Airline Sentiment Visualization, Consumer Loyalty Measurement and Prediction using Twitter Data

Rida Khan, Siddhaling Urolagin

Department of Computer Science
BITS Pilani, Dubai Campus, Academic City
Dubai, United Arab Emirates

Abstract—Social media today is an integral part of people's daily routines and the livelihood of some. As a result, it is abundant in user opinions. The analysis of brand specific opinions can inform companies on the level of satisfaction within consumers. This research focus is on analysis of tweets related to airlines based in four regions: Europe, India, Australia and America for consumer loyalty prediction. Sentiment Analysis is carried out using TextBlob analyzer. The tweets are used to calculate and graphically represent the positive, negative mean sentiment scores and a varying mean sentiment score over time for each airline. The terms with complaints and compliments are depicted using visualization methods. A novel method is proposed to measure consumer loyalty using the data gathered from Twitter. Furthermore, consumer loyalty prediction is performed using Twitter data. Three classifiers are employed, namely, Random Forest, Decision Tree and Logistic Regression. A maximum classification accuracy of 99.05% is observed for Random Forest on 10-fold cross validation.

Keywords—Consumer loyalty measurement; consumer loyalty prediction; sentimental visualization; airline consumer analysis

I. INTRODUCTION

People are increasingly using social media platforms as a place to review products and share product or service specific opinions [1]. As a result, businesses have become aware of the importance of social media as part of their marketing strategies. It can be used in communicating with consumers on a one-to-one basis and receiving immediate feedback [2]. The data from social media can be used to firstly understand the standing of a company within consumers and in the industry with respect to social media sentiment scores [3]. Secondly, the data can be used to infer consumer feedback, which is essential information for companies. The feedback informs the general public opinion about a company's product or service. People are constantly posting about brands they are using, their satisfaction with products and sharing these opinions with friends or family [4]. Hence, it is important for companies to know what the consumers are saying and whether it is positive or negative. Traditional forms of data collections could be replaced with analysis of social media platforms [5]. Data from surveys is commonly used to compute consumer loyalty measurements. We propose to use sentiment information to deduce loyalty measurement of a consumer towards a brand. The loyalty measurement will help to understand impact of consumer reviews on a brand [6].

Retaining customers and maintaining consumer loyalty is an essential marketing strategy, as customers are becoming increasingly important [7]. A loyal customer is an asset and companies today use a good amount of their marketing tools to maintain it. The currently available consumer loyalty measures use surveys and questionnaires but social media datasets are not used [8]. Since, social media platforms are full of user opinions pertaining to whether they are currently loyal or not and the reasons behind their position, social media datasets should be made an essential part of consumer loyalty measurement calculations [9], [10]. Discerning whether a consumer is loyal includes answering a set of questions, which could be in the form of a survey or questionnaire: How loyal is your customer? How likely is said customer to refer your brands to his friends or family? Is he likely to continue purchasing your products or services? Or is he looking for other potential options? Is he listening to pitches from your competitors? Is he willing to give you feedback and give you time to fix errors? Each of these questions is equally important in determining the loyalty of a certain consumer [11]. Along with these questions, there are many measures available that are used worldwide to calculate consumer loyalty measurement. These would include the Net Promoter Score (NPS) [12], Customer Loyalty Index (CLI), Upselling Ratio (UR) and Repurchase Ratio (RR) [13]. Companies use some or all of these measures in order to understand the current standing of consumers and to work towards retaining any or all of their consumers.

Our focus is on analyzing 'Tweets', which are Twitter [14] user posts of 280 characters or less. Along with the text of the tweet, other aspects are also collected like the username and number of people who have liked the tweet etcetera [15], [16]. Some categories of brands that are active and popular on Twitter include Airlines, Cars, Sports Teams, Entertainment, Finance, Retail and Food Industry among others [17]. This research work is focused on analysis of tweets pertaining to airlines. Our data consisting of tweets is collected using Tweepy [18]. We perform sentiment analysis [19] on tweets using TextBlob [20]. We infer the most common issues customers have with each airline and analyze the varying sentiments of the tweets using various visualization techniques [21]. We developed a new method to measure customer loyalty based on tweets. Furthermore, we carried out customer loyalty prediction based on sentiment data. In Section 2, we have provided the details about the dataset, in Section 3 we discussed the sentiment analysis of airline tweets. The experimental and graphical results are shown in Section 4 and Section 5 covers the conclusion.

II.   DATASETS

We have collected 'Tweets' for 18 airlines based in four selected regions, which are America, India, Europe and Australia. The specific airlines and tweet count are shown in Table I. Table II shows the sample tweets as an example.

