

Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies

Ali Mustafa Qamar, Suliman A. Alsuhibany
Computer Science Department
College of Computer, Qassim University
Buraidah, Saudi Arabia

Syed Sohail Ahmed
Computer Engineering Department
College of Computer, Qassim University
Buraidah, Saudi Arabia

Abstract—Twitter has attracted the attention of many researchers owing to the fact that every tweet is, by default, public in nature which is not the case with Facebook. In this paper, we present sentiment analysis of tweets written in English, belonging to different telecommunication companies in Saudi Arabia. We apply different machine learning algorithms such as k nearest neighbor algorithm, Artificial Neural Networks (ANN), Naïve Bayesian etc. We classified the tweets into positive, negative and neutral classes based on Euclidean distance as well as cosine similarity. Moreover, we also learned similarity matrices for kNN classification. CfsSubsetEvaluation as well as Information Gain was used for feature selection. The results of CfsSubsetEvaluation were better than the ones obtained with Information Gain. Moreover, kNN performed better than the other algorithms and gave 75.4%, 76.6% and 75.6% for Precision, Recall and F-measure, respectively. We were able to get an accuracy of 80.1% with a symmetric variant of kNN while using cosine similarity. Furthermore, interesting trends wrt days, months etc. were also discovered.

Keywords—sentiment analysis; social networks; supervised machine learning; text mining

I. INTRODUCTION

The social networking websites such as Twitter, Facebook, LinkedIn, Tumblr, Foursquare and Google+ have become an important part of everyday life. The users express everything related to their experiences, reviews and opinions on these websites. Similarly, since companies and organizations are also interested to know the feedback about their products and offered services, they could take a help from the social media. Although most of the data in social networking websites is private, the data in Twitter is public. This makes Twitter a good choice for research purposes. Using Twitter, users can share their opinion in a *tweet* having at most 140 characters. Currently, Twitter has 313M active users, who post 500M tweets each day¹.

Twitter has several features, such as hash tags (#), mentions (@). User can refer to events and companies in a tweet using hash tags, which could be used to retrieve the list of tweets relevant to a particular entity. On the other hand, Twitter analysis presents a lot of challenges such as a short message length, use of local references and use of non-standard language.

Sentiment analysis, sometimes also referred to as opinion mining, detects the sentiment from the data normally obtained

from the social media and helps in defining policies as well as providing better services.

A study by Qamar et al. [1] has developed a Similarity Learning Algorithm (*SiLA*) for nearest neighbor classification, which learned similarity matrices rather than distance based ones. *SiLA* is capable of learning diagonal, symmetric or square matrices. The similarity between two examples x and y could be calculated as:

$$s_A(x, y) = \frac{x^T A y}{N(x, y)} \quad (1)$$

where T represents the transpose, A is a $p \times p$ similarity matrix, and $N(x, y)$ stands for the normalization function. Replacing A with the identity matrix (I), one can get the cosine similarity. Furthermore, Ahmed et al. [2] used *SiLA* for prediction of popular tweets. They considered those tweets as popular which have been re-tweeted (equivalent of forwarding a message) at least once. However, to the best of our knowledge, *SiLA* has not been applied for sentiment analysis.

In this paper, we present the sentiment analysis of tweets belonging to different Saudi telecommunication companies such as *STC*, *Mobily* and *Zain*. In particular, tweets in English have been selected. We have not missed a single tweet in English belonging to the aforementioned companies. Our idea in this research is to detect the sentiment, which could be either positive, negative or neutral; from the data obtained from Twitter which could in turn help in defining policies as well as providing better services.

We performed feature selection using *CfsSubsetEvaluation* as well as *Information Gain*. The former proved to be a better choice than the latter. We got F-score of 75.6% using kNN. Furthermore, a symmetric variant of kNN performed better than the standard kNN. We were also able to get some insights about Twitter usage, such as popular days, months etc.

Our contributions include sentiment analysis of Saudi telecommunication tweets using various Machine Learning algorithms along with a good F-score, application of similarity and distance metrics, and finding interesting patterns in the Twitter data.

This paper is organized as follows: Section II presents the related work. The methodology used in our research is discussed in Section III, whereas Section IV discusses in detail the experiments. The results are analyzed in Section V. Section VI concludes the paper with future works.

¹<https://about.twitter.com/company>

II. RELATED WORK

Pak and Paroubek [3] gave a thorough description of Twitter as a data source for performing Sentiment analysis along with opinion mining. They distinguished the tweets into positive, negative and neutral classes where the tweets were only in English. The tweets were manually labeled whereby each one was labeled by three different people. Furthermore, they noticed that many of the tweets contain emoticons i.e. icons expressing the emotions of users such as ':)', ':(', '=)', '=(', ';)'. So as to express the users feelings toward an entity or a service. Three classification algorithms, namely, *Naïve Bayes* (NB), *Support Vector Machines* (SVM) and *Conditional Random Fields* (CRF) were used along with features like uni-grams, bi-grams, n-grams etc. in order to classify the tweets.

