RESEARCH ARTICLE

SENTIMENT ANALYSIS THROUGH TWEETS FOR 'DOUBLING FARMERS' INCOME' IN INDIA

[1]Ekta Hooda, [2]Akshay Hooda, [1,*]Hooda, B.K. and [2]NitinTanwar

[1]Department of Mathematics & Statistics, CCS HAU-Hisar, Haryana
[2]L M Thapar School of Management, Thapar Institute of Engineering & Technology, Derabassi Punjab (India)

ARTICLE INFO	ABSTRACT
	Twitter is one the fastest growing online social media where people express their views on various social and political issues. Public opinion and sentiments expressed through these tweets reflect the reactions and views of educated community on the topics of national importance. In the present study, recent tweets for the topics 'Agriculture' and 'Doubling Farmers'Income' in India have been analyzed and interpreted. These tweets have been extracted from a radius of 1000kms around Bhopal in Madhya Pradesh covering almost whole of India. Sentiment analysis of the tweets on the two topics has been performed for understanding minds of intellectual Indians. Analysis has been done using R software and its associated libraries. Sentiment scores, sparsity, bar plots and word clouds have been obtained for analysis of terms and key words in term document matrices for the Twitter data. Sentiment analysis showed highly positive sentiments for both topics, i.e. 'Agriculture' and 'Doubling Farmers'Income' followed by considerable negative sentiments.

INTRODUCTION

Present age is the era of internet. We can call it as the internet age because it has changed the way people express their ideas and views. These views and opinions are mainly expressed through social media websites and spread quickly among the internet and mobile users. These days, millions of people are using social media networking sites like Twitter, Facebook, Instagram and Google Plus to express their emotions, views and opinions. Due to the ever increasing interaction among the individuals on the internet, the social media websites are generating a large volume of sentiment rich data in the form of tweets, blog posts, comments, likes etc. Among the top social networking forums tweeter provides a quick and easy means for people to express their views. It is one of the richest source to know about public opinion and sentiments through social media. For each tweet it is important to determine the sentiment of the tweet whether is it positive, negative, or neutral. Moreover, tweets are limited to 140 characters in length and hence the Tweeter's emotion or feelings on particular topic remain consistent. The data generated through social media websites is vast and unstructured. However, it contains useful information which can be used for future planning.

*Corresponding author: **Hooda, B.K.,**
Department of Mathematics & Statistics, CCS HAU-Hisar, Haryana (India)

Government agencies and multinational companies are making use of these data for their policy decisions and improvements in their offerings. The users of social media websites also get an opportunity to connect the business persons and *vis-vis* the companies get an opportunity to know their customers more closely. Customers are also taking advantage of social media websites to check reviews of a particular product from some specific brand. Sentiment analysis is the process that converts views and opinions expressed in tweets and database sources with the help of Natural Language Processing (NLP), Kharde and Sonawane (2016). It categorizes the different opinions in text into three categories that are positive, negative and neutral. Sentiment analysis is also known by other names such as subjective analysis, appraisal extraction and opinion mining. It has been used as a Natural Language Processing task at many levels of granularity. Starting from a document level classification task (Turney, 2002), it has been used at the phrase level (Agarwal *et al*., 2009). Pang and Lee, (2008) described the existing techniques and approaches for an opinion-oriented information retrieval and reviewed the existing work on opinion mining and sentiment analysis. Sentiment analysis of Twitter data have been undertaken by Go *et al*. (2009), Barbosa and Feng (2010) and Bermingham and Smeaton (2010). Pak and Paroubek (2010) proposed a model to classify the tweets as objective, positive and negative. They created a twitter corpus by collecting tweets using Twitter API and automatically annotating those tweets using emoticons.

