Indian J.Sci.Res. 20 (2): 341-345, 2018

ISSN: 0976-2876 (Print)
ISSN: 2250-0138 (Online)

POLARITY ANALYSIS OF INDIAN BANKS THROUGH TWITTER

PIYUSH BHARDWAJ[a1], RAMESH KUMAR[b] AND PAYAL PAHWA[c]

[ac]Department of Computer Science and Engineering, BPIT, GGSIPU, Delhi, India
[b]Associate Professor, Department of CSE, OPJS University, Churu, Rajasthan, India

ABSTRACT

Polarity analysis tells us about the opinions and emotions of any sentence. This analysis can be used for decision-making and strategic planning. In our research, we have performed polarity analysis on one of the biggest service industries in the world: Banking Industry. This work focuses on 8 best banks of India. There is a massive competition among these major banks to come to top. Our research deduces the most trusted banking company in the India. The data source used is Twitter and research is conducted on the basis of our approach defined in this paper. According to our results, SBI is the most preferred as well as most trusted bank in India.

KEYWORDS:

India is the second largest country in this world in terms of population. With such a large number of people, there are many banks available in India for providing all the banking needs to their customers. In the era of social media, many people judge a bank by reviews given by its customers on social media. In this paper, we have taken eight largest banks in India for our analysis. The source used for analysis is twitter and the method is executed in statistical package R.

Paper (Bilal *et al.*, 2016) researched on Roman-Urdu opinion mining using three known algorithms, Naive Bayes, Decision Tree and KNN using WEKA tool. Opinions were extracted from a blog. There results concludes that Naive Bayes is better than Decision Tree and KNN in terms of more accuracy, precision, recall and F-measure.

In paper (Altawaier and Tiun, 2016), the authors identified a simple and workable approach for Arabic sentiment analysis on Twitter. They used three techniques, Naive Bayes, Support Vector Machine and Decision Tree. Their results conclude that Decision Tree is better then the rest of two algorithms.

The author in (Salas-Zárate *et al.*, 2017) proposes an aspect-level sentiment analysis method based on ontologies in the diabetes domain. For calculating efficiency of their method, they used a Twitter corpus. They concluded that N-gram around method was the best. (Caton *et al.*, 2015) formalized Social Observatory for observing and measuring social indicators. The authors analyzed 54,665 posts and 231,147 comments. The authors were able to conclude how users interacted, with whom and at what volume.

The author (Kanavos *et al.*, 2017) develops a Machine leaning algorithm that exploits all hashtags and emoticons inside a tweet. The system was proved efficient, robust and scalable. Paper (Sodanil, 2016) proposes sentiment analysis for hotel Their results proved support vector machines as the most accurate algorithm.

(Desai *et al.*, 2012) studied tweets from Kidney Week 2011 to increase public awareness of kidney disease. The (Holmberg *et al.*, 2014) studied people's publishing and tweeting frequency, use of hashtags, language, and emotions on Twitter. Their analysis shows that astrophysicists address the different groups but that they do not talk to each other. (Cody *et al.*, 2015) determines collective sentiment for climate change news, events, and natural disasters. They concluded that natural disasters, climate bills, and oil- drilling decrease in happiness while climate rallies, a book release, and a green ideas increase in happiness. (Novak *et al.*, 2015) provided an emoji sentiment lexicon, called the Emoji Sentiment Ranking and draws a sentiment map of the 751 most frequently used emojis. The author also proposes Emoji Sentiment Ranking. Study (Chew and Eysenbach, 2010) monitor usage of terms ‘‘H1N1’’ versus ‘‘swine flu’’ over time, analysis of ‘‘tweets’’and validate Twitter as a real-time content, sentiment, and public attention trend-tracking tool. Tweets can be used for real-time content analysis and knowledge translation research, allowing health authorities to respond to public concerns.

In (Clark *et al.*, 2016), the author studies tweets of advertisements of electronic cigarette. They concluded decrease in positivity of Organic tweets as compared with automated tweets. In (Bian *et al.*, 2016), the author discusses trend and sentiment analysis of IoT from multiple Twitter data sources and validated these trends with Google Trends. The author concludes that people were positive about IOT.

The paper is organized as follows: Section 2

[1]**Corresponding author**

explains the methodology used for polarity analysis Section 3 analyses the results. Section 4 is the concluding section. Finally, future scope is given in Section 5.

