Sentiment analysis of Turkish tweets by data mining methods

Adem Karahoca
Bahçeşehir University

International Journal of Mechanical Engineering and Technology (IJMET)
Volume 10, Issue 01, January 2019, pp. 915–925, Article ID: IJMET_10_01_094
Available online at http://www.iaeme.com/ijmet/issues.asp?JType=IJMET&VType=10&IType=01
ISSN Print: 0976-6340 and ISSN Online: 0976-6359

SENTIMENT ANALYSIS OF TURKISH TWEETS BY DATA MINING METHODS

A Karahoca
Professor, Bahcesehir University, Software Engineering Department, Turkey

D Karahoca
Assistant Professor, Bahcesehir University, Health Management Department, Turkey

Erkut Evirgen
Bahcesehir University Bahcesehir University

ABSTRACT

Twitter, has fast emerged as one of the most powerful social media sites which can sway opinions. Sentiment or opinion analysis has of late emerged one of the most researched and talked about subject in Natural Language Processing (NLP), thanks mainly to sites like Twitter. In the past, sentiment analysis models using Twitter data have been built to predict sales performance, rank products and merchants, public opinion polls, predict election results, political standpoints, predict box-office revenues for movies and even predict the stock market. This study proposes a general frame in R programming language to act as a gateway for the analysis of the tweets that portray emotions in a short and concentrated format. The target tweets include brief emotion descriptions and words that are not used with a proper format or grammatical structure. Majority of the work constituted in Turkish includes the data scope and the aim of preparing a data-set. There is no concrete and usable work done on Turkish Tweet sentiment analysis as a software client/web application. This study is a starting point on building up the next steps. The aim is to compare five different common machine learning methods (support vector machines, random forests, boosting, maximum entropy, and artificial neural networks) to classify Twitters sentiments.

Keyword head: Turkish Tweets, Emotion Analysis, Text Mining, Sentiment Analysis, Classification, Support Vector Machines (SVM), Random Forest, Boosting, Artificial Neural Networks.

Cite this Article: A Karahoca, D Karahoca and Erkut Evirgen , Sentiment Analysis of Turkish Tweets by Data Mining Methods, International Journal of Mechanical Engineering and Technology, 10(01), 2019, pp.915–925
http://www.iaeme.com/IJMET/issues.asp?JType=IJMET&VType=10&Type=01

1. INTRODUCTION

Others' opinions have always mattered to mankind. Whether it was to wage wars, or make a simple choice as picking a cola from the local grocery store, we have always looked at what others think about the choice we are about to make. Perhaps it emanates from an inherent conforming-with-the-majority attitude, but the bottom line is, that opinions do matter. More so in today's digital world, where thanks to the reach and penetration of the internet, opinions at a global scale are available. That is all that it takes today to make a difference. The micro blogging site called Twitter, has fast emerged as one of the most powerful social media sites which can sway opinions with 140 characters.

Sentiment or opinion analysis has of late emerged one of the most researched and talked about subject in Natural Language Processing (NLP), thanks mainly to sites like Twitter. In the past, sentiment analysis models using Twitter data have been built to predict sales performance, rank products and merchants, public opinion polls, predict election results, political standpoints, predict box-office revenues for movies and even predict the stock market. There are a plethora of start-up companies which have emerged who are very vociferously engaging in sentiment/opinion analysis for maximizing their revenues. However, gathering opinion or analyzing sentiment in not as straight forward as it seems. Like mentioned, above, it is one of the most challenging problems in NLP.

With the recent surge in the availability of data, companies the world over, are leveraging the power of gaining insights from data to solve real world problems or for achieving business goals. The volume, velocity and variety of data being generated have reached unprecedented rates. Not only has this called for newer platforms like Hadoop, to handle big data, but also new machine learning techniques and algorithms to derive insights from the data. The focus of this project is on one such technique to handle both structured and unstructured data. Sentiment Analysis defined as "the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials." The basic idea behind Sentiment Analysis is to extract an opinion. And opinions do matter. In today's highly connected world, social networking sites rule the roost. The number of "likes", "dislikes", "retweets", ratings etc. sway the core thinking of human beings. Today, product managers are more concerned about opinions on social networking sites of their products, rather than feedback provided on their site. Movie reviews and restaurant reviews on such social networking sites dictate the amount of profit the movie producer or restaurant owner is likely to rake in. Sentiment Analysis essentially looks at classifying the polarity of text, emoticons and now days, even images and videos. The aim is to find out whether the source material is positive, negative or neutral. A lot of the sentiments expressed through social networking sites may not get captured by more traditional survey questions. Sentiment Analysis bridges this gap.