TABLE I.   DATA COLLECTED

Airlines			
American	*Count*	*European*	*Count*
Delta	6000	**Lufthansa**	507
United	1600	**Air Berlin**	5822
Spirit	1600	**Turkish Air**	4637
Southwest	2600	**KLM**	2404
Jet Blue	3000	**Easy Jet**	1465
Indian	*Count*	*Australian*	*Count*
Indigo	1600	**Virgin**	3800
Air India	1080	**Qantas**	1600
Jet Airways	1100	**Tiger air**	5000
Vistara	1100	**Jet Star**	9994
Spice Jet	2392	**Sharp**	630

TABLE II.   SAMPLE DATASET

User name	Date	Tweets	Likes	Retweets
sashane_gordon,	2018-04-10 11:33,	Am sorry but I am really disappointed @JetBlue . I was a loyal customer. The way I advertise this airline you would think I work for the place but you guys showed me today that you don't care. Whether my family is dying or not as long as you have the money. Things happen everyday,	0	0
KW_Consultants,	2018-04-09 23:58,	@FlyAirNZ is the best airline I have ever flown! I am a loyal @AmericanAir customer , but the #skycouch stole my heart! https://twitter.com/flyairnz/status/983283384951754752	1	0
pberrini,	2018-02-28 02:18,	@WorldMark (Michael D Brown) I'm a corporate retiree and worked for organizations where I was empowered to make a difference for the customer as long as I followed 'core values'. I am a LOYAL Southwest Airline s customer for the same reason. I would love to feel same 4 Wyndham.	0	0
sevenpitches,	2018-02-25 04:04,	This is one of the many reasons I am a loyal Delta customer . Proud they are my hometown airline . Thanks, @Delta https://twitter.com/Delta/status/967391778897891328	3	0
ScottBurges,	2017-12-02 16:54,	Thought you were commenting on the on board staff ! Now they have become a budget airline , there are so many better quality and cheaper alternatives out there, and I was a loyal customer for over 20 years.	1	0

Each airline has a help desk handle; we have collected tweets that users have directed to these particular handles to create Dataset 1. In the case that an airline does not have such an account, we have used their official Twitter handle instead. Tweet collection has been carried out using Tweepy [18], which is a python library used for accessing Twitter API and is widely accessible to all Twitter users. Moreover, we formed Dataset 2 for consumer loyalty analysis by collecting tweets using the search queries "loyal flyer" and "loyal to airline" as well as "left airline" and the response of 10,000 tweets are gathered in CSV form. These 10,000 tweets are collected from 1048 users, out of which 524 have used the term "loyal to airline" or "loyal flyer" and 524 have used the term "left airline". Each collected tweet consists of the ID, permalink, date and time stamp, the text contained within the tweet, username, retweets and likes.

III.   ANALYSIS OF AIRLINE TWEETS

Twitter is an online platform that connects people through a social networking environment. Each user can create an account with a unique handle and post "Tweets" [22]. The main aspect of Twitter is text; however, pictures and videos are also shared. It should, therefore, be vital for companies to ensure any posts regarding their brand are positive. Millions of people discussing and mentioning a brand is only a good thing if the tweets are positive [23]. Sentiment analysis becomes important for identifying positive or negative tweets and determining the consumer voice. It is the use of natural language processing, as well as analysis of text and computational linguistics to study subjective data. It discerns whether a particular text or piece of writing – tweet, for example – is positive, negative or neutral. It uses context of the piece as well as tone, emotion and vocabulary. Sentiment analysis can aid companies in marketing strategies by understanding a general public opinion and in succession improving their customer service [24]. A company can determine the public opinion, analyze the customer satisfaction towards their products and be open to listening to the issues. This analysis can, not only help companies to know how their customers think, but can also aid them in competitor advantage. If companies are aware of the sentiments of competitors, it can aid them in comparison with their own sentiments and they can plan strategies to improve accordingly [25]. This analysis can also help in retaining customers as consumer loyalty can be determined and predicted which can greatly influence a company's organizational decisions [26].

A. Sentiment Information Visualisation

Our focus is on the opinions that users post on Twitter directed to airline Help Desks or to the official Twitter handles of various airline companies. These tweets range from compliments to complaints and issues that various consumers have with any airline. Analyzing the sentiment of tweets has an extra level of complication because the anatomy of tweets includes more textual aspects than an average written piece [27]. There are images, links, emoticons and other forms of media included. Hence, the first step of our analysis is to clean the tweets we have collected [28]. Tokenization is also difficult due to the body of the text. We would need to make sure that the @-mentions, emoticons, links and #hash-tags are preserved as individual tokens and not ignored, since these are equally important aspects of the analysis [27], [28]. In this research we follow the method as shown in Fig. 1. Airline tweets are first gathered from the Twitter API, which are then cleaned and tokenized. We then perform sentiment analysis on the Tweets, giving each tweet a score using TextBlob. Here, a score of 1

indicates most positive and -1 indicates most negative while zero means a neutral tweet or term. This analysis is carried out using a python library called TextBlob, which is used for processing textual data [20].

From each Tweet, its sentiment score TS_i is computed using TextBlob and the tweet is segregated into positive type or negative type as given in (1).

$$T_i = \begin{cases} \text{Positive if } TS_i > 0 \\ \text{Negative if } TS_i < 0 \\ \text{Neutral if } TS_i = 0 \end{cases} \tag{1}$$

where TS_i is the sentiment score of ith tweet.

