# Twitter Sentiment Analysis on Airline Tweets in India Using R Language

**Sreeja I, Joel V Sunny, Loveneet Jatian**

i.sreeja@mba.christuniversity.in, joel.sunny@mba.christuniversity.in, loveneet.jatian@mba.christuniversity.in

**Abstract.** Every day, passengers from all parts of the world are carried over to various destinations in about 100,000 flights. And their experience in the journey is almost unknown to the airlines. Gaining knowledge on customer opinions on the airlines is very crucial as it is very difficult for the Indian Airlines to collect customer feedback. Also, the understanding of the degree to which customers are satisfied is vital for the strategy development for the company. And hence, twitter data is utilized for the same purpose. Twitter is the best possible to way to obtain such information. This paper focuses mainly on the analysis of customer views on Indian Airlines services. Data visualization is used here for displaying customer feelings which can be anger, fear, anticipation, trust, surprise, sadness, joy and disgust. Whole together their opinion can be sentiments such as positive and negative.

## 1. Introduction

The online activity of people across the world has increased rapidly in the past few years. People post their opinions on a variety of subjects. Companies are making use of this information as a competitive advantage through social media. Many people are expressing their emotions, opinions, feelings and disclose about their daily life. Around million posts are being updated every second on social media. This results in tremendous amounts of data that can be put into great use for various gains. One such platform is twitter. People often resort to express their emotions about a brand, a product or a service. The twitter data is put into use by the companies for the understanding of their customers.

Customer feedback is one of the important factors which can help to improve the airline services. Collection of this feedback using questionnaires can be a time-consuming process and may also encounter false information. And hence Sentiment analysis is used to help the companies in various ways.

Sentiment Analysis is one of the text analysis techniques which distinguishes and extracts abstract data from the source. It helps to understand the emotions of the people on the service/product. Sentiment analysis can also be called as opinion mining because people opinions are mined through it. This paper will help us know which Indian Airlines are more positively/negatively known by their customers and which airlines resonates well in the audience. Sentiment analysis may provide answers to many of the company's questions regarding its success and growth or decline of sales. It can help the airlines to know their service performance and handle customer complaints through which strategy analysis can be performed. Also, effective marketing campaigns can be planned accordingly by taking this sentiment

analysis into consideration. Sentiment analysis is widely used to analyse customer needs and delivering to the customer expectations. Business problems can be anticipated and solved by the airline's companies using sentiment analysis.

The sentiment analysis on Indian Airlines is done using R programming in this paper by using a step by step approach. There are many packages in R which can be made of proper use for the extraction, pre-processing and analysing the data from twitter which is why R tool has been chosen in this paper. R is great for creating plots of exploratory data analysis and data visualization. R Programming by offering a set of inbuilt functions and libraries lets us visually analyse the data through data visualization.

## 2. Literature review

Previous papers talk about Sentiment Analysis being a process of extracting the information about the emotions of the people. It analyses the feelings of the people through their comments (positive or negative), questions and requests.

Sentimental analysis has many applications for different domains which were discussed earlier. Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity.

Text understanding is a significant problem to solve. Some machine learning techniques, including various supervised and unsupervised algorithms, are being utilized for the same purpose.

The previous researches gave an idea about how natural language processing can be used to extract twitter data and the various processes of analysing the sentiments from text data. Apart from the process, the research has shown the importance of the public opinions in decision making of any organization. This research paper is based upon the sentiment analysis of the tweets about the Airlines in India. The analysis is on the feelings of the people for the Airlines operating in India. Different variables which express the feelings of the people through the tweets they tweet are analysed in this study. Variables used are anger, disgust, joy, fear, negative, positive, surprise and trust as these are the emotions expressed by customers according to the research. These variables help to analyse the feelings of people of India and what they think about the Airlines operating in their country. Sentiments of the people towards Airlines would be a good area to analyse the feelings of the people as it plays a major role in travelling. Specially for the employees who are working for business organizations, this analysis provides revealing insights as they usually have to travel from one place to other in and around the country and sometimes outside India as well. Middle class people who travel occasionally also want to get the best services from the Airlines in the least price possible. Discounts available for the tickets also affect the sentiments of the people towards the Airlines. The data for this research is collected from Twitter as it is now-a-days, one of the best sources to collect the data about the sentiments of the people. Twitter messages and tweets tell a lot about the feelings and sentiments of the people towards any topic. Polarity of the tweets towards positivity or negativity can provide a good data to analyse the sentiments and views of the people towards the Airlines. The three airlines on which this paper does a sentiment analysis is on Air India, Indigo, and SpiceJet as these three are one of the well-known Indian airlines. Previous researches talk about numerous methods of cleaning of textual data and also several techniques used for sentiment analysis. Further analysis is done in this paper to obtain the degree to which their customers express their feelings on social media and this degree is measured in terms of variables such as anger, disgust, joy, fear positive, negative, surprise and trust.