2. BACKGROUND

A comparative analysis on techniques used for sentiment analysis shows techniques utilizing both lexicon and non-lexicon based approaches for polarity identification. The paper also reflects on a multilingual approach and concludes by stating that no existing technique is language independent, thereby prompting a case for a generalized Sentiment Analyzer [1]. Xia and Li (2015) claim about dual sentiment analysis. The study proposes a model to handle the polarity shift problem and brings out the inadequacies of the bag of words (BOW) approach. The proposed model creates reversed reviews that are sentiment-opposite to the original reviews, and make use of the original and reversed reviews in pairs to train a sentiment classifier and make predictions [2]. Similar to the organization of Gmail's inbox, Shetty et. al (2014), propose classification of Facebook news feeds based on sentiment analysis. The proposed model automatically identifies "important" feeds which reduce manual survey work which is done for drawing conclusions on opinions posted on Facebook [3]. Islam (2014) describes a procedure of obtaining a unified user rating system for Google Play Store apps by sentiment analysis on written reviews as well as the starred ratings [4]. The use of sentiment analysis of social media content for forecasting election results (Pakistan general elections 2013) has been shown in Razzaq 2014. The results obtained have shown remarkable accuracy to the actual outcome of the elections [5].

Another unique use of Twitter data has been to gauge users' response to popular smart phone brands and their underlying operating systems. The results show that although the Twitter data does provide some information about users' sentiments to the popular smart phone brands and their underlying operating systems, the amount of data available for different brands varies significantly. This limitation makes the comprehensive analysis of users' response somewhat more challenging for some brands compared to others and consequently makes the comparison between brands almost impossible [6]. Multi-dimensional sentiment analysis on restaurant ratings has been undertaken to add special contexts and pricing as two other major aspects in the AAA Restaurant Diamond Rating Guidelines to rate a restaurant. The data for this study was scraped from http://Yelp.com. Results from fitting a multilevel model showed that the sentiments about each of these five aspects alone explained about 28% of the explainable between-restaurant variances, and 12% of the explainable within-restaurant variances of the restaurants' star ratings [7].

The ability to identify opinions in the presence of diverse modalities is becoming increasingly important. This study experiments with several linguistic, audio, and visual features, and shows that the joint use of these three modalities significantly improves the classification accuracy as compared to using one modality at a time [8].

Similarly Wollmer et al., focuses on automatically analyzing a speaker's sentiment in online YouTube videos containing movie reviews. In addition to textual information, this approach considers adding audio features typically used in speech-based emotion recognition, as well as video features encoding valuable valence information conveyed by the speaker [9].

The attention on Twitter for sentiment analysis is immense. This study reports on the design of a sentiment analysis, extracting a vast amount of tweets. Prototyping is used in this development. Results classify customers' perspective via tweets into positive and negative, which is represented in a pie chart and html page [10]. Bhuta et al. (2014) reviews a number of techniques, both lexicon-based approaches as well as learning based methods that can be used for sentiment analysis of Twitter text [11].

Use of an SVM classifier combined with a cluster ensemble offers better classification accuracies than a stand-alone SVM. The study proposes an algorithm which can refine tweet classifications from additional information provided by clusters, assuming that similar instances from the same clusters are more likely to share the same class label [12].

Singh et al. (2013) present an experimental study on a new kind of domain specific feature-based heuristic for aspect-level sentiment analysis of movie reviews. The methodology adopted analyses the textual reviews of a movie and assigns it a sentiment label on each aspect. The scores on each aspect from multiple reviews are then aggregated and a net sentiment profile of the movie is generated on all parameters [13].

Detection of anomalies in tweets in a timely manner can be very beneficial. In this study, the authors survey existing anomaly analysis as well as sentiment analysis methods and analyses their limitations and challenges. To tackle the challenges, an enhanced sentiment classification method is proposed and discussed [14].