Davidov *et al*. (2010) proposed an approach to utilize Twitter user- defined hash tags in tweets as a classification of sentiment type using punctuation, single words, n-grams and patterns as different feature types, which are then combined into a single feature vector for sentiment classification. Several papers have been published in recent years on sentiment analysis of tweeter data and news articles. Most of these articles have devoted to methodological aspects or analysis of twitter data on camera, movie, mobile etc. Hooda and Hooda (2018) performed sentiment analysis of tweets for 'Agriculture' from three major of India. Agriculture sector is directly linked with the world's food security where crop forecasting is considered to a valuable tool for framing policies to address food security and poverty. Agriculture is the back bone of the developing countries including India. However, not much work has been done on analysis and understanding of public sentiments on the very important issue like agriculture and government policies. Also, India is an agriculture economy and present Indian government is exploring every nook and corner for doubling the income of. Therefore, in this study we have collected and analyzed data from twitter.com related to the recent tweets for the themes "Agriculture" and "Doubling of Farmers' income" from India.

MATERIAL AND METHODS

Two-recent datasets one each on 'Agriculture' and 'Doubling Farmers' Income' of 2393 and 1584 tweets, respectively were extracted using twitte R package in R studio during August 2018. The tweets were retrieved using the 'geocode' function of R by choosing Bhopal (Madhya Pradesh) as the central location and choosing a radius of 1000kms, assuming to cover majority of India. The retrieved datasets contained data regarding the actual text of the tweet, date and time of creation, tweet ID, screen name of the username, retweet count and the latitude- longitude of the source location of the tweet. R Studio GUI has been used for the analysis and the packages used are namely, twitter R, ROAuth, plyr, Stringr, tm (for text mining), word cloud and httr. A twitter application was created that provided an interface to connect R studio console with twitter.com using Twitter API. This API facilitates analysis of tweets posted by users and for extraction of tweets and underlying meta-data. The required tweets were extracted using search Twitter function in R Studio. The tweets data obtained by extraction was highly unstructured and needs to be cleaned for numbers, symbols, strip of white spaces, URLs, punctuations, emoticons etc. This made the unstructured data usable for further analysis. After satisfactory cleaning, the data converted into a Term-Document Matrix (TDM) to further list down the terms against their occurrences in the documents (tweets). The TDM was then used to form the wordcloud, barplot etc. of the frequent terms. Sentiment analysis was done via the nrc_sentiment_library to get the sentiment scores in the data.

RESULTS AND DISCUSSION

Analysis of Tweets for Agriculture

The TDM for 'Agriculture' contained 3485 terms in 2393 documents. The sparsity in TDM was near 100% indicating large number of cells with zero entries. The bar plot diagram (Fig. 1) the most frequent terms, here for those with frequency greater than70 for the Agriculture dataset shows the most frequent terms to be Israel, techniques, farmers, followed by

new, green, deserts, learnt, Jharkhand etc. with almost similar frequencies.
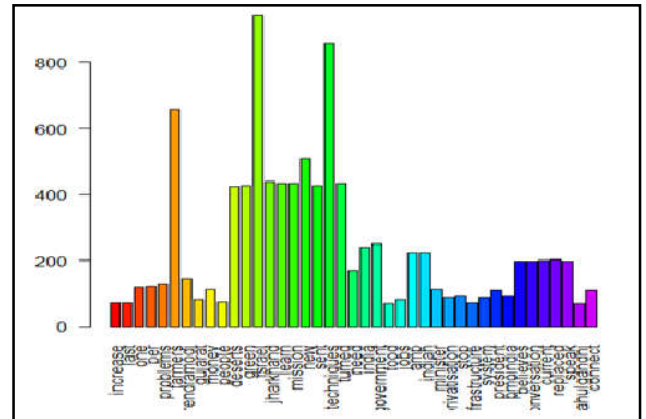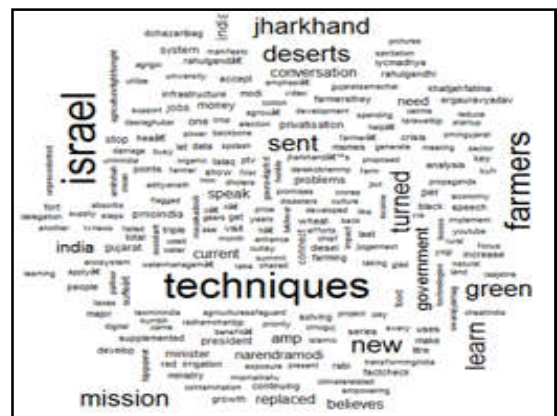