We calculate a mean sentiment score, MS_i for each airline using the sentiment score of the airline tweets.

$$MS_i = \frac{1}{n} \sum_{i=1}^{n} TS_i \tag{2}$$

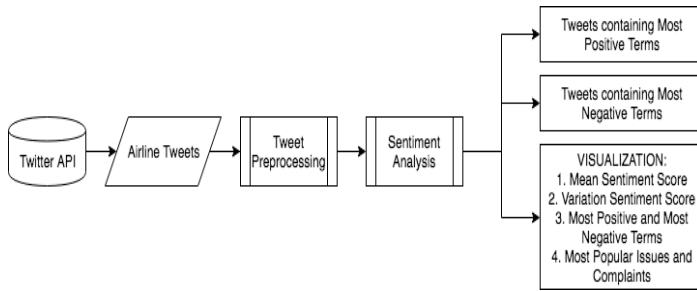where TS_i is the sentiment score of ith tweet and n is number of tweets for the airline.

Fig. 1. Tweet sentiment analysis.

B. Airline Passenger Loyalty Measurement

Consumer loyalty analysis is carried out using a second dataset, which consists of tweets containing phrases like "loyal flyer", "left airline" etcetera. The data downloaded with a tweet includes likes, which is the amount of people who have liked a tweet and retweets, which is the number of people who have shared a tweet. This can be used along with the number of followers a user has as well as the sentiment score of their tweets to calculate a loyalty measurement. We collect usernames of passengers who have explicitly stated they are loyal and those who say they are not loyal to an airline. Airline related tweets are gathered for each user. The tweets are subjected to TextBlob to compute the sentiment score, TS_i. The tweets related to a user are segregated to positive type and negative type. From positive tweets, the mean is computed as P^j and from negative tweets; the mean is computed as N^j for the jth person, which is given in (3), (4).

$$P^j = \frac{1}{n} \sum_{i=1}^{n} TS_i \tag{3}$$

where $TS_i > 0$ and n is number of positive tweets.

$$N^j = \frac{1}{n} \sum_{i=1}^{n} TS_i \tag{4}$$

where $TS_i < 0$ and n is number of negative tweets.

For the jth person, the mean of his Likes, L^j and Retweets, R^j are also calculated which make the influencer score In^j, and number of followers, F^j is gathered.

$$L^j = \frac{1}{n} \sum_{i=1}^{n} TL_i \tag{5}$$

where TL_i is number of likes of ith tweet and n is total number of likes for j^{th} person.

$$R^j = \frac{1}{n} \sum_{i=1}^{n} TR_i \tag{6}$$

where TR_i is number of retweets of ith tweet and n is total number of retweets for j^{th} person.

$$In^j = L^j + R^j \tag{7}$$

Hence, each user has positive, negative and influence score. Consumer loyalty measurement, CM^j is calculated as given in (8).

$$CM^j = \lfloor P^j + N^j \rfloor \times In^j \times F^j \tag{8}$$

Fig. 2 shows the method to compute consumer loyalty measurement. The tweets are queried with the search terms "loyal to airline", "loyal flyer", "left airline" and searched for usernames. The number of followers for each username is gathered along with the airline related tweets for each user. A sentiment score is computed for the tweets using TextBlob. The tweets are then segregated into positive and negative as given in (1). The positive, negative and influence scores are calculated for each user as given in (3), (4) and (7). The loyalty measurement is computed based on positive, negative, influence and follower scores as given in (8).
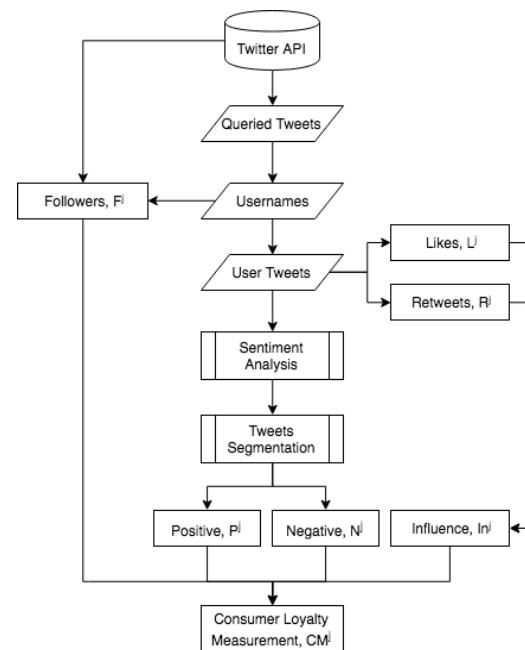

Fig. 2. Consumer loyalty measurement.