Beltagy and Ali (2013) highlight the major problems and open research issues that face sentiment analysis of Arabic social media. The paper also presents a case study the goal of which is to investigate the possibility of determining the semantic orientation of Arabic Egyptian tweets and comments given limited Arabic resources. One of the outcomes of the presented study is an Egyptian dialect sentiment lexicon [15].

Duwairi (2015) investigates sentiment analysis in Arabic tweets with the presence of dialectical words. The study uses machine learning techniques to determine the polarity of tweets written in Arabic with the presence of dialects. Dialectical Arabic is abundantly present in social media and micro blogging channels. Dialectical Arabic presents challenges for topical classifications and for sentiment analysis [16].

For the Turkish language, Demirci (2014) has carried out an emotion analysis on Turkish Tweets. Rather than doing a sentiment analysis, the author talks about emotion analysis. He classifies emotions as joy, sadness, anger, fear, disgust and surprise. He also brings out that the Turkish language is an agglutinative language. This creates problems as far as analysis is concerned because each derivational suffix has the possibility of changing the meaning of the word, and hence to obtain the real meaning of a word each derivational suffix must be examined. Using Twitter is beneficial because of the sheer volume of the data generated. The author has chosen Zemberek (https://code.google.com/p/zemberek/) as a multiple morphological analyser for the study. Zemberek is an open source, platform independent, general purpose Natural Language Processing library and toolset designed for Turkic languages, especially Turkish. In the study, the author has utilized several types of classifiers. Bayesian classifier is selected since it gives the probability of the instance to be in a class. SVM is selected to utilize both linear and non-linear models. k-NN, which uses similarities of the instances, is chosen as an inherently multi-class classifier. Tools used in the study are Zemberek, LIBSVM and WEKA.

For the purpose of data collection, the author has compiled a list of words and phrases for each emotion class. Keyword based search is applied to construct a data set with ready to use class labels. In the pre-processing phase, all non-emotional content is removed. This includes removal of hashtags, URLs, links and user names. Thereafter, these are subjected to morphological analysis to correct mistyped words. For feature vector construction, different combinations of ngrams, part of speech tagging, emoticons and punctuation marks is tried out with term count and tf-idf statistics. Feature selection is implemented through information gain with term count and term frequency-inverse document frequency as the weighing factors. As mentioned earlier, WEKA, LIBSVM and Matlab are the tools used for the classifiers - Naive Bayes, Compliment Naive Bayes and k-nearest neighbours. The study also uses pronouns and textual numbers as stop words [17].

Though the work done in this study is indeed creditable, it lacks the capability of handling a wide range of very diverse tweets primarily because of the absence of a comprehensive keyword list. A data set containing more diverse and a greater number of tweets may better generalize the tweets in Turkish. Also the work does not cater for neutral tweets. Elimination of such non-emotional tweets will help in better classification.

3. DATA AND METHODS

This research is examining the ability of different machine learning classifiers in classifying sentiment in twitter feeds. Five common machine learning methods for supervised classification of data include support vector machines, random forests, boosting, maximum entropy, and artificial neural networks. These methods have demonstrated ability to classify data including text data with desirable performance outputs across a wide range of problem types [18].

3.1. Data Gathering and Pre-processing

Twitter feeds from hepsiburada.com and pegasus.com were obtained. The 717 tweets from hepsiburada.com and 129 tweets from pegasus.com (846 total tweets) consisted of opinions about travel. The language of the twitter feeds was Turkish. In addition, a lexicon of 3579 words was generated with sentiment ranging from very negative (-5) to very positive (+5), including a variety of sentiments in-between. The lexicon ratings were performed by the author. Of these words in the lexicon, 2,821 had a score of zero, not having any positive nor negative sentiment.