Fig. 1. Barpiot of words with frequency>70

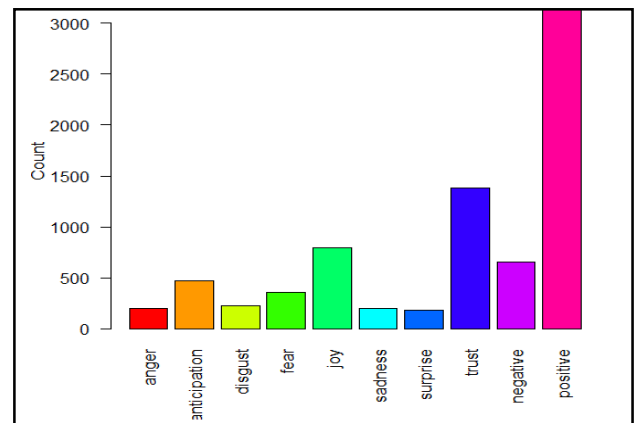**Fig. 2. Word cloud for 'agriculture'**

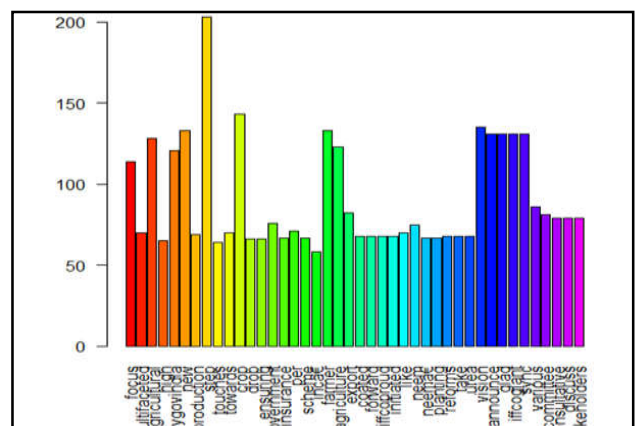**Fig. 3. Sentiments scores for 'agriculture'**

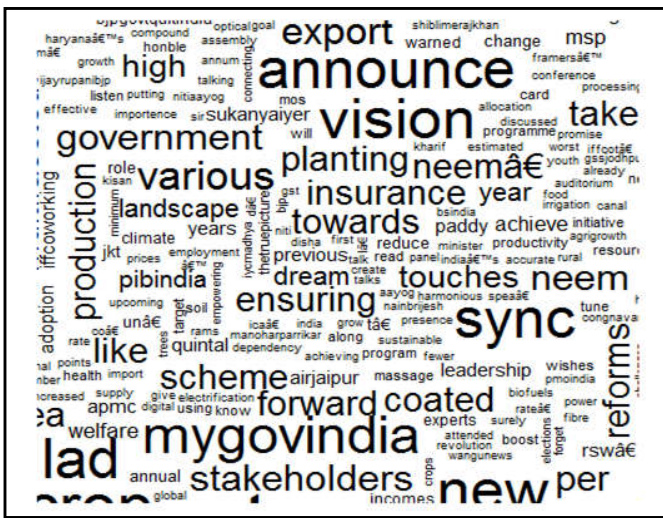**Fig. 4. Barplot of words with frequency > 50**

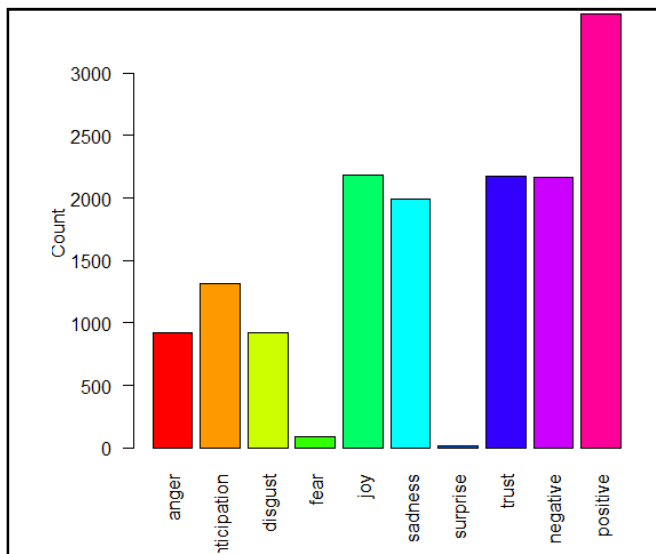**Fig. 5. World cloud for Doubling farmers 'Income'**

