

Evoking Emotions from Tweets: lexicon based method of Sentimental Analysis

Dr. Uttamkumar M. Kinange,

Professor, Kousali Institute of Management Studies, Karnatak University, Dharwad.

Vijay Joshi,

Research Scholar, Kousali Institute of Management Studies, Karnatak University, Dharwad

Abstract

Twitter a micro-blogging platform has gained huge popularity for its instant and quick information diffusion feature. Marketers are increasingly using social media to disseminate information related to their products and promotions on the social media. The robust growth of banking sector and its convergence with the new technology has brought in the revolutions in the economy. The purpose of the study is to ascertain the sentiment of the customers of India's leading private bank on the popular social media platform "Twitter". The study collects 5000 recent tweets of hdfc bank from its customer care twitter handle. The article uses R software for analysis and R studio for graphical interface to perform NLP - sentiment analysis using the "syuzhet" package given by saif mohammad which is based on NRC lexicon. Findings of the study reveal that people have positive sentiment towards hdfc bank and customers trust the bank. The findings also reveal the presence of considerable negative sentiment and display of sadness emotion. Emotions of anticipation are also on the higher side which state that people have higher expectations from the bank and its products and services. The study provides insights to managers of hdfc bank by suggesting them to reduce negative sentiment and sadness emotions of the customers by handling issues and complaints more effectively and indulge more in promotional activities which could surprise and delight their customers since these emotions are on a lower score. The study also states its shortcomings and also hints future researchers by providing limitations and avenues of future research which is across platforms and sector.

Keywords: sentiment analysis, consumer engagement, Twitter, social media, Retweets, emotion, R software.

1. Introduction

Digital 2020 report by Hootsuite states that with a 50% penetration, there are 687 million internet users in India as on January 2020 which clocked a 23 percent increase from the previous year of 2019, out of which the users of social media stood at 400 million. The average daily time that an Indian internet user spends on using internet is 6 hours and 30 minutes out of which the time spent on social media is 2 hours 24 minutes and Twitter has a reach of 11.45 million people (kemp, 2020). Social media plays an active role in the day to day activities of people in various aspects of life like family, friends, colleagues, social life, the choices and likes, exchanging information about knowledge, politics, sports, entertainment, products and services and so on.

Banking has always been an important need for all categories of people and become a requirement of almost all major financial transactions. Public sector banks have a major role to play in Indian economy but private banks are also not lagging behind. On the basis of market capitalization, HDFC bank is India's leading bank with nearly 6,000 billion rupees followed by kotak Mahindra bank having 2,800 billion market capitalization and ICICI bank having 2,600 market capitalization (kemp, 2020).

49% of social media users are on the popular micro-blog Twitter. The instant responsiveness and quick diffusion of information and micro-blogging feature of the twitter increased the popularity of the platform. For the present study we have extracted the data from the official customer care handle of HDFC bank "@HDFCBank_Cares". We have selected the customer care handle of the bank because since our aim to ascertain the customer sentiment and the customer care handle majorly deals and interacts with the customers directly and promotes more user generated content.

The study aims to analyse the tweets of India's leading private bank and understand the sentiments of its customers by performing sentimental analysis resulting into categorization of emotions of the collected tweets. These emotions would provide an insight to understand the attitude and intentions of the customer behavior. A service satisfaction would lead to positive behavior whereas the service failure would result in negative emotions and negative electronic word of mouth (Achar et al., 2016; Jean, 2019).

2. Related Literature

Due to the advent of technological revolution and digitalization, enormous amount of data is being generated and distributed virtually across various platforms in the fields like politics, social environment, sports, businesses, and stock market and so on. This huge data comprises both qualitative and quantitative characteristics. The analysis of quantitative data is not as tedious and critical as the qualitative data. Hence the sentiment analysis of qualitative data plays an important role in this data era (Saad & Saberi, 2017).