Recently, Giachanou and Crestani [4] have conducted a thorough survey on *Twitter Sentiment Analysis* (TSA) methods. They identified four different types of (textual) features which have been used so far (semantic, syntactic, stylistic and Twitter-specific). Semantic features include opinion words, sentiment words, negation etc. and could be extracted in a manual or semi-automatic manner from opinion and sentiment lexicons. Many researchers have taken help from lexicons which have been developed for other domains, for example, SentiWordNet [5]. Similarly, syntactic features include *uni-grams*, *bi-grams*, *n-grams*, *terms' frequencies*, *Part Of Speech* (POS). Together with semantic features, they are the most widely used ones. Whereas some researchers preferred binary weighting score based on presence/absence, others considered term frequencies. *Stylistic* features come from the non-standard writing style such as emoticons, use of slang terms and punctuation marks. Lastly, Twitter-specific features include *hash-tags*, *re-tweets*, *replies* and *user names*. Many researchers such as Hong et al. [6] have considered their presence/absence or their frequency.

Natural Language Processing (NLP) techniques have also been used in content analysis. One of the simplest techniques determines the presence of a sentiment lexicon (word expressing positive or negative sentiment) in an entity, such as tweets. Asur and Huberman [7] used the tweets in order to forecast the revenue for movies. They used 3 Million tweets and constructed a linear regression model. Similarly, Zhou et al. [8] developed a *Tweet Sentiment Analysis Model* (TSAM) which was able to successfully determine the societal interest as well as general peoples' opinions with respect to a social event (Australian federal elections). Sriram et al. [9] classified the tweets using a small set of domain-specific features extracted from the authors profile along with the text.

On the other hand, many researchers have used Twitter in order to determine twitter users political influence. For instance, Stieglitz and Xuan [10] performed sentiment analysis of political tweets and analyzed re-tweet behavior, where a tweet is simply forwarded. Razzaq et al. [11] gathered tweets belonging to different political parties just before the Pakistani General Elections 2013 and tried to predict the winner. However, they were not very much successful since a majority of the population did not use Twitter at all. Thus, claims about the general public based on a pattern observed in Twitter, should be made carefully.

Moreover, Burgess and Bruns [12] discusses the challenges

in the filed of Big data with respect to its application on Twitter data.

Go et. al. [13] presented an approach to classify tweets based on positive and negative sentiments. Their approach used different machine learning algorithms for classifying tweets. The results showed that these algorithms offer above 80% accuracy, if trained with emotions data. Uni-grams and bi-grams, in combination, provide better results with Naïve Bayes and MaxEnt classifier algorithms. Qasem et. al. [14], also provide sentiment classification but on stock related tweets. The goal of this work is to compare logistic regression and neural network machine learning strategies in providing positive, negative and neutral tweets by training the classifier using a data set based on 42000 tweets. Uni-gram TF-IDF and Bi-gram TF were used, as feature extractors in the experiment, out of which uni-gram provides better performance. Furthermore, Khan et al. [15] classified tweets into positive, negative and neutral classes using various approaches such as emot-icons, bag of words and SentiWordNet.

Zimbra et al. [16] performed brand-related sentiment analysis using feature engineering. They used only seven features and obtained accuracy above 80% along with very good recall rates. They conducted three-class as well as five-class sentiment classification.

Recently, Latifah and Cristea [17] worked on Arabic tweets to predict the satisfaction of Saudi telecommunication companies' customers. However, they only presented a plan and their research is expected to complete by the year 2022.

III. RESEARCH METHODOLOGY

This paper focuses on analyzing tweets written in English language. Therefore, we gathered all tweets written in English language belonging to different telecommunication companies of KSA, namely, Zain, STC and Mobily. A total of 1331 tweets were found. A majority of them, 75.2% i.e. 1001 out of 1331, belong to *STC*, the largest telecommunication company in Saudi Arabia. Similarly, 207 tweets (15.5%) belong to *Mobily*, whereas 124 (9.3%) are for *Zain* as shown in Fig. 1. The official handles for the aforementioned companies are *@STC_KSA*, *@Mobily* and *@ZainKSA* whereas the number of their followers on Twitter are 3.16 M, 3.07 M, and 1.32 M, respectively. Nevertheless, the number of tweets for *Mobily* are more than 3 times than that for *STC*.