## 3. Proposed work

Natural languages are different from programming languages. The semantic or the meaning of a statement depends on the context, tone and a lot of other factors. Unlike programming languages, natural languages are ambiguous. Text mining deals with helping computers understand the meaning of the text.

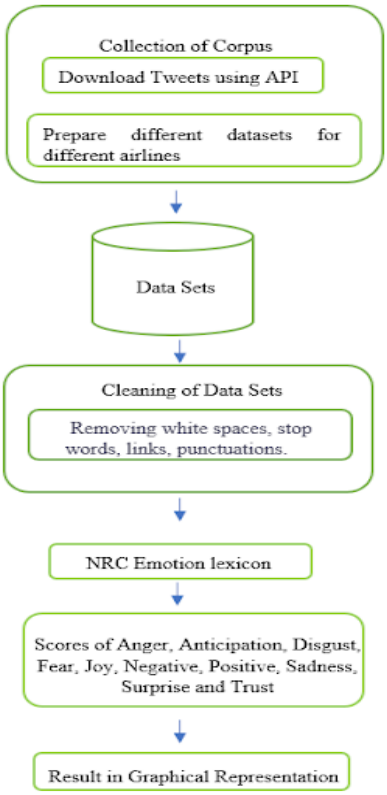


**Figure 1.** The figures shows the methodology to mine the twitter data and to carry out the analysis.

*3.1. Twitter Authentication*

Before mining any data from Twitter using APIs, Twitter authentication needs to be done using an application created on twitter. Once the application is created, the access to consumer key, consumer secret, access token, and access secret are obtained using which the API has to authenticate itself with Twitter Authentication server.

consumer_key<-'xxxxxxxxxxx'
consumer_secret<-'xxxxxxxxxxx
access_token<-'xxxxxxxxxxx'
access_secret<-'xxxxxxxxxxx'

*3.2. Access twitter data sets*

Once API is authenticated with Twitter Authentication service, a token is generated and is made available to API for every transaction with twitter server. Using this token, tweets are mined using hashtags which involve the names of the airlines. The function searchTwitter() is applied to access the data.

*3.3. Collection of corpus*

The datasets are directly downloaded from the microblogging website Twitter using the Twitter API and "twitteR" package in R. Since the different airlines in India are considered, tweets about different airlines like Air India, Indigo and Spicejet generally contains hashtags such as "#indigo","#airindia" and "#spicejet". The tweets having these hashtags are first downloaded. After this, the tweets are divided into different datasets.

Then, from each of the datasets, 1000 tweets are randomly selected which consists of all emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and sentiments (positive, negative).

This method of corpus collection can be used for collecting any kind of datasets in many different languages because twitter API allows to select the languages of the tweets to be collected.

*3.4. Text pre-processing*

Before analyzing the data, text pre-processing has to be done. Text data contains white spaces, punctuation, stop words etc. These characters do not convey much information and are hard to process. For example, English stop words like "the", "is" etc. do not tell give much information about the sentiment of the text, entities mentioned in the text, or relationships between those entities.

The following steps were performed to make sure that the text mining in R which is dealt here is clean:
• Convert the text to lowercase, so that words like "write" and "Write" are considered the same word for analysis
•Remove numbers
•Remove English stop words e.g "the", "is", "of", etc
•Remove punctuation e.g ",", "?", etc
•Eliminate extra white spaces
•Stemming the text

Stemming is the process of reducing inflected (or sometimes derived) words to their word stem, base or root form. e.g changing "aeroplane", "aeroplanes", "aeroplane's", to "aeroplane". This can also help with different verb tenses with the same semantic meaning such as fly, flying, and fly. "Tm" package is used to perform this operation. The main structure for managing documents in tm is called a Corpus, which represents a collection of text documents.