IV. EXPIREMENTAL RESULTS

In this research work, we have conducted sentiment analysis, customer loyalty measurement and loyalty prediction on tweets collected from airlines. We have dataset 1 and dataset 2 of tweets from various airlines from four regions, namely India, Europe, America and Australia. Dataset 1 consists of tweets collected from airline handles for region India: 6172 tweets, for region Europe: 14835 tweets, for region America: 13200 tweets and for Australian airlines: 21024 tweets. Searching terms "loyal flyer", "loyal to airline" and "left airline" forms Dataset 2 and total 10000 tweets are gathered. We use the data to calculate mean sentiment scores for each airline. The airlines can use these depictions to understand areas of improvement, successful strategies and can utilize these insights in retaining customers.

A. Tweet Sentiment Visualisation

Sentiment analysis is performed on the tweets from dataset 1 using TextBlob. The tweets are then separated into "Positive tweets" and "Negative tweets". The mean sentiment score for positive and negative scores is calculated for each airline. We also compute sentiment score over time to depict the variation for selected airlines. From the gathered tweets, we also search for the most frequently occurring positive and negative terms along with the corresponding tweets. Fig. 3(a) graphically represents the positive and negative mean sentiment scores for five Australian airlines. The mean sentiment score is vital for an airline to understand a general consumer opinion about their services at a point in time. We have observed variations in positive and negative sentiments for the various airlines. An airline would want to make sure that their positive sentiment score is greater than their negative sentiment scores. These scores also aid in competitor advantage as an airline can work towards making their positive scores greater and negative

scores lesser than their competing airlines. Fig. 3(b), 3(c) and 3(d) show similar results for American, Indian and European airlines, respectively.

The brand sentiment scores over time are important for companies, since they indicate whether customers have been talking about your brand positively and their attitude towards your brand is improving or whether they have been dissatisfied and the score has been reducing. The tweets used to calculate the airline variation score range over a month's time. The mean sentiment score per week is calculated for each airline. Fig. 4 graphically represents airline variation scores that show the increases or decreases in consumer satisfaction over a time period. These variations can be studied in order to understand the consumer's satisfaction over a period of time. Fig. 4(a) depicts Jet Airways variation score that starts at 0.15 in Week 1 but decreases down to 0.025 by Week 3.

The most common negative and positive terms for all airlines are counted and a list is made. The negative terms have the highest score of -1.0 and are worst, awful, pathetic, disgusting, terrible and horrible. The positive terms have the highest score of 1.0 and are awesome, excellent, delicious, perfect, superb and wonderful. Fig. 5(a) depicts the most frequently occurring negative terms and their respective frequency. An example of a frequent word with a frequency of 150 is 'worst'. The tweets with this term can be indicators of areas that need improvement for the respective airline. Fig. 5(b) shows similar results for positive terms. The tweets with positive terms can be indicators of areas that are incurring a positive sentiment and the work in these areas can be maintained in order to keep or improve positive scores and consumer satisfaction. Some of the tweets where these terms occur are shown in Table III with their sentiment score.
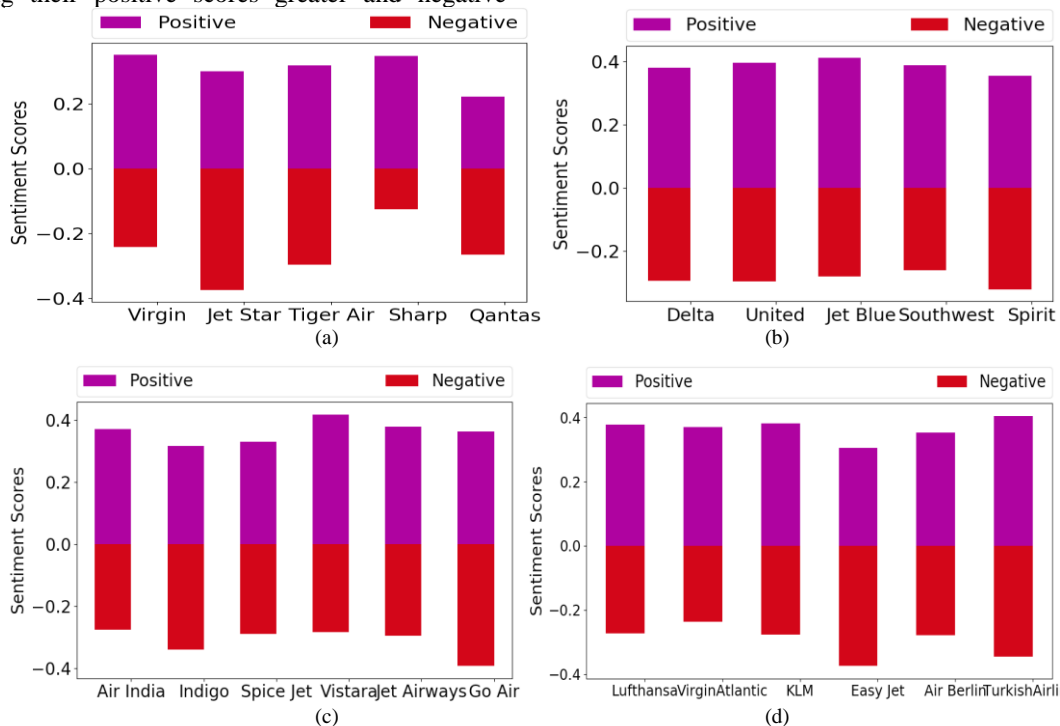

Fig. 3. Mean sentiment score: (a) Australia (b) America (c) India (d) Europe.