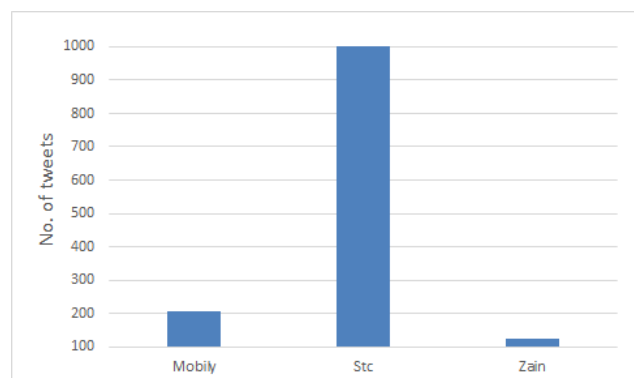


Fig. 1. Proportion of Tweets belonging to different companies

In order to determine the class of each tweet, we asked three different persons to classify the tweets into either *positive*, *negative* or *neutral* class. The class was determined based on a majority vote. Most of the tweets 57.25% (762/1331) expressed *negative* sentiment, where as positive and neutral sentiment was found in 203 (15.25%) and 366 (27.50%) tweets, respectively. Fig. 2 depicts the polarity of tweets belonging to different companies. Moreover, Table I provides a sample of tweets belonging to the three classes.

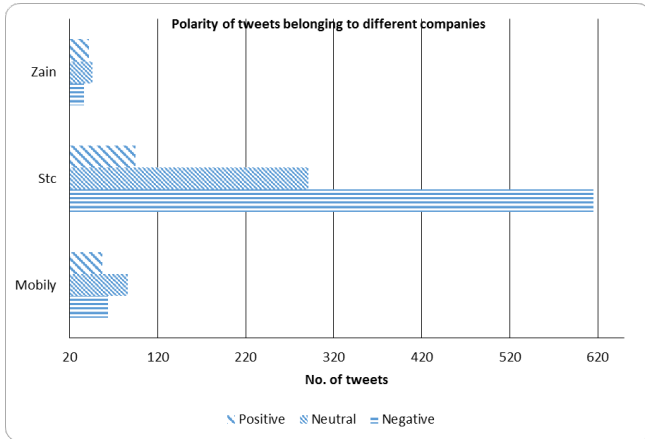


Fig. 2. Polarity of Tweets

TABLE I. SAMPLE TWEETS FOR THE THREE CLASSES

Tweet	Class
"If there is anyway I can bypass even if I have to pay more I will... DSL down everywhere... Total"	<i>negative</i>
"New: Saudi Telecom Company (STC) to Release iPhone February 11 URL"	<i>neutral</i>
"I am using the 3G network intensely today & It is perfect . USER not just the DSL buddy everything is ok today ? WEIRD!!"	<i>positive</i>

A. Supervised Machine Learning Algorithms

A number of supervised machine learning algorithms was used for classification. Their brief details are presented as follows:

k Nearest Neighbor (kNN) algorithm is a very simple yet efficient classification algorithm. It belongs to *lazy* classifiers since it does no real work until classification time. In order to classify a new example x , its k nearest neighbors are found from a set of already classified examples. Afterwards, x is assigned to the class which is the most represented in the set of nearest neighbors. The proximity is determined based on either distance or similarity. Whereas the distance is always greater than or equal to 0, similarity has the range [-1, 1]. *SiLA* uses two prediction rules: *kNN-A* based on learning the similarity matrix A using standard *kNN*, and symmetric *kNN* (*SkNN-A*), where k nearest neighbors are found from different classes. The similarity is calculated with each class (sum of similarity between x and its k nearest neighbors in the class). This is followed by assigning x to the class having maximum similarity.

Naïve Bayes (NB) classifier is based on the application of Bayes' theorem with strong (naïve) independence assumptions among the features. By default, it uses a normal distribution. However, one can also use a *kernel estimator* for numeric attributes. In multinomial Naïve Bayes, the feature vectors are the frequencies with which certain events have been generated by a multinomial.

B. Pre-Processing

The data pre-processing is required in order to remove duplicate tweets, hash tags along with repeated symbols. In particular, following pre-processing tasks were performed:

- User-ids, preceded by '@' sign were converted into *USER*.
- URLs were also replaced with the keyword *URL*.
- An unsupervised filter was applied at the attribute level so as to convert the tweet text in a word vector (*StringToWordVector*). All words were converted into lower case.
- Stemming helps convert a word to its word stem or root form, e.g. *fishing*, *fished*, and *fisher* are reduced to the word *fish*. Stemming was performed using *LovinsStemmer*.
- A stoplist was used in order to remove common words such as *a*, *an*, *the*, *as* etc. which have no influence in finding the sentiment of a tweet.
- *TF-IDF* (term frequency - inverse document frequency) measure was also applied on the data. This reflects how important a word is to a tweet in the collection of tweets. It is represented as:

$$f_{ij} \log \frac{\text{number of tweets}}{\text{number of tweets that include word } i} \quad (2)$$

where f_{ij} is the frequency of word i in tweet j . The idea is to reduce the weightage of the words appearing in more tweets, since they are useless as discriminators [18].