The twitter feeds were then pre-processed by using `R` Statistical Software (version 3.2.3) with packages for text-mining (`tm`, version 0.6-2) and text classification (`RTextTools`, version 1.4.2). A document-term matrix was created, excluding words with less than two letters, removing numbers, removing punctuation, and converting punctuation to lower case. The matrix consisted of 3,732 words (listed as columns) across the 846 tweets (listed as rows) with number of occurrences in the tweet as values. Most words were represented in a tweet just once with 138 tweets included a word twice and 3 tweets included a word three times. An example of a document-term matrix based on three tweets is described in Table 1. The tweets are “mobil alisveris yukseliste...” (Document 1), “umarim yarin teslimatim yapilir..” (Document 2), and “resmen kandirildim cok yazik” (document 3). Each tweet is considered a “document” and are represented by rows. Each word is considered a “term” and represented in the column with each number representing a word ordered alphabetically (e.g., “alisveris” = 1, “cok” = 2, “kandirildim” = 3, “mobil” = 4, etc.). The cells represent the number of times a term (i.e., a word) is mentioned in the document (i.e., a tweet). The document term matrix is a sparse matrix, with many cells containing no information. If not empty, most of the cells in the table are represented by the value of 1, which is the number of times that a term is represented in a document as similar in.

Table 1 Sample document term matrix

Document	1	2	3	4	5	6	7	8	9	10	11
1	1			1							1
2						1	1	1	1		
3		1	1		1					1	

3.2. Scoring of Sentiment of Tweets

For each tweet, a total sentiment score was calculated based upon the matrix and the lexicon. The total sentiment score for a tweet was equal to the sum of the sentiment score of each word times the number of the times the word was represented in eq. 1.

$$S_{tweet} = \sum(s_n w_n) \qquad (1)$$

Where Stweet denotes the total sentiment score for a tweet, sn is the sentiment score of a word n and wn is the number of times word n appears in a tweet. Equation 1 could be applied to Table 1. As mentioned previously, most words had a sentiment score of 0. For the tweets used to create Table 1, “umarim(hope so)” (Term 7) had a score of +1, “cok(much)” (Term 2) had a score of +1, and “kandirildim(deceived)” (Term 3) had a score of -4. The total sentiment score for document 1 would be 0 (0*1 + 0*1 + 0*1), for document 2 would be 1 (1*1 + 0*1 + 0*1 + 0*1), and for document 3 would be -3 (1*1 + -4*1 + 0*1 + 0*1). From the Table 1 example, documents 1, 2, and 3 would have raw scores of 0, 1, and -3, respectively. Under the scoring system from Equation 1, the sentiment scores for a tweet ranged from -11 to +11 (Table 2). The raw scores were further processed to develop two alternative scores. A tweet score was converted to a simple positive/negative scoring system (-1, 0, +1) based on the total positive, neutral, and negative scores. Tweets with a total score of -10 or -3 were both considered negative tweets and the magnitude of the negative or positive score was ignored. Also, a tweet score was converted to a scaled score (-2, -1, 0, +1, +2) whereby scores of -11 to -5 were reassigned a score of -2, scores of -4 to -1 were reassigned a score of -1, scores of 1 to 4 were reassigned as score of 1, and scores of 5 to 11 were reassigned a score of 2. This scoring differentiated the very negative or very positive scores. The simple positive/negative scoring system for documents 1, 2, and 3 would be 0, 1, and -1, respectively. The scaled scores for documents 1, 2, and 3 would also be 0, 1, and -1, respectively. These two processed scores (positive/negative and scaled) were used as the response variable for the supervised classification step. The 846 tweets were reshuffled to create a random ordering of tweets for further analysis. The tweets were separated into a class for training the models (700) and for testing (146).

Table 2 Distribution of sentiment scores for tweets

Score	Number of Tweets	Score	Number of Tweets
-11	1	1	107
-9	2	2	64
-8	5	3	31
-7	11	4	21
-6	9	5	17
-5	39	6	7
-4	34	7	1
-3	55	8	2
-2	72	9	1
-1	136	10	1
0	229	11	1

The model training and testing were performed using the RTextTools package [19]. Within the RTextTools package, classification algorithms used for the analysis included support vector machines (SVM), random forests, boosting, maximum entropy (MAXENT), and artificial neural networks (ANN). Models were generated using the 700 tweets for model training and the predictions based on the 146 tweets for testing were compared with the actual sentiment values. For all final model outputs the proportion accuracy (eq. 2), precision (eq. 3), and recall (eq. 4) were calculated.