Sentiment analysis is also known as mining of emotions, review, sentiment, opinion and text. But the most widely used words many a times interchangeably denoted by opinion mining or sentiment analysis (Bing, 2012). Saad and saberi define sentiment analysis as "Opinion Mining (OM) or Sentiment Analysis (SA) can be defined as the task of detecting, extracting and classifying opinions on something" (Saad & Saberi, 2017).

Companies in various industries like aviation industry have been using the twitter not only to diffuse information about their services but also they are using Twitter to ascertain the service quality with the help of user generated content available on social media platforms (Martin-Domingo et al., 2019). Data from blogs have been extracted to ascertain the level of satisfaction of the customers of airport industries on different airport (Gitto & Mancuso, 2017).

People seek opinions of friends, family and experts before their purchase decision and these opinions include emotions, information, entertainment and the like. Opinions induce the behavior of the customers (Soong et al., 2019).

A systematic arrangement of social media opinions is required as they provide inclusive information generated by the consumers hassle free and in huge volume. Such a systematic arrangement with advanced tools called sentiment analysis is continuously improving with its newer dimensions (Yue et al., 2019).

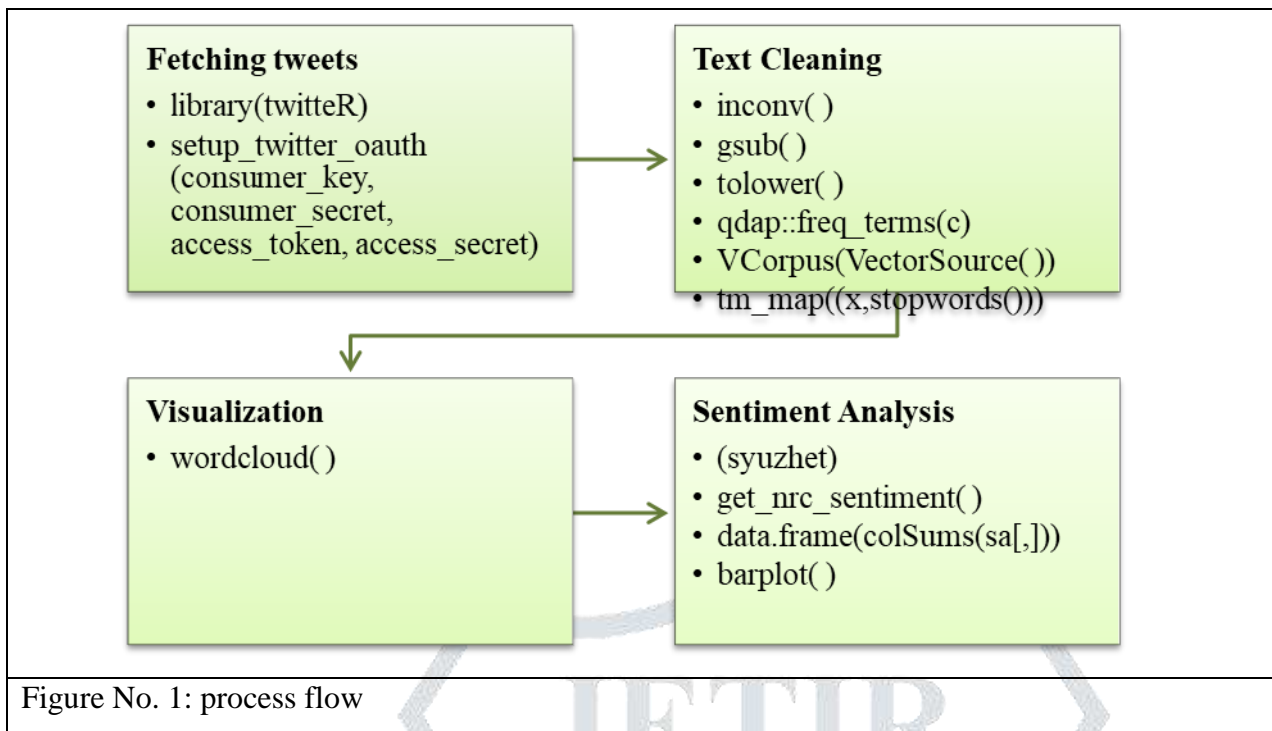
Twitter platform has gained interest of the academicians in the recent past as huge amount of growing activity is seen both by the marketers and the consumers on the platform. Since Twitter poses a limitation on the length of tweet on the basis of number of characters, analysis of twitter sentiment becomes a challenging job. Sentiment on twitter can be analysed with four different classes machine learning, lexicon based, hybrid (lexicon and machine learning) and graph based (Giachanou & Crestani, 2016). Twitter sentiment analysis initially began with only polarizing the tweets with either positive or negative or neutral. Later many data scientists have developed algorithms to skim other emotions and sentiments out of the tweets (Tang et al., 2014).