- The initial number of attributes was more than 2500. In order to reduce this, *CfsSubsetEval*, which is an attribute selection method, was applied to reduce the number of features to 40. It evaluates the worth of a subset of attributes by taking into account the individual predictive ability of each feature along with the degree of redundancy between them. This method prefers subsets of features that are highly correlated with the class as compared to the ones having low intercorrelation. Some of the selected features include *dsl*, *googl*, *crap*, *telecommunic*, *fail*, *worst*, *stupid*, *damn*, *proper* and *reach*. Once the features were reduced, a number of rows (809) just contained all zeros. Such rows were removed, giving rise to a **smaller** dataset.
- Furthermore, 68 attributes were also selected based on *Information Gain* (*InfoGainAttributeEval* in *WEKA*) as the evaluator, and *Ranker* having a threshold of 0 as a search method.

IV. EXPERIMENTS

This section describes the used software and various metrics.

Waikato Environment for Knowledge Analysis (WEKA), an open-source software containing implementation for a number of machine learning algorithms was used for most of the algorithms. Sentiment analysis is primarily a supervised learning process, thus belongs to supervised machine learning. 5-fold cross-validation was used for the experiments, in which case the data is divided into 5 equal parts. One part is selected for testing, where as rest of the 4 parts are used for training. Afterwards, another part is selected for training. Thus, each example is selected exactly once for testing and 4 times for training. The results obtained with various algorithms are compared based on precision, recall and F-measure.

Table II shows the confusion matrix for sentiment detection. *True Positives* (TP) indicates the instances that were predicted as positive and were indeed positive. Similarly, *False Positives* (FP) refers to the tweets which were wrongly classified as positive. *True Negative* (TN) and *False Negative* (FN) are defined in the similar manner for the negative class.

TABLE II. CONFUSION MATRIX FOR ANALYZING THE PERFORMANCE OF SENTIMENT CLASSIFICATION METHODS

	classified as <i>Positive</i>	classified as <i>Negative</i>
Are <i>Positive</i>	TP	FN
Are <i>Negative</i>	FP	TN

Accuracy is one of the most frequently used metric [4] and calculates the ratio of the true predictions to the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

Precision shows the exactness of a method and is defined as the percentage of tweets predicted to be of class *X* which actually belong to class *X*.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

On the other hand, Recall, also known as sensitivity, is the percentage of tweets which actually have class *X* and which have been correctly predicted to have class *X* by the algorithm. It is defined as the fraction of positive instances which were predicted as positive. The recall is given as:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

F-measure is the harmonic mean of precision and recall. This is calculated as:

$$F\text{-measure} = \frac{2 * Precision * Recall}{Precision + Recall} \quad (6)$$

TABLE III. RESULTS FOR THE ORIGINAL DATA SET WITH CFSUBSETEVALUATION (IN PERCENTAGE)

Algorithm	Accuracy	Precision	Recall	F-measure
NB	57.2	54.1	57.2	51.9
NB (with Kernel Estimator)	62.1	60.4	62.1	60.2
NB (Multinomial)	63.8	63.0	63.8	60.6
Simple Logistic	62.9	62.7	62.9	60.6
kNN (k=1)	61.0	62.2	61.0	58.5
kNN (k=3, distance weighting)	60.9	61.6	60.9	57.7

V. CLASSIFICATION RESULTS

Classification algorithms were applied on both original as well as smaller dataset. Table III shows the results on the original data set. The best results comprise 63.8% accuracy, 63% Precision, 63.8% Recall and 60.6% F-measure. These results were obtained by *Naïve Bayes*, while using its multinomial variant.

The smaller data set contained only 522 instances. In particular, 327 belong to the *negative* class, whereas 113 had *neutral* sentiment and only 82 displayed *positive* sentiment. Table IV shows the results obtained by various algorithms on the smaller data set while using *CfsSubsetEval* and without using *N-grams* model. *kNN* appears to be the best having better values for accuracy (76.6%), precision (75.4%), recall (76.6%) as well as F-measure (75.6%). It can be easily observed that the results on smaller data set are way better than the ones obtained with original i.e. larger data set. One of the primary reasons is that the smaller data set is void of tweets which do not contain any of the selected features (words).

Table V contains the results on the smaller data while using *N-grams* and *Information Gain*. The maximum size for *N-grams* was selected as three; giving rise to uni-grams, bi-grams and tri-gram. *Information Gain* was selected for feature selection. Although NB with *Kernel Estimator* got the best Precision of 68.3%, yet because of poor Recall (48.6%), F-measure was just 47.7%. On the other hand, *kNN* got F-measure of 63.7%.

