machine	PTI	78%
	PMLN	65%
	MQM	53%

We can see that there is no significant difference in the accuracies of the algorithms, with respect to the political parties. Still, CHAID gives better accuracy for PMLN and MQM. We thus select CHAID because it also outputs prediction rules which help us to understand the positive (Pro) and negative (Anti) sentiments of Twitter users regarding political parties. We enlist them in the following sections.

A. PTI Predictive Rules

The PTI classification rules which are more important for prediction are enlisted below:

- Tweets containing only "IK" are 76% Pro-PTI
- Tweets containing only "PTI" are 91% Pro-PTI
- Tweets containing "PTI" and "Support" are 70% Pro-PTI
- Tweets containing "PTI" and "NayaPakistan", or "PTI" and "IMRANKHAN" are 100% Pro-PTI
- Tweets containing "PTI" and "IK" are 75% Pro-PTI
- Tweets containing "PTI", "IK" and "Support" are 72% Pro-PTI
- Tweets containing "PTI", "IK" and "NayaPakistan" are 100% Pro-PTI
- Tweets containing "PTI", "IK" and "IMRANKHAN" are 100% Pro-PTI
- Tweets containing "PTI" and "Reject" are 100% Anti-PTI

B. PMLN Predictive Rules

The PMLN classification rules which are more important for prediction are enlisted below:

- Tweets containing "Nawaz Sharif" and "vote" are 75% Pro-PMLN
- Tweets containing "pmln" and "vote" are 64% Pro-PMLN
- Tweets containing "pmln" and "join" are 60% Pro-PMLN
- Tweets containing only "Nawaz Sharif" are approximately 50% Pro-PMLN and 50% Anti-PMLN
- Tweets containing "pmln" and "Nawaz Sharif" are 63% Anti-PMLN, while Tweets containing "pmln", "Nawaz Sharif" and "pti" are 63% Pro-PMLN
- Tweets containing "pmln" and "shame" are 86% Anti-PMLN
- Tweets containing "pmln" are 77% Anti-PMLN Tweets containing "pmln" and "pti" are 68% Anti-PMLN
- Tweets containing "pmln", "vote" and "pti" are 70% Anti-PMLN
- Tweets containing "pmln", "support" and "shame" are 100% Anti-PMLN

C. MQM Predictive Rules

The MQM classification rules which are more important for prediction are enlisted below:

- Tweets containing "MQM" and "support" are 63% Pro-MQM
- Tweets containing "MQM" and "Altaf" are 65% Pro-MQM
- Tweets containing "MQM" and "WaadaByMQM" are 100% Pro-MQM
- Tweets containing "MQM" and "TeamMQM" are 97% Pro-MQM
- Tweets containing "MQM" and "join" are 58% Pro-MQM
- A tweet not containing "MQM" is 68% Anti-MQM, while tweets containing only "MQM" are 57% Anti-MQM
- Tweets containing "MQM" and "vote" are 63% Anti-MQM

D. A Comparison of Predictive Rules

From the aforementioned rules, we extract the following important information:

- A large majority of PTI and MQM tweets are in favor of the party, and the Pro percentages for these tweets are higher as compared to those for

[1] All percentages in this paper are rounded to the nearest integer

PMLN. In other words, a high positive sentiment is associated with PTI and MQM, as compared to PMLN.

- There is a higher positive sentiment for PTI as compared to MQM
- PMLN has been primarily criticized by PTI supporters and vice versa (this fact is associated with the occasional vociferous speeches by each party's candidates against the others)
- The word "MQM" (by itself) has largely been used to express negative sentiments about MQM
- The word "PTI" (by itself) has largely been used to express positive sentiments about PTI
- The birth of a new Pakistan ("NayaPakistan") was a prime motivation for Twitter users to express support for PTI
- Users have used the name of the PTI prime minister candidate (ImranKhan) to largely express positive sentiment, and the name of PMLN prime minister candidate (Nawaz Sharif) to express positive and negative sentiment equally; MQM candidate Altaf Hussain was not that much significantly mentioned.
- The promises made by MQM and its unifying spirit for the supporters were primary motivations to express support for MQM.