Fig. 6. Sentiments scores for 'Doubling Farmers' Income'

This similarity in the frequency pattern can be attributed to a recently trending tweet regarding 26 farmers from Jharkhand sent to Israel for learning new techniques regarding agriculture using their example of turning their deserts green. Also, Israeli techniques in agriculture sector are in much discussion amongst the public due to this latest venture, hence justifying the trend. The wordcloud (Fig. 2) obtained from the most frequent terms also visualises the same trend with the word size increasing with the increase in the frequency of the trending word such as Israli, techniques, farmers etc. The sentiment analysis scores (Fig. 3) for the "Agriculture" dataset shows a highly positive sentiment overall amongst the public keeping in view of the recent data with a count of above 3000 positive sentiments and a slight above 500 for the negative sentiment. Additionally, when sentiments are further classified in 8 categories, the trust sentiment takes a lead alongwith joy and anticipation, all of which are positive sentiments. Fear and disgust also follows but the positive sentiments prevail over a much larger part.

Analysis of Tweets for 'Doubling Farmers' Income: The second dataset comprises of 1584 tweets regarding the 7-point strategy of the Prime Minister for putting farmers first and thus aspiring to doubling of farmer's income by the year 2022. A tweet by MyGovIndia leads the trend and talks about a step taken towards doubling farmers' income as agricultural production touches new high in 2017-18 and also talks about a 'New Crop Insurance Scheme' with a slogan 'from per drop more crop (Fig. 4 & Fig. 5)'. Another trending tweet is regarding IFFCO taking a step forward encouraging neem-coated urea for the above mentioned initiatives. The sentiment scores (Fig. 6) for the theme of 'Doubling of Farmer's income' show a mixed sentiment but mostly towards the positive side with a count of nearly 4000 positive sentiments against a count of nearly 2100 negative sentiments as per the recent dataset. The positive sentiments of joy and trust have a good count but are closely followed by sadness, anticipation and disgust probably indicating that there is sense of trust for the initiative but in general the sentiment for the fulfillment of the targets is mixed amongst the public. Surprise and fear are ranked the least but the anger sentiment has a high count.

REFERENCES

Agarwal A., Biadsy F., and Mckeown K. 2009. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, March, 24–32.

Barbosa L. and Feng J. 2010. Robust sentiment detection on twitter from biased and noisy data. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 36–44.

Bermingham A. and Smeaton A. 2010. *Classifying sentiment in microblogs: is brevity an advantage is brevity an advantage? ACM*, 1833–1836.

Davidov D., Tsur O. and Rappoport A. 2010. Semi-supervised recognition of sarcastic sentences in twitter and amazon. *In Proceedings of the Fourteenth Conference on Computational Natural Language Learning, CoNLL '10*. Stroudsburg, PA, USA, 2010. Association for Computational Linguistics, 107–116.

Go A. Bhayani R. and Huang L. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford.

Hooda, E., and Hooda, A. 2018. Sentiment Analysis Through Tweets for 'Agriculture' in India. *Advances in Computer Science and Information Technology,* 5(2) : 86-91.

Kharde, V. A. and Sonawane, S.S. 2016. Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications,* 139(11): 5-15.

Pak A., and Paroubek P. 2010. Twitter as a corpus for sentiment analysis and opinion mining. *In Proceedings of LREC.*

Pang and Lee L. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, *In Proceedings of LREC.* 2(1-2):1–135.

Turney P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. ACL.

****