THE PROPOSED APPROACH

A lexicon is a vocabulary of a person or a language. We have defined a lexicon-based approach that makes use of a dictionary of positive and negative words. The dictionary is prepared by understanding the semantics of the language. All positive and negative words in the sentence are assigned sentiment values from the dictionary. Next step is to pass the assigned values in our proposed algorithm. Finally, a combining function is applied to predict the overall polarity of a text or message.

There are many researchers who have defined their own lexicon-based approach. Many companies who collect feedback about their products directly from the users use these approaches.

We have defined our modified lexicon-based approach for sentiment analysis. In this approach, a dictionary of positive and negative words is prepared.

The algorithm for calculating the polarity of any tweet is as follows:

1. Clean the sentence (spelling correction, replace, modify, delete incorrect data etc.) using data cleaning tool.
2. Remove all articles, auxiliary verbs, numbers, symbols and stop words from the sentence.
3. Let P represents positive and N represents negative.
4. For every adjacent pair of remaining words in the sentence:

4.1 Compare both the word with positive and negative dictionary.

4.1.1 If first word is in positive dictionary and second word is in positive dictionary: P=P+2

4.1.2 If the first word is in negative dictionary and second word is also is negative dictionary then: P=P+1

4.1.3 If the first word is in negative dictionary and second word is in positive dictionary then: N=N+1

4.1.4 If the first word is in positive dictionary and second word is in negative dictionary then: P=P+1; N=N+1

5 Return P and N.

The above algorithm is applied to tweets related to eight largest banks in India. The results are further analyzed in next section.

RESULTS AND ANALYSIS

Our system is executed for eight largest banks in India. The time duration taken for execution is from 1st March 2017 to 20th April 2017. Fig.1 to Fig. 8 gives a sample of automated graphs generated for various banks.

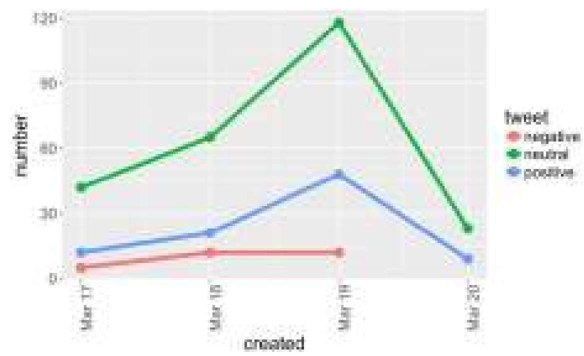

Figure 1: Sample of automated graph generated for HDFC bank in India

The above figure is an automated graph generated from our system to reveal a polarity opinion for HDFC bank in India. The results show more neutral tweets then positive and negative on various occasions.

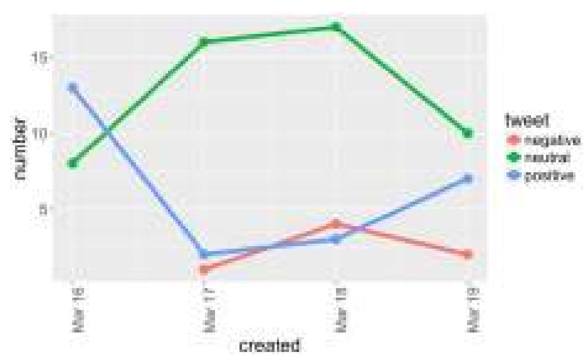

Figure 2: Sample of automated graph generated for State Bank of India in India

This automated graph shows more neutral tweets for State Bank of India in India from 16th March 2017-19th March 2017. Neutral tweets are followed by positive tweets and then by negative tweets.

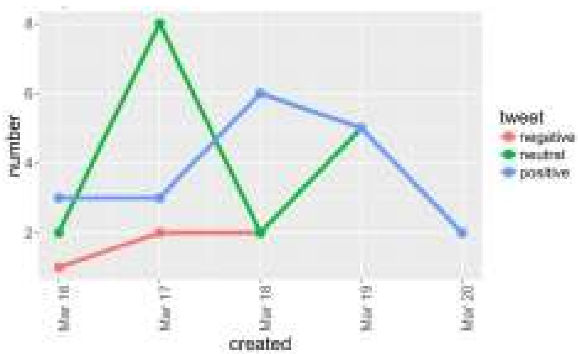

Figure 3: Sample of automated graph generated for Axis Bank in India

This graph shows that there are many positive and neutral tweets as compared to negative tweets for axis bank in India.