Table VI shows the accuracy along with standard deviation obtained with *kNN* and *SiLA*. For *SiLA*, 80% of the examples were used for training while 20% were used for testing purposes. The results with symmetric variant of *kNN* i.e. *SkNN* are the best. Another interesting thing is that the *Euclidean distance* appears to be working well with textual data. The accuracy for all approaches except *SkNN-A* (*SkNN* with *SiLA*) is better than the ones reported for larger and smaller data sets in Tables III and IV. The results were also evaluated for statistical significance i.e. whether one method is significantly better than the other one, using *s-test* [19]. In case

TABLE IV. RESULTS FOR THE SMALLER DATA SET WITHOUT N-GRAMS AND CFSUBSETEVALUATION (IN PERCENTAGE)

Algorithm	Accuracy	Precision	Recall	F-measure
NB	69.2	68	69.2	68.4
NB (with Kernel Estimator)	72.4	72.3	72.4	72.1
NB (Multinomial)	74.7	75.6	74.7	75.1
Simple Logistic	75.3	73.5	75.3	73.8
kNN (k=1)	76.6	75.4	76.6	75.6
kNN (k=3, distance weighting)	74.9	75.4	76.6	75.6

TABLE V. RESULTS FOR THE SMALLER DATA SET WITH N-GRAMS AND INFO GAIN (IN PERCENTAGE)

Algorithm	Precision	Recall	F-measure
NB	64.9	52.1	53.6
NB (with Kernel Estimator)	68.3	48.6	47.7
NB (Multinomial)	61.1	61.8	58.2
Simple Logistic	65.6	63.7	57.0
MLP	60.5	60.6	59.6
kNN (k=1)	63.4	64.0	63.6
kNN (k=3 and distance weighting)	63.4	64.5	63.7

the P-value is less than or equal to 0.01, this means that the difference is much more significant and is represented as \gg . Consequently, a lower level of significance occurs when the P-value lies in between 0.01 and 0.05 ($>$). In case, the P-value is greater than 0.05, the results are considered equivalent ($=$). We observed that *kNN-Euclidean*, *kNN-cosine*, *SkNN-cosine* and *kNN-A* performed *significantly* better than *SkNN-A*. Similarly, all of the methods except *SkNN-A* were *statistically* equivalent. This can be expressed as follows:

$$(kNN-Euclidean = kNN-cosine = SkNN-cosine = kNN-A) \gg SkNN-A$$

Fig. 3 shows the impact of different values of k on the accuracy of *kNN* for various algorithms: *SiLA* with cosine (*kNN-A*), *SiLA* with symmetric *kNN* (*SkNN-A*), *kNN* with Euclidean distance, *kNN* with cosine, *SkNN* with cosine. As k increases from 1 to 7, accuracy with *SkNN-cos* increases from 0.62 to 0.81. On the other hand, it increases from a value of 0.60 to 0.76 while employing *SkNN-A*. Furthermore, the

TABLE VI. RESULTS FOR THE SMALLER DATA SET WITH SIMILARITY LEARNING (IN PERCENTAGE)

Algorithm	Accuracy \pm sd
<i>kNN - Euclidean</i>	79.68 \pm 3.35
<i>kNN - Cosine</i>	79.42 \pm 4.01
<i>SiLA - kNN-A</i>	79.10 \pm 4.29
<i>SkNN - Cosine</i>	80.13 \pm 4.17
<i>SiLA - SkNN-A</i>	75.58 \pm 7.27

accuracy with Euclidean distance decreases from 0.79 to 0.65 and eventually to 0.63 before increasing to 0.79.

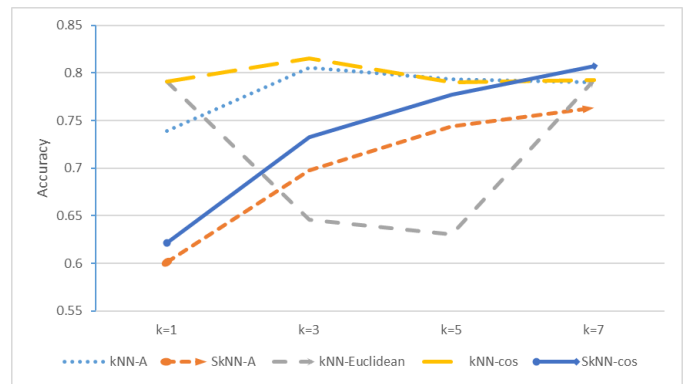


Fig. 3. Accuracy for different values of k in *kNN* along with various algorithms

The data set was further analyzed and it turned out most of the tweets are from 2010 and 2011 (more than 70%) as shown in Fig. 4. Moreover, it was found out that most of the tweets were written in the month of January as shown in Fig. 5. Months like March, July and November saw less number of tweets as compared to the other months. While looking closely at the different years, we noticed that October contributed most of the tweets for 2010.