Here, for Indigo corpus - all of Indigo tweets are stored in the variable "indigo_tweets.text.corpus", for Air India corpus - all of Air India tweets are stored in the variable "airindia_tweets.text.corpus" and for Spicejet corpus - all of Spice Jet tweets are stored in the variable "spicejet_tweets.text.corpus".
Modification of the words within the tweets is done using with the techniques which were discussed above. These include stemming, stopword removal and others as mentioned. The 'tm' library is used for this purpose. Transformations are done via the tm_map() function which applies a function to all elements of the corpus. Basically, all transformations work on single text documents and tm_map() just applies them to all documents in a corpus. To convert all the text of Indigo's tweets, Air India tweets and Spice Jet tweets into lowercase at once, the tm library and the techniques mentioned below are used to do so easily.

*3.5. Technique used - sentiment analysis*

Sentiment analysis is the process of determining whether a piece of writing is positive, negative or neutral. Here, "syuzhet" package is worked upon to the purpose of this analysis. This library uses NRC Emotion lexicon for the classification of the tweets after text preprocessing. The NRC emotion lexicon is a list of words and their associations with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The get_nrc_sentiment() function returns a data frame in which each row represents a sentence from the original file. The columns include values for each emotion type as well as the positive or negative sentiment valence drawn from the tweets. It allows us to take a body of text and return the emotions the text represents and also whether the emotion is positive or negative.

*3.6. Graphical representation*

Now the sentiment analysis can be represented in graphical modes, with the usage of R-studio. There are a rich set of graphical tools supported by R packages which can be used to represent the effective and attractive outcomes of the sentiment analysis. In this paper, ggplots are used as they give a clear representation with less number of lines of code and also providing access to modifications such as color and size.

## 4. Experiments and results

In this research analysis, data is collected from twitter for the Airlines and the mainly focused three airlines here are SpiceJet, Indigo and AIR India. Analysis is on the basis of the variables used in the data. Variables are anger, positive, negative, disgust, anticipation, fear, joy, surprise, sadness and trust. Analysis is done on the sentiments of the people for the three airlines using the mentioned variables. The problems faced by the people who were not satisfied with the experience of the airlines are also considered in the analysis. The following are the sentiments of the people for each airline:

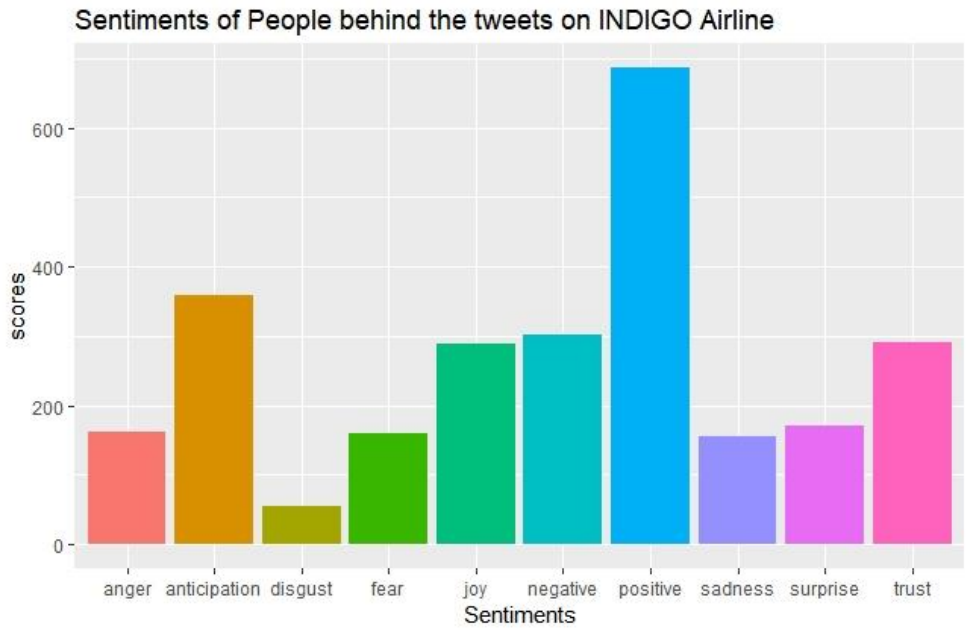


**Figure 2.** This figure shows the sentiment scores of Indigo Airline tweets.

The above graph is for the INDIGO airlines. This shows the sentiments of the people towards Indigo airlines. Most responses of the people are positive towards Indigo. Positive sentiment has a score of more than 400. It indicates that most of the people had a good experience and tweeted positively about the Indigo airlines. Anticipation, joy, and trust are the sentiments which also have a high score here. This indicates that the people are excited about Indigo, trust the airlines, enjoying their services and having a good experience with Indigo. People who had a negative response may be due to the services they are not happy with such as a delayed flight, lost luggage of customers etc. These responses can be converted into positive if these kinds of bad experiences are taken care of by the company. Customers trust Indigo as some assurance is given by the company of reaching the destination on time.

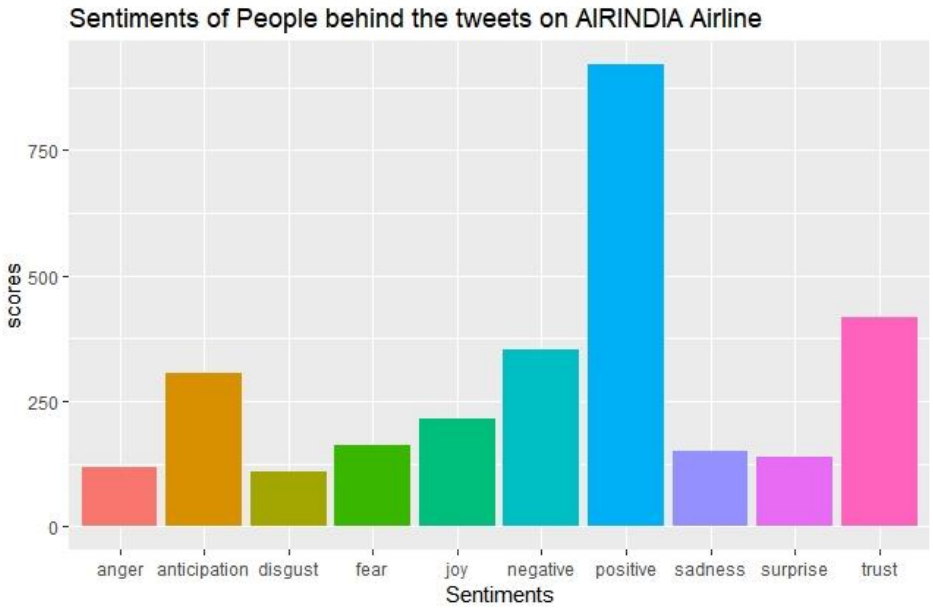


***Figure 3.*** This figure shows the sentiment scores of AirIndia Airline tweets.

The above graph of AIR India airlines shows that most responses are for the trust. People trust AIR India a lot. Next highest score is for the positive sentiment. People have had a positive experience with AIR India and hence it is building the trust for AIR India amongst the customers. Air India used a cross section of its actual users to endorse itself which helped this airline gain a lot of trust in their customers. Also, Air India's International reach and rich heritage got the airlines a good number of positive responses from the people. Negative sentiments and sadness are on a lower score on the graph. So, therefore AIR India is a trusted airline amongst the people.