For all airlines, we have represented the most common bigrams in the form of a pie chart in Fig. 6 where each term and its respective frequency are depicted. These bigrams are the two most common terms that occur together most frequently in negative and positive tweets. These terms show the most common problems and the most common praises. Fig. 6(a) shows the most common positive feedback within positive tweets and their respective percentage in terms of frequency within the dataset. The most frequently occurring praise is "customer service" indicating that successful customer service incurs positive sentiment scores. Fig. 6(b) depicts the most common issues found within negative tweets.

TABLE III. TWEETS CORRESPONDING TO POSITIVE AND NEGATIVE TERMS

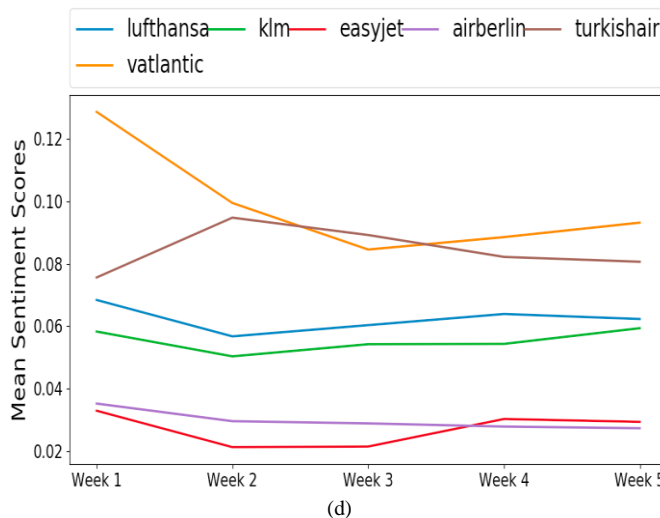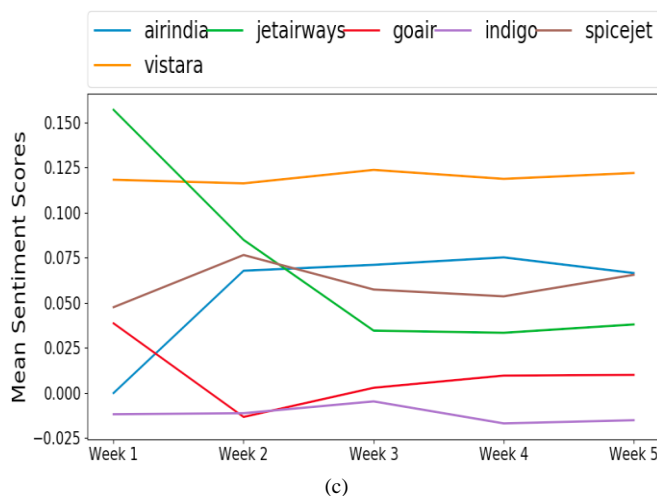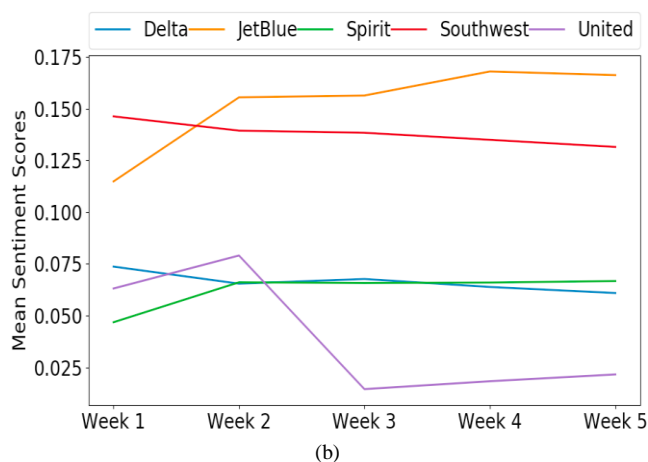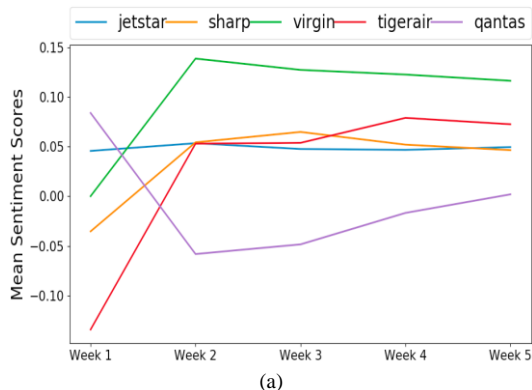| Term	Tweet	Score
Negative	is the **worst airline** ever My flight been scheduled to leave since 11 am it 3 pm now smh	-1.0
Negative	has no respect for its customers and zero professional pride Terrible service The entire flight is still **waiting for luggage pathetic** You're losing a customer	-0.475
Negative	And I get to sit in yet another **disgusting seat** No thanks I will never recommend your carrier	-0.55
Negative	My 1st **terrible airline flight** personal items going missing in the plane and cabin crew super unhelpful refused to help find my passport money housekeys shrugging shoulders is not what I expected Conspiring with ground staff disappointed	-0.286115
Positive	Emergency landing in St John s Canada pilot made a **perfect landing** Passenger taken to hospital and is reported to be in good hands bei St John s International Airport YYT	0.56666666
Positive	Such super **delicious food** and awesome hospitality at 38000 ft Just loved it You will surely rule the sky one day Hope to fly frequently with you Thank You	0.55416666
Positive	Flew with for the first time today **superb experience** Great service and hospitality on board Fantastic leg space too Kudos	0.61249999
Positive	Had a **wonderful flight** Never expected such a delicious snack more leg space in domestic flights UK899 See you soon	0.425