Extending the analysis to the days of a week, it was noticed that most tweets were written on Wednesday (the day before the weekend) as shown in Fig. 6.

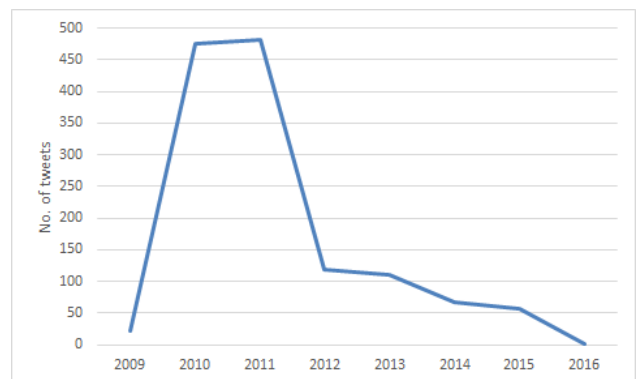


Fig. 4. Number of Tweets from 2009 to 2016

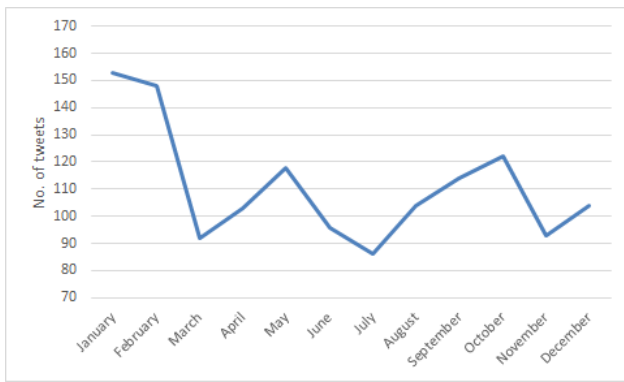


Fig. 5. Variation in the number of tweets for different months

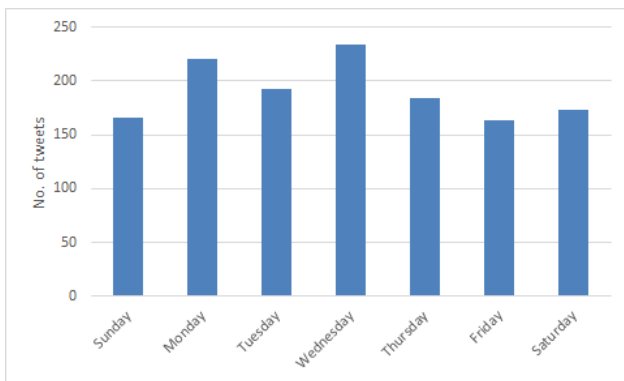


Fig. 6. Tweets for different days of a week

It may be noted that the weekend in Saudi Arabia comprised of Thursday and Friday till June 28, 2013. We also observed the number of tweets for different days over the years as shown in Fig. 7. Wednesday saw the most number of tweets for the years 2010, 2011 as well as 2013. In 2012, most of the tweets appeared on Tuesday.

Fig. 8 shows the number of tweets by different users. Interestingly, only one user wrote more than 80 tweets, while four users tweeted more than 20 times.

A number of issues were faced while conducting the experiments:

- Some of the tweets contained words from other languages. In such cases, tweets' sentiment was determined as if the word was not present.
- Conflicting sentiments: There were some tweets which contained conflicting sentiments e.g. the tweet *#Vodafone UK u r a breath of fresh air. #ZainKSA shame on u* contains both positive as well as negative sentiment. However, since the sentiment towards Saudi telecommunication company is negative, therefore, the tweet was considered as negative. One can also note that *you* has been written as *u* and *are* as *r*.

VI. CONCLUSION

In this paper, we presented sentiment analysis of tweets written in English, belonging to the various telecommunication companies (*Mobily*, *STC* and *Zain*) of Saudi Arabia.

Three classes, namely, *positive*, *negative* and *neutral* were considered. We made sure that none of the related tweets were missed. A number of machine learning algorithms like *ANN*, *k nearest neighbor (kNN)*, *Naive Bayesian* were used for classification. *kNN* got the best results including *F-measure* of 75.6%. Furthermore, different metrics such as *Euclidean distance* and *cosine similarity* were used with *kNN*. The results with *cosine* were slightly better than the ones obtained with its counterpart. We also applied *Similarity Learning* algorithm (*SiLA*). However, the results were not improved. Our results also showed that increasing the value of *k* has a positive impact on the accuracy for some of the algorithms. Lastly, we found out that the maximum tweets were written in the months of *January* and *February* during the years 2010 and 2011. We also observed that most of the tweets were written on the day before the weekend (*Wednesday*).