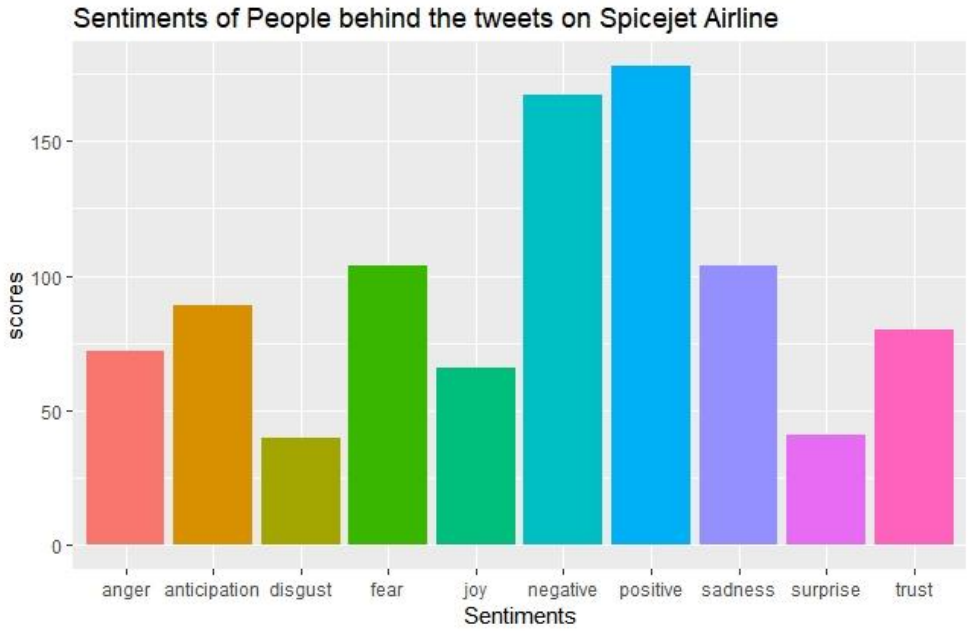


**Figure 4.** This figure shows the sentiment scores of SpiceJet Airline tweets.

This graph is for SpiceJet airlines. In this graph there are more responses for the negative sentiments. Positive responses are also at a good score but negative sentiments are on a greater score than positive sentiments. Fear and sadness also have a major contribution in the graph indicating that people don't trust SpiceJet and fear their service even more. Also, customer satisfaction seems to be at the bottom for this airline in recent times giving an increasing number of negative responses. Many customers have had a bad experience with SpiceJet service due to various reasons and this could affect the value and share of the company. Customers having bad experiences many times is not a good sign.

## 5. Conclusion

The paper has focused on using twitter data to analyse the sentiments of people on the Indian Airlines. Twitter sentiment analysis is developed to analyse customer's¬ perspectives towards the critical success factors in the marketplace. The program used here for research involves a machine-based learning approach which is more accurate with natural language processing techniques to analyse a sentiment.

The opinions of people on the airlines are obtained using a promising tool R programming for extraction and analysis purposes. The methodology proposed in the paper can be used for analysis of various other trending topics by applying sentiment analysis and hence serves the purpose of the research. Also, the results in the paper showcase a variety of emotions of customers which are expressed on twitter. The emotions such as fear, surprise, trust, disgust, anticipation and anger apart from positive and negative give very useful insights to the company on customer satisfaction and customer attitude towards their airlines.

Future work can be done in the area of analysing and differentiating positive and negative words from twitter as sentiment analysis has a major drawback of improper efficiency of classifying a tweet which can be a sarcastic opinion.

## References

[1] Changhyun Byun, Lee, H., Kim, Y., & Kim, K. K. (n.d.). Twitter data collecting tool with rule-based filtering and analysis module.

[2] D'Avanzo, E., Pilato, G., & Lytras, M. (2017). Using Twitter sentiment and emotions analysis of Google Trends for decisions making.

[3] K.Arun, Srinagesh, A., & Ramesh, M. (2017). Twitter Sentiment Analysis on Demonetization Tweets in India Using R. International Journal of Computer Engineering and Information Technology.

[4] Singh, A. K., & Gupta, D. K. (2017). Sentiment Analysis of Twitter User Data on Punjab Legislative Assembly Election. I.J. Modern Education and Computer Science.

[5] Li X, Li J, Wu Y (2015) A Global Optimization Approach to Multi-Polarity Sentiment Analysis.

[6] Kumar, A., & Sebastian, T. M. (2010). Sentiment Analysis: A Perspective on its Past, present and future. I.J. Intelligent and Application, 1-14

[7] Thelwall, M. (2017). Gender bias in sentiment analysis. Emerald Insights. https://doi.org/10.1109/ACCESS.2019.2

[8] Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pages 36–44.

[9] T. Wilson, J. Wiebe, and P. Hoffman. 2005. Recognizing contextual polarity in phrase level sentiment analysis. ACL.

[10] Hamid Bagheri, Md Johirul Islam, Computer Science Department Iowa State University, Sentiment analysis of twitter data.

[11] Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. P rocedia Computer Science, 17, 26-32.

[12] P.Lai, "Extracting Strong Sentiment Tendfrom Twitter". Stanford University, 2012.

[13] P. Nakov, Z. Kozareva, A. Ritter, S. Rosenthal, V. Stoyanov, T. Wilson, Sem Eval-Sentiment Analysis in Twitter (Vol, 2, pp. 312-320 ,2013).