(a)

(b)

(c)

(d)

Fig. 4. Variation sentiment score: (a) Australia, (b) America, (c) Europe, (d) India.
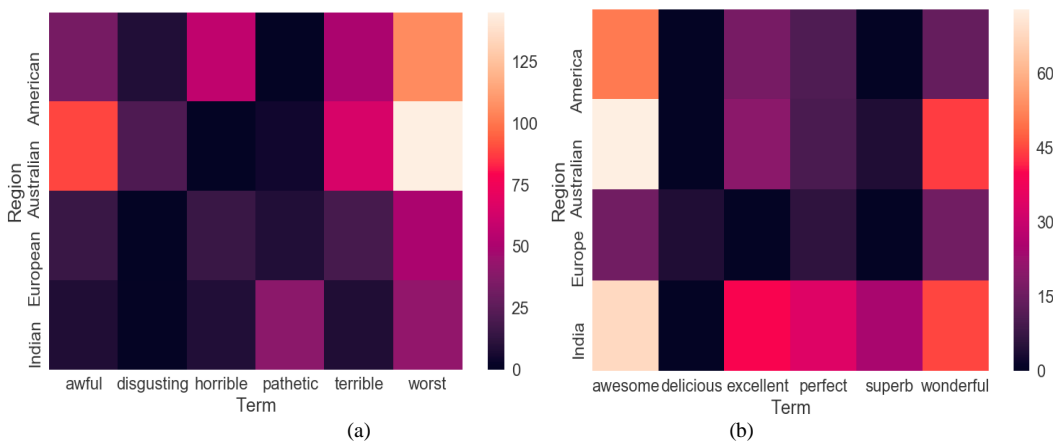
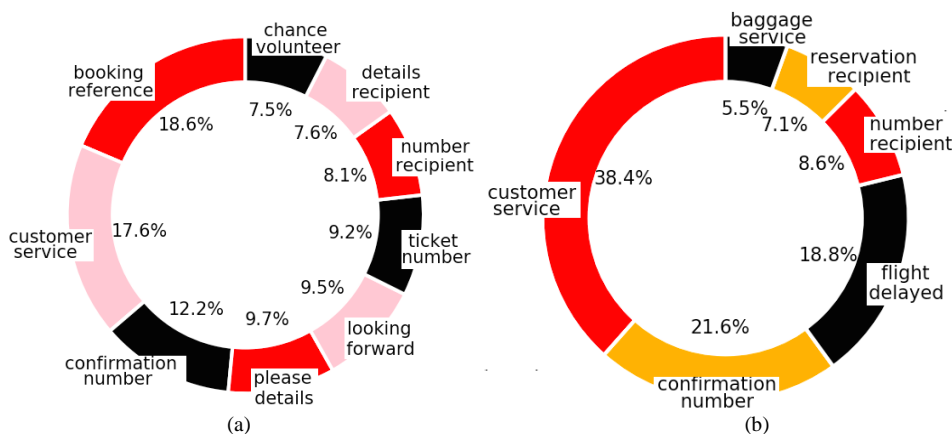Fig. 5.   Most common terms: (a) Negative terms, (b) Positive terms.

Fig. 6.   (a) Common praises,(b) Common issues.

B. Airline Passenger Loyalty

We gathered tweets using search queries such as "loyal flyer", "loyal to airline" as well as "left airline". There are 1048 users and 10000 tweets in the dataset 2. From the users, 524 have explicitly said they are loyal to an airline and the other 524 have said they are not loyal or have left an airline. The positive, negative user scores are calculated using (3), (4) along with mean likes and retweets using (5), (6) as described in section 3.2. These values are used to calculate the consumer loyalty measurement as given in section 3.2 using (8). The normalized loyalty measurements are depicted in Fig. 7. The normalization is performed by dividing the difference between maximum and minimum loyalty score.