$$\text{ACCURACY} = \frac{\text{TRUE POSITIVE+TRUE NEGATIVE}}{\text{POSITIVE+NEGATIVE}} \tag{2}$$

$$\text{PRECISION} = \frac{\text{TRUE POSITIVE}}{\text{TRUE POSITIVE+FALSE POSITIVE}} \tag{3}$$

$$\text{RECALL} = \frac{\text{TRUE POSITIVE}}{\text{TRUE POSITIVE+FALSE NEGATIVE}} \tag{4}$$

$$\text{SPECIFICITY} = \frac{\text{TRUE NEGATIVE}}{\text{FALSE POSITIVE+TRUE NEGATIVE}} \tag{5}$$

In addition, receiver operating characteristic (ROC) curves were determined u using the pROC package (version 1.8) within R Statistical Software. Within the ROC curve, the x-axis values represent (1 – specificity (eq. 5)). The y-axis values represent sensitivity, which is the same as recall (eq. 4).

4. FINDINGS

4.1. Twitter Data

Overall, the twitter data had a higher degree of negativity. The training data had tweets that were 43% negative, 26% neutral, and 31% positive. On a five-category scaled breakdown, the tweets were 9% very negative, 32% negative, 33% neutral, 23% positive, and 3% very positive. The test data had tweets that were 41% negative, 33% neutral, and 26% positive. On a five-category scaled breakdown, the tweets were 8% very negative, 36% negative, 26% neutral, 27% positive, and 4% very positive.

4.2. Accuracy Assessment

The examination of sentiment using five different algorithms resulted in very similar results for the analysis of positive/negative sentiment scores and for the scaled sentiment scores. The accuracy based on the training data set was higher than the accuracy determined from the test data set for both the positive/negative sentiment scores and the scaled sentiment scores. With support vector machines, random forests, maximum entropy, and artificial neural networks, the difference between the training and test data accuracy was over 0.4, whereas with boosting, the difference was no greater than 0.25. For the accuracy assessment of the positive/negative sentiment scores in Figure 1, the values ranged from 0.49 to 0.54. The precision and recall values were similar. The most accurate classification algorithm was the random forest algorithm at 54%. The least accurate classification algorithm was the boosting algorithm.

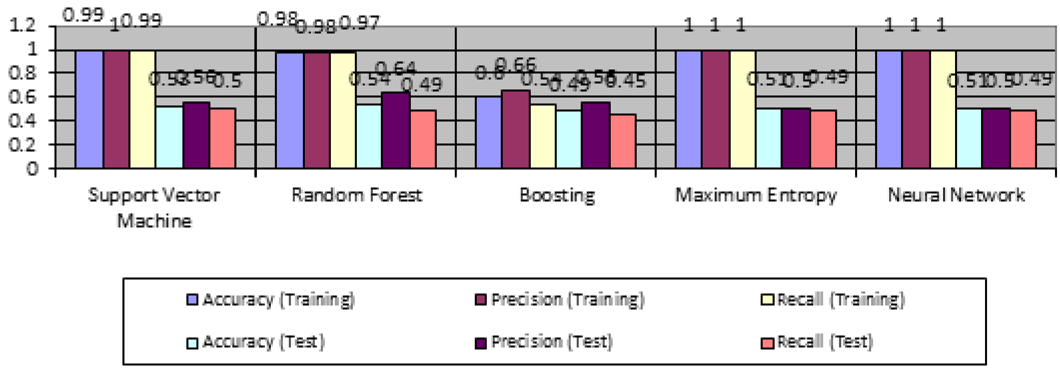

Figure 1 Accuracy assessment of the positive/negative sentiment scores

For the accuracy assessment of the scaled sentiment scores (Fig. 2), the values ranged from 0.38 to 0.40. The precision and recall values were similar. The classification algorithm for the scaled sentiment scores had lower accuracy values (10-15% less) compared to the positive/negative sentiment score classification. The most accurate classification algorithms were support vector machine, random forest, and the maximum entropy algorithms. The least accurate classification algorithm was the boosting algorithm.