In the future, sentiment could be deduced from tweets written in arabic language. This would help increase the size of the data set as well, since most of the tweets related to the telecommunication companies of Saudi Arabia are in arabic. Moreover, various methods such as *Support Vector Machines (SVM)*, *ensemble* techniques could also be employed. One could also define a sentiment on a 5–10 scale, for example, –1 (not negative) to –5 (extremely negative) and 1 (not positive) to 5 (extremely positive).

ACKNOWLEDGEMENT

The authors gratefully acknowledge Qassim University, represented by the Deanship of Scientific Research, on the material support for this research under the number 3289 during the academic year 1436 AH / 2015 AD.

REFERENCES

- [1] A. M. Qamar, E. Gaussier, J.-P. Chevallet, and J. H. Lim, "Similarity learning for nearest neighbor classification," in *Proceedings of International Conference on Data Mining (ICDM)*, Pisa, Italy, December 15-19, 2008, pp. 983–988.
- [2] H. Ahmed, M. A. Razzaq, and A. M. Qamar, "Prediction of popular tweets using similarity learning," in *Proceedings of IEEE International Conference on Emerging Technologies (ICET)*, Islamabad, Pakistan, December 9-10, 2013.
- [3] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA), May 2010.
- [4] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," *ACM Computing Surveys (CSUR)*, vol. 49, no. 2, pp. 28:1–28:41, Jun. 2016.
- [5] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," in *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06)*, 2006, pp. 417–422.
- [6] L. Hong, O. Dan, and B. D. Davison, "Predicting popular messages in twitter," in *Proceedings of the 20th International Conference Companion on World Wide Web*, ser. WWW '11. New York, NY, USA: ACM, 2011, pp. 57–58.
- [7] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, ser. WI-IAT '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 492–499.
- [8] X. Zhou, X. Tao, J. Yong, and Z. Yang, "Sentiment analysis on tweets for social events," in *Computer Supported Cooperative Work in Design (CSCWD)*, 2013 *IEEE 17th International Conference on*, June 2013, pp. 557–562.

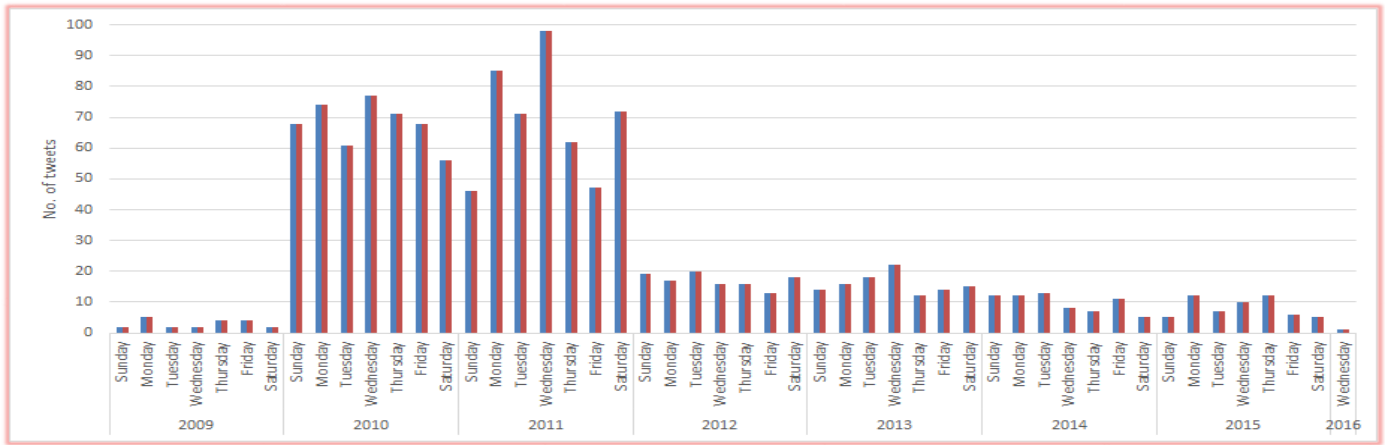


Fig. 7. Variation in the number of tweets for different days of a week over the years