E. *Precision for Pro and Anti Classes*

Using CHAID, we determine the precision for each of the two classes, i.e., Pro and Anti. The results are shown in Tables IV, V and VI. The highest precision for Pro is obtained for PTI (78%) followed by MQM (71%) and PMLN (69%). Anti precisions are lesser than their Pro counterparts, for each party, with the highest for MQM (59%), PMLN (53%) and then PTI (42%). Clearly, Pro is being predicted better than Anti for each party, although the difference is the least for MQM, followed by PMLN and then PTI (highest difference). This is probably due to the lesser frequency of Anti statements available for PTI and PMLN data. Note that the "recall" measure, i.e., the number of retrieved relevant instances, is not much useful in a cross-validation scenario of data training [18].

Table IV Class Precision for PTI

	True Pro	True Anti	Precision
Predicted Pro	238	67	79%
Predicted Anti	7	5	42%

Table V Class Precision for PMLN

	True Pro	True Anti	Precision
Predicted Pro	203	95	69%
Predicted Anti	30	33	53%

Table VI Class Precision for MQM

	True Pro	True Anti	Precision
Predicted Pro	120	51	71%
Predicted Anti	286	400	59%

F. *Predicting the Winning Party*

As outlined in Section IV, we constructed test matrices for PTI, PMLN and MQM and used them to make predictions for 8^{th} May – 11^{th} May, by using the learnt decision trees for each party. The predictions for PTI, PMLN and MQM are shown in Tables VII, VIII and IX respectively. In each table, we first list the number of tweets for each day, followed by the number of Pro predictions (Pro Pred.) and Anti predictions (Anti Pred.) for that day. All predictions are expressed as percentages.

Table VII Electoral Prediction for PTI

	Tweets	Pro Pred.	Anti Pred.
8^{th} May	16	100%	0%
9^{th} May	31	97%	3%
10^{th} May	32	97%	3%
11^{th} May	39	100%	0%

Table VIII Electoral Prediction for PMLN

	Tweets	Pro Pred.	Anti Pred.
8^{th} May	19	26%	74%
9^{th} May	15	0%	100%
10^{th} May	16	6%	94%
11^{th} May	73	23%	77%

Table IX Electoral Prediction for MQM

	Tweets	Pro Pred.	Anti Pred.
8^{th} May	16	12%	88%
9^{th} May	20	15%	85%
10^{th} May	39	7%	93%
11^{th} May	117	16%	84%

As can be seen, PTI has the highest percentage of Pro predictions for each day, which considerably outnumber the Pro predictions for both PMLN and MQM. Consequently, both PMLN and MQM have very high Anti predictions. These results clearly show that twitter sentiments predict PTI as the winning party, and at the same time, predict PMLN and MQM to lose heavily. In order to further verify our results, we manually labeled the tweets in each test matrix (as Pro or Anti) and compared these labels with the predicted ones. The results are shown in Table X. We made manual Anti predictions for MQM and PMLN, and manual Pro prediction for PTI. Also, the predictions are expressed as average over 4 days (8^{th}-11^{th} May). We can see that the Pro prediction is still the highest for PTI, followed by PMLN and the MQM. This result strongly reinforces our statement that Twitter sentiments predict PTI as the winning party. Nevertheless, we do observe a drastic increase in the Pro prediction for PMLN, which signals a type of "bias" in learning. However, this bias is associated with the users' sentiments, and may required extensive experimentation with different parameters and algorithms (which is outside the scope of this work).

Table X Comparison of Predictions with Manual Labels

	Avg. Manual	Avg. Actual
PTI Pro Prediction	76%	99%
PMLN Anti Prediction	44%	87%
MQM Anti Prediction	88%	63%

As the elections are already over, we must comment that the actual winning party was PMLN (with respect to seats in the top-level National Assembly), followed by PPP (which was not at all focused on Twitter), PTI, and finally MQM (refer to http://hamariweb.com/pakistan-election-2013/). However, PTI achieved landslide victory in the Khyber Pakhtunkhaa province, and grabbed seats from several important constituencies across the country. Moreover, PTI supporters issued rigging complaints from the NA-250 constituency, in which re-election lead to PTI victory. From this, we derive the following conclusions:

- The educated category of political analysts was primarily supporting PTI, possibly because of the party's vision to implement reforms in a better way than the other parties.
- Notwithstanding our limited number of Twitter users and their tweets, our predictions are still valid for one whole province.
- From a generic perspective, Twitter users do not represent the significant voting population of Pakistan.
- Twitter may not always significantly impact election results, as PPP won without any considerable Twitter usage.