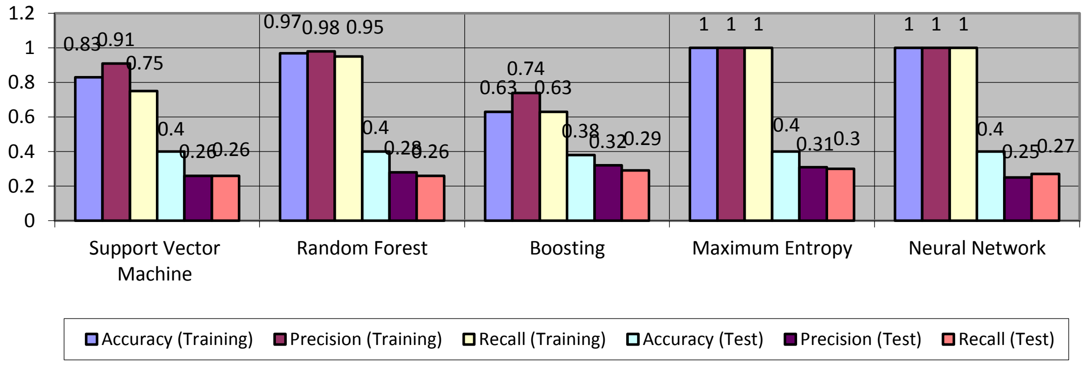

Figure 2 Accuracy assessments of the scaled sentiment scores

4.3. Receiver Operating Characteristic Curve

Figure 3 shows the ROC curves for each of the five classification algorithms for the positive/negative scores and for the scaled scores. The area under the curve for the positive/negative scores ranged from 0.545 to 0.644. The boosting algorithm provided the lowest AUC and the maximum entropy provided the highest AUC. The area under the curve for the scaled scores ranged from 0.531 to 0.588. The random forest algorithm provided the lowest AUC and the maximum entropy provided the highest AUC.

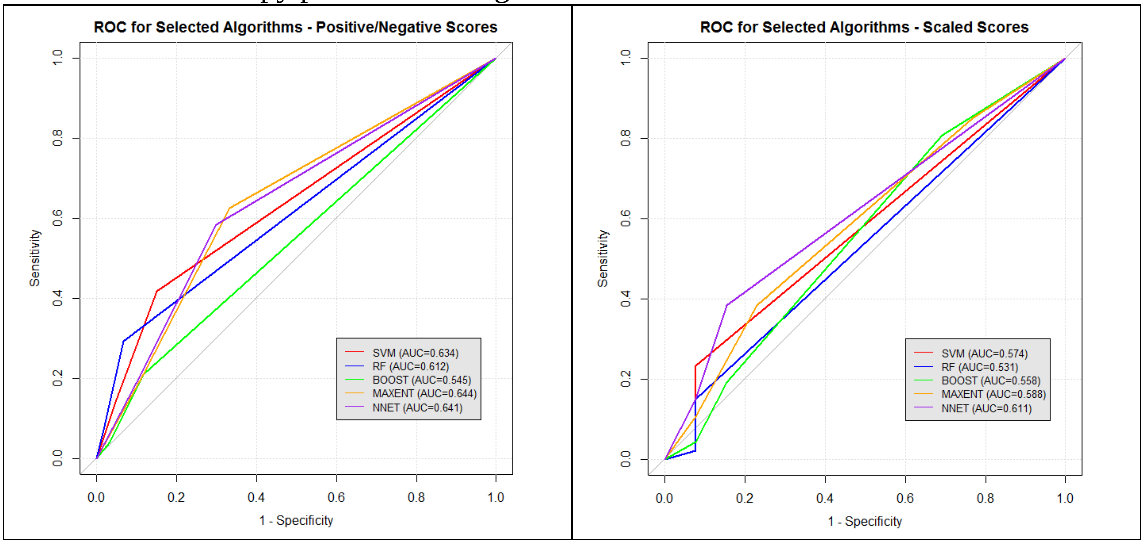

Figure 3 Receiver operating curves for classification of twitter sentiments using support vector machines, random forests, boosting, maximum entropy, and artificial neural networks for a) positive/negative scores and b) scaled scores

5. DISCUSSION AND CONCLUSION

The examination of accuracy from training and test data indicated the algorithms had a large degree of overfitting. This was especially true for support vector machines, random forests, maximum entropy, and artificial neural networks. Boosting had a lesser degree of overfitting. These results reinforce the need to have accuracy assessments on separate data from the data used to develop the model in order to avoid over-optimistic assessments of accuracy.