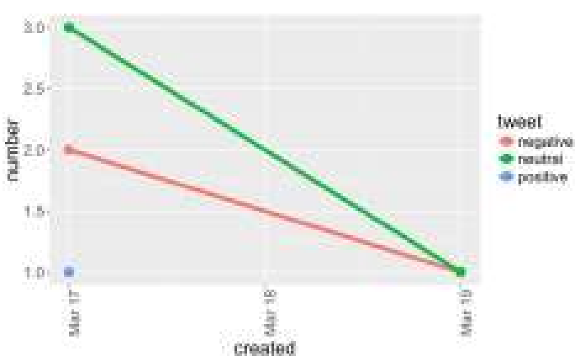

Figure 4: Sample of automated graph generated for ICICI Bank in India

The above graph depicts that people are neutral in their opinion about ICICI bank followed by negative opinion.

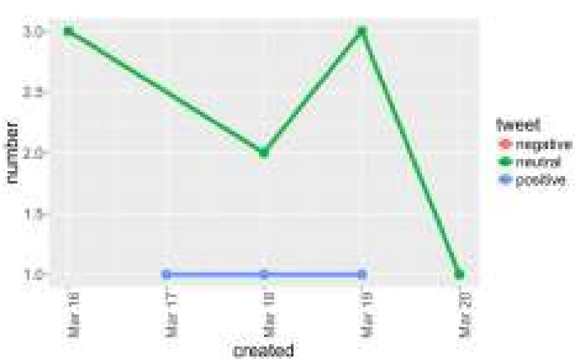

Figure 5: Sample of automated graph generated for IndusInd in India

In India, IndusInd bank shows aneutral polarity for most of the time followed by positive tweets. There are no negative tweets observed in the research duration.

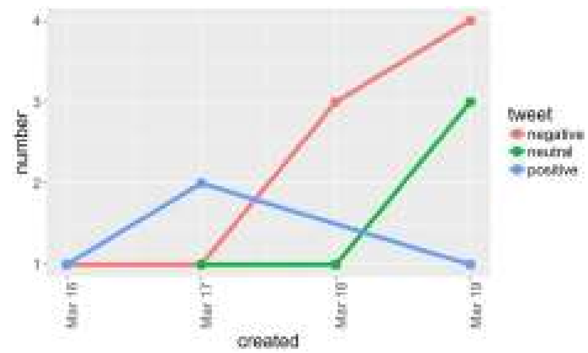

Figure 6: Sample of automated graph generated for Kotak Mahindra Bank in India

The above-automated graph depicts the polarity of tweets for Kotak Mahindra Bank in India for a given time frame. The graph is self concluding with results showing a negative tweets overpowering positive and neutral tweets.

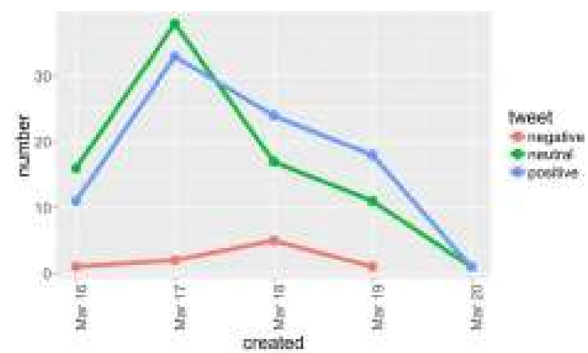

Figure 7: Sample of automated graph generated for Yes Bank in India

The above graph present an opinion for Yes Bank in India. It shows a maximum amount of neutral and positive tweets.

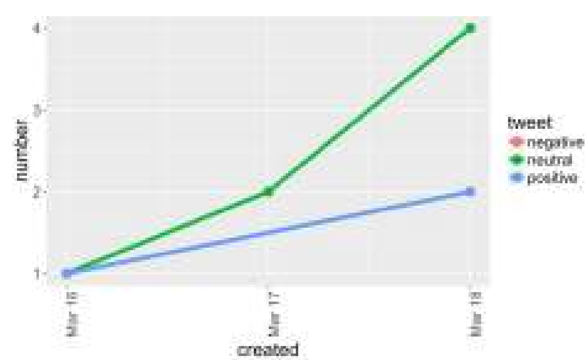

Figure 8: Sample of automated graph generated for Punjab national Bank in India

The above graph concludes the polarity of tweets for Punjab national Bank in India. A very minimal negative tweets and a good number of positive and neutral tweets makes Punjab National Bank a more strongerbank to search for.