Fig. 7 represents the consumer loyalty measurements for 524 loyal passengers and 524 disloyal passengers. The passengers who have used the term "left airline" have a loyalty measurement varying between 15 and 21. The passengers who have used the terms "loyal to airline" or "loyal flyer" have a loyalty measurement varying between 250 and 300. These measurements can be used to cluster consumers as loyal or not loyal based on their Twitter data. We used K-Means clustering [29] which is an unsupervised learning algorithm with the number of clusters set to two. The various values used to calculate the loyalty measurement are graphically represented in different combinations using the k-means clusters to depict the loyal and not loyal passengers.
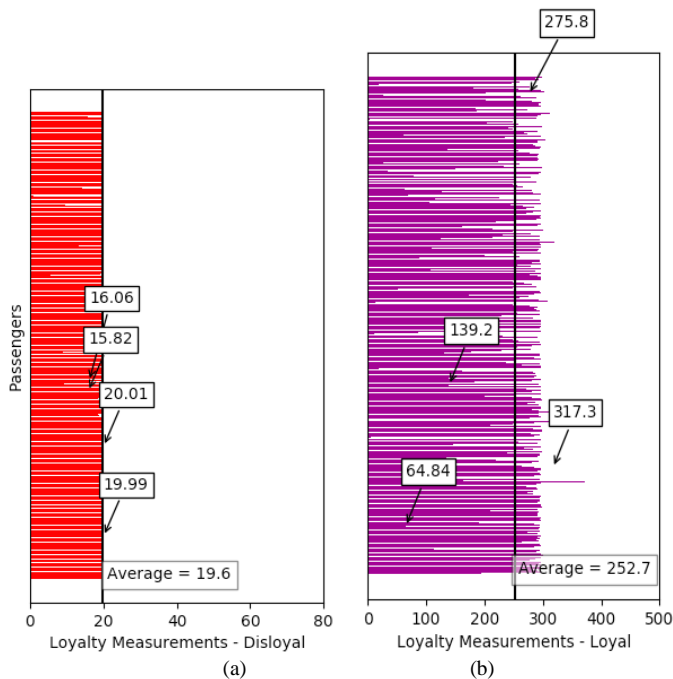

Fig. 7.   Airline passenger loyalty measurements: (a) Disloyal passengers, (b) Loyal passengers.
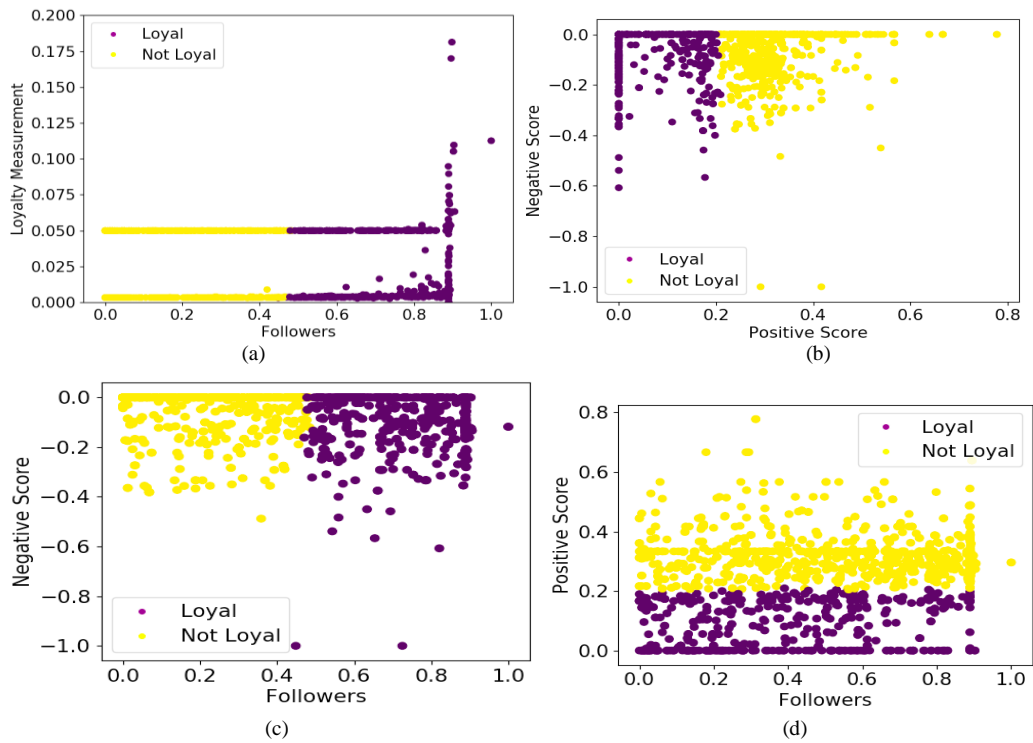
Fig. 8. Clustered passengers (a) Loyalty measurement vs followers, (b) Negative score vs positive score, (c) Negative score vs followers, (d) Positive score vs followers.
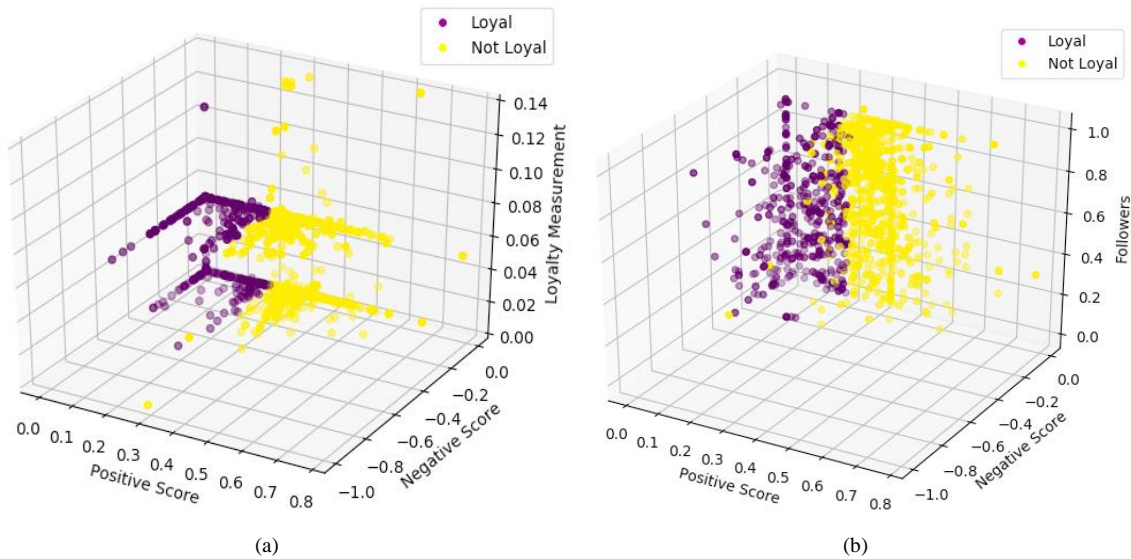