Both supervised and unsupervised machine learning is used to analyse the sentiments through the lexicon based method to polarize the sentiments. Many people have contributed by constructing sentiment lexicons and lexicons are successful only when they can predict emotions as their out accurately (Vo & Zhang, 2016).

Mahajan and Rana used NRC lexicon approach and investigated the polarity of emotions and arrived at the words representing the emotions which are called as sentiment tokens and classified them into positive, negative and neutral. They used syuzhet package in R to arrive at sentiments by assigning plus five to minus five scores and listed eight emotions from their corpus (Mahajan & Rana, 2018).

3. Methodology

We follow NRC emotion lexicon method to analyse tweets with R software with R studio as interface given by saif mohammad (Mohammad & Turney, 2010). The study follows the qualitative approach by conducting sentiment analysis of text in tweets. The study used R software and R studio interface to fetch the tweets of hdfc bank. 5000 recent tweets were collected from the bank's official customer care twitter handle "@HDFCBank_Cares" through twitter API. The steps of sentiment analysis, process flow and the functions used to analyse the codes are shown in the figure no. 1.



4. Data Analysis

Data Extraction

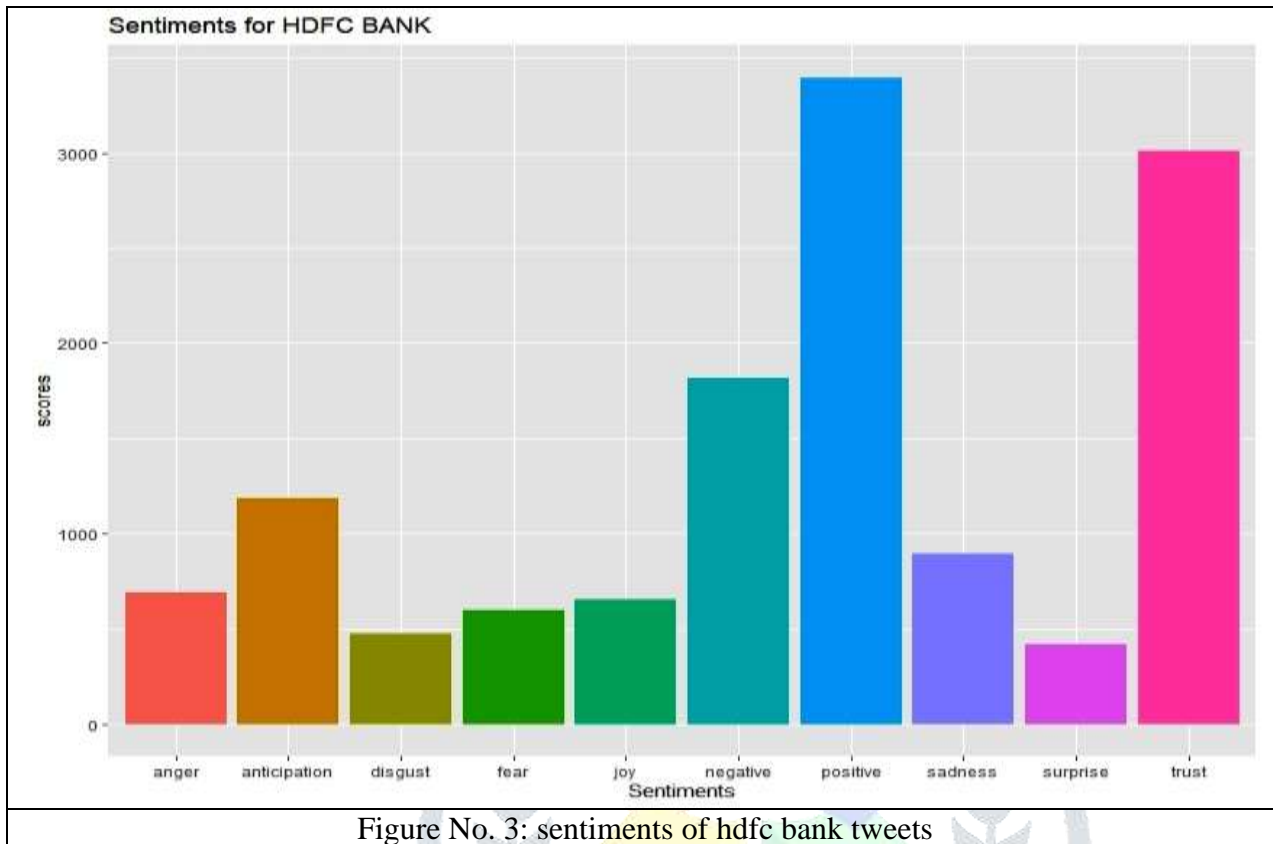
The data was converted into data frame and tweets were extracted by using “twListToDF()” and “x=a\$text” functions.