The level of accuracy declined with the increase in number of classes. The positive/negative score classification involved three classes (positive, neutral, and negative) and the scaled score classification involve give classes (very positive, positive, neutral, negative, and very negative). As more classes are added, the spatial extent of the classes was reduced, resulting in a smaller classification target footprint. The consequence would be greater likelihood of misclassifications.

The percent accuracy for all algorithms were similar and around 0.5-0.55 for the positive/negative scores and the 0.4 for the scaled scores. These accuracy scores were similar to what was observed before for other studies classifying text using machine learning algorithms. Previously, Hsu et al. (2010) recorded 47% accuracy for a five-category sentiment analysis project using SVMs. Similarly, Socher (2014) recorded 49.7% accuracy for the same five-category sentiment analysis. Also, Go et al. (2009) recorded 80.5% and 82.2% accuracy for classifying positive and negative (no neutral) sentiments using maximum entropy and SVMs, respectively. In addition, Pang et al. recorded accuracies of 72.8% to 82.9% using support vector machines on positive/negative sentiments (no neutral sentiments).

The main distinction between the higher accuracy measurements in past sentiment analyses with machine learning algorithms with the current effort is the number of classes. The studies recording accuracies in the 80% range dealt with only positive and negative comments (two classes) and did not consider neutral comments. The current effort include neutral comments. Thus, it was understandable why the accuracies were not as great with the current effort. The other studies that looked at five classes also reported less accurate assessments.

In light of the results from other studies and in consideration of the differences, the accuracy of the classifications from the current effort were within reason. With three classes, about 50-55% accuracy would be appropriate. After factoring in the advanced work performed to fine-tune the parameters, the level of accuracy for classification of sentiment using this dataset was high highest level expected.

Another phenomenon was the convergence in performance among the five algorithms. The differences among algorithms were relatively small. Other comparative studies have shown that different machine learning algorithms have had markedly different results [18-21]. The current effort did show differences in performance, but the differences suggest similar performance. There are natural questions on whether the five algorithms used were the best algorithms. These five represent the highest performing machine learning algorithms [18]. It is unlikely that another algorithm would have drastically greater level of performance for the given data set. The algorithms used within this study were controlled through commands within the RTextTools package. The commands within RTextTools facilitated pre-processing of the twitter feeds, removing two-letter words, punctuation, and numbers. The commands within RTextTools called upon machine learning algorithms in existing R packages, facilitating the use of multiple machine learning packages within a systematic approach. However, within R Statistical Software, alternate packages not called upon by RTextTools may result in different performances. For artificial neural networks, the neuralnet package may offer some improvements. For boosting, the gbm package may offer greater fine-tuning features and improvements. For random forests, the conditional inference forests of the party package may offer improvements over the known biases of the randomForest package.

A methodological issue may have been the scoring of the lexicon. The lexicon sentiment scores were assigned by one individual. Without having an established lexicon with associated emotion and sentiment values, the assignment of sentiment scores by just one individual was needed. Due to this methodological limitation, there may have been a bias. If the lexicon was developed by a larger group of people or have been peer-reviewed, the scoring may have been different. This may have resulted in a more clearly defined, consistent scoring system that would have led to clearer classifications.

The current work classified twitters sentiments with estimated accuracies of up to 55% for three-class outputs using five machine learning algorithms including support vector machines, random forests, boosting, maximum entropy, and artificial neural networks. The outputs were similar across the different algorithms. The current work advances sentiment analysis work using a comparative evaluative approach. Future work would involve ways to improve the outputs. The current work involved five machine learning algorithms. Whether these are the best algorithms needs further consideration in light of alternative versions of a particular algorithm. Future work could also use additional machine learning algorithms. In light of the comprehensive approach undertaken with the current work, which includes a diversity of algorithms and thorough efforts to fine-tune the algorithms, drastic improvements may be unlikely. Future work could also focus on improving the lexicon and the scoring within the lexicon. The sentiment scores assigned to the lexicon used for this study were not meant to be the final authority. Future improvements on scoring with more individuals would facilitate creation of a more robust lexicon and scoring system.