CONCLUSION

Banking industry is one of the biggest service industries in the world. Our research provides a platform for billions of Indians to study fellow Indian'semotions and opinion about various banks. Our research concludes that State Bank of India is the most trusted bank in India, followed by HDFC and ICICI. Punjab National Bank falls last under this category with least number of positive tweets. The research helps banks to self-analyze their processes and products. The banking industry benefits from this research by learning their weaknesses in terms of positivity and negativity.

FUTURE WORK

The research opens a new way of analyzing tweets in banking industry. There is a large scope of making new discoveries in this field. The work can be taken further by analyzing sentiments and opinions on various banking companies in various other countries.

REFERENCES

Altawaier M.M. and Tiun S., 2016. Comparison of machine learning approaches on Arabic twitter sentiment analysis. International Journal on Advanced Science, Engineering and Information Technology, **6**(6): 1067-1073.

Bilal M., Israr H., Shahid M. and Khan A., 2016. Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques. Journal of King Saud University-Computer and Information Sciences, **28**(3): 330-344.

Bian J., Yoshigoe K., Hicks A., Yuan J., He Z., Xie M. and Modave F., 2016. Mining twitter to Assess the public perception of the "Internet of Things". PloS one, **11**(7): e0158450.

Clark E.M., Jones C.A., Williams J.R., Kurti A.N., Norotsky M.C., Danforth C.M. and Dodds P.S., 2016. Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter. PloS one, **11**(7): e0157304.

Caton S., Hall M. and Weinhardt C., 2015. How do politicians use Facebook? An applied Social Observatory. Big Data & Society, **2**(2): 2053951715612822

Cody E.M., Reagan A.J., Mitchell L., Dodds P.S. and Danforth C.M., 2015. Climate change sentiment on twitter: an unsolicited public opinion poll. PloS one, **10**(8): e0136092.

Clavel C. and Callejas Z., 2015. Sentiment analysis: From opinion mining to human-agent interaction. IEEE Transactions on Affective Computing, **7**(1): 74-93, from http://ieeexplore.ieee.org/document/7122903/

Chew C. and Eysenbach G., 2010. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PloS one, **5**(11): e14118.

Desai T., Shariff A., Shariff A., Kats M., Fang X., Christiano C. and Ferris M., 2012. Tweeting the meeting: an in-depth analysis of Twitter activity at Kidney Week 2011. PloS one, **7**(7): e40253.

Holmberg K., Bowman T.D., Haustein S. and Peters I., 2014. Astrophysicists' conversational connections on Twitter. PloS one, **9**(8): e106086.

Hua W., Wang Z. and Wang H., 2016. Understand short texts by harvesting and analyzing semantic knowledge. IEEE Transactions on Knowledge and Data Engineering, **29**(3): 499-512, from http://ieeexplore.ieee.org/document/7476863/

Kanavos A., Nodarakis N., Sioutas S., Tsakalidis A., Tsolis D. and Tzimas G., 2017. Large Scale Implementations for Twitter Sentiment Classification. Algorithms, **10**(1): 33.

Medhat W., Hassan A. and Korashy H., 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, **5**(4): 1093–1113, from http://www.sciencedirect.com/science/article/pii/S2090447914000550

Novak P.K., Smailović J., Sluban B. and Mozetič I., 2015. Sentiment of emojis. PloS one, **10**(12): e0144296

Salas-Zárate M.D.P., Medina-Moreira J., Lagos-Ortiz K., Luna-Aveiga H., Rodríguez-García M.Á. and Valencia-García R., 2017. Sentiment Analysis on Tweets about Diabetes: An Aspect-Level

Approach. Computational and mathematical methods in medicine, 2017.

Sodanil M., 2016. Multi-Language Sentiment Analysis for Hotel Reviews. In MATEC Web of Conferences, EDP Sciences, **75**: 03002.