Data cleaning

Text cleaning was done by converting tweets from ASCII to tackle using “iconv()” function. Retweets and mentions were removed and text was converted to lower case. Numbers, single words, special characters, tabs, links, punctuations, user names and blank spaces in at the beginning and at the end were removed. After cleaning the data we noted most occurring frequencies as in the table no 1.

Sl. No	WORD	FREQ
1	to	1993
2	the	1606
3	my	1357
4	is	1185
5	you	1042
6	and	944
7	have	899
8	for	879
9	of	781
10	not	779
11	card	737
12	hdfc	728
13	on	698
14	your	656
15	in	637
16	bank	566
17	from	546

anticipation was found to be associated with 1,185 words. 892 words were associated with sadness whereas 691 words displayed anger emotion. Sentiment of joy was found to constitute by 650 words and 599 words represented fear. Disgust and surprise were recorded to be associated with lesser number of words that is 471 and 419 respectively.



6. Discussions and Managerial Implications

The word cloud represents the repetitive words where we can find out “card” and “hdfc” are the most used words in tweets revealing the focus towards the brand name (bank name) and the products (card) followed by “bank”, “credit” and “customer”. The second layer of sentiment flows with the words like “please”, “account”, “service”, “emi”, “moratorium” and further “issue”, “care”, “loan”, “response” and “team” where in the focus is towards request and exchange of information about the services and about issues raised and about solving them (“team”). Positive sentiment is reflected by the third layer of sentiment like “thanks”. Some negative sentiment is also highlighted in the word cloud like “unable”, “worst” and “complaint”. Thus we can conclude that the overall sentiment of the bank is highly on the positive side and customers trust the company and its products and services, though issues in services cannot be ignored as the third highest indicator of the sentiment on the negative side. Anticipation is on a higher note which says that people have lot of expectations on the products and services of the bank before sales and this sentiment is followed by “sadness” which might indicate that the bank is not performing as per the expectations of the customers.

The study reveals various insights to the managers of the hdfc bank by analysing the sentiments of the tweets. Managers are advised to retain the higher level of positive sentiment and trust expressed by the customers via the

twitter platform so that the bank sustains its top position in the market. Managers are advised to be careful about the negative sentiment and sadness expressed by the customers on the twitter platform. Words highlighted in the word cloud like “issue”, “worst” and “complaint” alarm the need to provide better service to the customers and resolve their complaints quickly and smoothly since there are lot of expectations on the bank. Sentiments like “surprise” and “joy” are on the lower side which is not a good sign hence the managers can push in the promotional strategies which could surprise its customers and bring them delightful feeling.