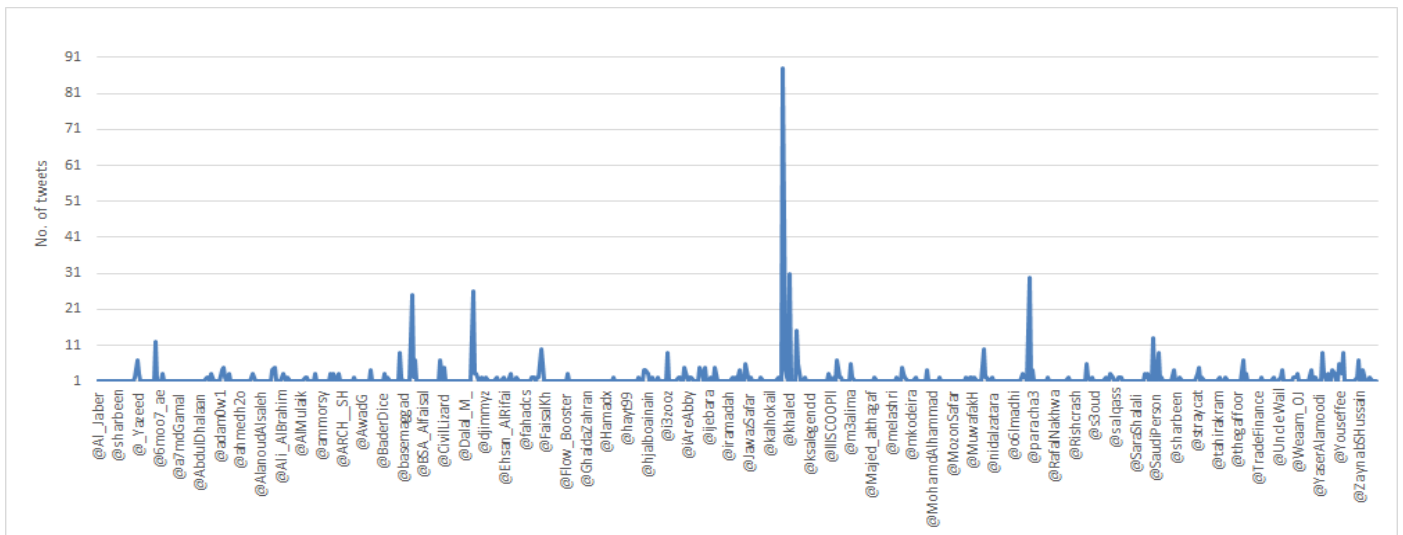


Fig. 8. Number of tweets by different users

[9] B. Sriram, D. Fuhrý, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '10. New York, NY, USA: ACM, 2010, pp. 841–842.

[10] S. Stieglitz and L. Dang-Xuan, "Political communication and influence through microblogging—an empirical analysis of sentiment in twitter messages and retweet behavior," in *Proceedings of the 2012 45th Hawaii International Conference on System Sciences*, ser. HICSS '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 3500–3509.

[11] M. A. Razzaq, A. M. Qamar, and H. S. M. Bilal, "Prediction and analysis of pakistan election 2013 based on sentiment analysis," in *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM, Beijing, China, August 17-20, 2014*, pp. 700–703.

[12] J. Burgess and A. Bruns, "Twitter archives and the challenges of "big social data" for media and communication research," *M/C Journal*, vol. 15, no. 5, pp. 1–7, October 2012.

[13] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *Processing*, pp. 1–6, 2009.

[14] M. Qasem, R. K. Thulasiram, and P. Thulasiraman, "Twitter sentiment classification using machine learning techniques for stock markets." in *ICACCI*, J. L. Mauri, S. M. Thampi, M. Wozniak, O. Marques, D. Krishnaswamy, S. Sahni, C. Callegari, H. Takagi, Z. S. Bojkovic, V. M., N. R. Prasad, J. M. A. Calero, J. Rodrigues, X. Que, N. Meghanathan, R. Sandhu, and E. Au, Eds. IEEE, 2015, pp. 834–840.

[15] F. H. Khan, U. Qamar, and M. Y. Javed, "Sentiview: A visual sentiment analysis framework," in *International Conference on Information Society (i-Society 2014)*, Nov 2014, pp. 291–296.

[16] D. Zimbra, M. Ghiassi, and S. Lee, "Brand-related Twitter sentiment analysis using feature engineering and the dynamic architecture for artificial neural networks," in *49th Hawaii International Conference on System Sciences, HICSS 2016, Koloa, HI, USA, January 5-8, 2016*, 2016, pp. 1930–1938.

[17] L. Almqren and A. I. Cristea, "Twitter analysis to predict the satisfaction of telecom company customers," in *Late-breaking Results, Demos, Doctoral Consortium, Workshops Proceedings and Creative Track of the 27th ACM Conference on Hypertext and Social Media (HT 2016)*, Halifax, Canada, July 13-16, 2016.

[18] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.

[19] Y. Yang and X. Liu, "A re-examination of text categorization methods," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1999, pp. 42–49.