Recall from Section II that many electoral predictions using Twitter were not exactly precise [17]. However, our message in this paper is that Twitter did create a unifying community of PTI supporters which did have some type of positive impact on the result. In other words, the use of social networking can have a large say in determining the election result, which is what we observed for the US Presidential 2012 Elections.

VI. CONCLUSIONS

In this paper, we have used Twitter to predict the winning political party for the Election held in Pakistan in May 2013. The primary motivation was the positive impact which Twitter analysis had on helping Obama to win US Presidential 2012 elections. We identified a number of relevant users (mainly political analysts), extracted their tweets, and the attributes most useful to gauge the positive and negative user sentiments. Using Rapid Miner, we learnt decision tree predictive models with the CHAID algorithm, as it had better predictive accuracy compared to two other standard algorithms, i.e., Naïve Bayes and Support Vector Machines. We extracted useful rules regarding tweeting trends associated with positive and negative opinions. Our predictions reveal PTI to win the elections. In reality, even though PMLN was the actual winner, PTI did obtain a landslide victory in one province, and bagged other important seats across the country. Our results demonstrate the importance of Twitter in creating some type of positive impact for winning elections, which is reminiscent of Obama's victory in US 2012 elections. However, in Pakistan, Twitter alone cannot be relied heavily for electoral prediction due to its developing Pakistani user-base. Our study does not include the opinions of the rural population of Pakistan which runs in large numbers.

REFERENCES

[1] Twitter Statisitics, http://www.statisticbrain.com/twitter-statistics/, Last Updated: 5.7.2013

[2] C. Spengler, W. Werth, and R. Sigrist, "360º Touchpoint Management - How important is Twitter for our brand", Marketing Review, St. Gallen, 2010

[3] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network", Expert Systems with Applications, Vol. 40, Issue 16, pp 6266-6282, November 2013.

[4] Y. Bae and H. Lee, "Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers", Journal of the American Society for Information Science and Technology, Vol. 63, Issue 12, pp. 2521-2535, December 2012.

[5] M. Hao, C. Rohrdantz, H. Janetzko, U. Dayal, D. A. Kiem, L.E.. Haug, M. C. Hsu, "Visual Sentiment Analysis on Twitter Data Streams", IEEE Symposium on Visual Analytics Science and Technology, October 23-29, Providence, Rhode Island, USA.

[6] Tweet Feel, http://www.tweetfeel.com/

[7] S. Chandrasekar, E. Charon, and A. Ginet, "Predicting the US Presidential Election using Twitter data", CS 229 Machine Learning Project report, Stanford University, Autumn 2012.

[8] J. M. Soler, F. Cuartero, and M. Roblizo, "Twitter as a Tool for Predicting Election Results", In Proceedings of the 2012 IEEE/ACM Inernational Conference on Advances in Social Network Analysis and Monitoring, 2012.

[9] Pakistan General Elections, http://en.wikipedia.org/wiki/Pakistani_general_election,_2013, Last Updated: 19.11.2013

[10] Twitter Political Index, http://www.topsylabs.com/election/, Last Updated: 18.11.2013

[11] A new barometer for the election – Twitter Political Index, https://blog.twitter.com/2012/new-barometer-election, Last Updated: 01.08.2012

[12] Tumasjan, Andranik, Timm Oliver Sprenger, Philipp G. Sandner, and Isabell M. Welpe. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment." *ICWSM* 10 (2010): 178-185.

[13] Belgian Elections – Twitter Opinion Mining, http://www.clips.ua.ac.be/pages/pattern-examples-elections

[14] Gillingham, Gwendolyn, Conway, Michael, Chapman, Wendy, Casale, Michael, AND Pettigrew, Kathryn. "#wheezing: A Content Analysis of Asthma-Related Tweets", *Online Journal of Public Health Informatics* [Online], Volume 5 Number 1 (24 March 2013)

[15] Twimemachine, http://www.twimemachine.com/

[16] Murphy Choy, Michelle L.F. Cheong, Ma Nang Laik, and Koo Ping Shung, "A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction", Journal CoRR, Vol. abs/1108.5520, 2011

[17] D. Gayo-Avello, "A meta-analysis of state-of-the-art electoral prediction from Twitter data", Journal CoRR, Vol. abs/1206.5851, 2012

J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques", 2nd Edition, The Morgan Kaufmann Series in Data Management Systems, July 6, 2011.