7. Limitations and Future Research

The study exclusively analyses the recent tweets of hdfc bank’s customer care handle on twitter and reveals the sentiments of those tweets. The study suffers with certain limitations some of which might open an opportunity to conduct further research based on these limitations. The scope of the study is limited to only one bank which belongs to the private sector. The study is based on Twitter platform of social media and ignores other platforms like Facebook and Instagram where the customers might be more active with their sentiments. The study includes tweets generated by both the bank and its customers. With these limitations, the study directs future researchers to study the content generated by users only. Future research can also be carried out on other platforms of social media like Facebook and also can compare both the platforms. Future researchers also can take up a comparative study of different banks in the private sector or may compare with the public sector banks.

8. References

- Achar, C., So, J., Agrawal, N., & Duhachek, A. (2016). What we feel and why we buy: the influence of emotions on consumer decision-making. *Current Opinion in Psychology*, 10, 166–170. <https://doi.org/https://doi.org/10.1016/j.copsyc.2016.01.009>
- Bing, L. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Giachanou, A., & Crestani, F. (2016). Like It or Not: A Survey of Twitter Sentiment Analysis Methods. *ACM Comput. Surv.*, 49(2). <https://doi.org/10.1145/2938640>
- Gitto, S., & Mancuso, P. (2017). Improving airport services using sentiment analysis of the websites. *Tourism Management Perspectives*, 22, 132–136. <https://doi.org/https://doi.org/10.1016/j.tmp.2017.03.008>
- Jean, H.-W. L. (2019). The effect of consumer emotions on outcome behaviors following service failure. *Journal of Services Marketing*, 33(3), 285–302. <https://doi.org/10.1108/JSM-04-2018-0124>
- kemp, simon. (2020). *DIGITAL 2020: INDIA*. <https://datareportal.com/reports/digital-2020-india>
- Mahajan, P., & Rana, A. (2018). No Title. *Journal of Sociotechnology and Knowledge Development (IJSKD)*, 10(1), 219–224. <https://doi.org/10.4018/IJSKD.2018010104>
- Martin-Domingo, L., Martín, J. C., & Mandsberg, G. (2019). Social media as a resource for sentiment analysis of Airport Service Quality (ASQ). *Journal of Air Transport Management*, 78, 106–115. <https://doi.org/https://doi.org/10.1016/j.jairtraman.2019.01.004>
- Mohammad, S., & Turney, P. (2010). Emotions Evoked by Common Words and Phrases: Using Mechanical

- Turk to Create an Emotion Lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*.
- Saad, S., & Saberi, B. (2017). Sentiment Analysis or Opinion Mining: A Review. *International Journal on Advanced Science, Engineering and Information Technology*, 7(5), 1660–1666. <https://doi.org/10.18517/ijaseit.7.5.2137>
- Soong, H., Jalil, N. B. A., Ayyasamy, R. K., & Akbar, R. (2019). The Essential of Sentiment Analysis and Opinion Mining in Social Media : Introduction and Survey of the Recent Approaches and Techniques. *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 272–277. <https://doi.org/10.1109/ISCAIE.2019.8743799>
- Tang, D., Wei, F., Qin, B., Liu, T., & Zhou, M. (2014). {C}oooolll: A Deep Learning System for Twitter Sentiment Classification. *Proceedings of the 8th International Workshop on Semantic Evaluation (S{em}{E}val 2014)*, 208–212. <https://doi.org/10.3115/v1/S14-2033>
- Vo, D. T., & Zhang, Y. (2016). Don't count, predict! an automatic approach to learning sentiment lexicons for short text. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 219–224.
- Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A Survey of Sentiment Analysis in Social Media. *Knowl. Inf. Syst.*, 60(2), 617–663. <https://doi.org/10.1007/s10115-018-1236-4>