REFERENCES

[1] K. Ghag, and K. Shah. "Comparative analysis of the techniques for Sentiment Analysis." Advances in Technology and Engineering (ICATE), 2013 International Conference on. IEEE, 2013.

[2] R. Xia, et al. "Dual Sentiment Analysis: Considering Two Sides of One Review." IEEE Trans. Knowl. Data Eng. 27.8 (2015): 2120-2133.

[3] S. Shetty, et al. "System and method of implementing an object storage infrastructure for cloud-based services." U.S. Patent No. 9,135,269. 15 Sep. 2015.

[4] M. R. Islam. "Numeric rating of Apps on Google Play Store by sentiment analysis on user reviews." Electrical Engineering and Information & Communication Technology (ICEEICT), 2014 International Conference on. IEEE, 2014.

[5] M. A. Razzaq, A. M. Qamar, and H. S. M. Bilal. "Prediction and analysis of Pakistan election 2013 based on sentiment analysis." Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE Press, 2014.

[6] D. Arora, K. F. Li, and S. W. Neville. "Consumers." 2015 IEEE 29th International Conference on Advanced Information Networking and Applications (AINA). IEEE, 2015.

[7] Q. Gan, and Y. Yu. "Restaurant Rating: Industrial Standard and Word-of-Mouth--A Text Mining and Multi-dimensional Sentiment Analysis." System Sciences (HICSS), 2015 48th Hawaii International Conference on. IEEE, 2015.

[8] V. Pérez-Rosas, R. M., and L. Morency. "Utterance-level multimodal sentiment analysis." Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vol. 1. 2013.

[9] M. Wollmer, et al. Youtube movie reviews: In, cross, and open-domain sentiment analysis in an audiovisual context. IEEE Intelligent Systems, 2013, 99.1: 1.

[10] A. Sarlan, C. Nadam, and S. Basri. "Twitter sentiment analysis." Information Technology and Multimedia (ICIMU), 2014 International Conference on. IEEE, 2014.

[11] S. Bhuta, et al. "A review of techniques for sentiment analysis Of Twitter data." Issues and challenges in intelligent computing techniques (ICICT), 2014 international conference on. IEEE, 2014.

[12] N.F.F. da Silva, et al. "Using unsupervised information to improve semi-supervised tweet sentiment classification." Information Sciences 355 (2016): 348-365.

[13] V. K. Singh, et al. "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification." Automation, computing, communication, control and compressed sensing (iMac4s), 2013 international multi-conference on. IEEE, 2013.

[14] Z. Wang, et al. "Anomaly detection through enhanced sentiment analysis on social media data." Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on. IEEE, 2014.

[15] S. R. El-Beltagy, and A. Ali. "Open issues in the sentiment analysis of Arabic social media: A case study." Innovations in information technology (iit), 2013 9th international conference on. IEEE, 2013.

[16] R.M. Duwairi, "Sentiment analysis for dialectical Arabic." Information and Communication Systems (ICICS), 2015 6th International Conference on. IEEE, 2015.

[17] S. Demirci, Emotion analysis on turkish tweets. Diss. M. Sc. Thesis, Middle East Technical University, Ankara, 2014.

[18] R. Caruana, N. Karampatziakis, and A. Yessenalina. 2008. "An Empirical Evaluation of Supervised Learning in High Dimensions," in Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland.

[19] T.P. Jurka, L. Collingwood, and A.E., Boydstun. Grossman, E., and van Atteveldt, W. 2013. "RTextTools: a supervised learning package for text classification," The R Journal, 5(1), pp. 6-12.

[20] P. Pavan Kumar, N. Lakshmi Chandana, A. Sai Venu Madhav Reddy, Prof. D. Rajeswara Rao, A Comprehensive analysis on Sentimental Data Set Using Machine Learning Technique, International Journal of Mechanical Engineering and Technology 9(1), 2018, pp. 320–326. http://www.iaeme.com/IJMET/issues.asp?JType=IJMET&VType=9&IType=1

[21] Magesh G and Dr. P. Swarnalatha, Analyzing Customer Sentiments Using Machine Learning Techniques. International Journal of Civil Engineering and Technology, 8(10), 2017, pp. 1829–1842.
http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=10