Fig. 9. 3D visualizations: (a) Positive score vs negative score vs loyalty measurement, (b) Positive score vs negative score vs followers.

Fig. 8(a) depicts the followers of each passenger in comparison to their normalized loyalty measurement. Fig. 8(a) depicts loyalty measurements along the y-axis and number of followers along the x-axis. This informs an airline about each loyal or disloyal passenger and their influence. Fig. 8(b) depicts the negative score on the y-axis and positive score on the x-axis. Each point represents a user who is either loyal or not loyal. This represents positive versus negative scores with respect to consumer loyalty. Fig. 8(c) and 8(d) represent the negative, positive scores on the y-axis versus the number of followers on the x-axis. Both these figures aid airlines in understanding the influence of users versus their positive and negative scores.

We depict the loyalty measurement and few terms used in calculating this measurement in Fig. 9. Fig. 9(a) depicts positive score on the x-axis, negative score on the y-axis and normalized loyalty measurement on the z-axis. Airlines can understand whether their loyal or disloyal passengers have a positive or negative attitude towards their services at a point in time. Fig. 9(b) depicts positive score on the x-axis, negative

score on the y-axis and the number of followers on the z-axis. Both these figures show a clear distinction between the loyal and disloyal passengers. Fig. 9(b) can help inform airlines of the strength of people each loyal and disloyal could influence. The 3D pictorial graphs can be used in vital analysis by airline marketing teams to understand where each passengers stands. Also, these 3D graphs represent the influence of each passenger, negative, positive scores with respect to loyalty.

Next, we carried out consumer loyalty prediction. Previous works that have been carried out for consumer loyalty prediction include surveys (with questionnaires) [30], [31] and airport reports [32] as datasets. Social media is used as a data set but to predict an existing consumer loyalty measure, which is NPS [33]. This recent work can be seen in Table IV.

For consumer loyalty prediction, we used three prediction models, which are Random Forest [34], Decision Trees [35] and Logistic Regression [36] on dataset 2. The model is fitted using tweet related information such as positive sentiment score, negative sentiment score, mean of retweets, mean of likes and the number of followers. Two-class prediction is performed as either loyal or not loyal. The models are tested on 10-fold cross validation [37] and the accuracies are given in Table V. The maximum accuracy of 99.05% is observed for Random Forest.

TABLE IV. RECENT WORK IN CONSUMER LOYALTY PREDICTION

Author	Year	Title	Data sets	Accuracy
Mohamed Zaki, Dalia Kandeil, Andy Neely, Janet R. McColl-Kennedy	2016, [30]	The Fallacy of the Net Promoter Score: Customer Loyalty Predictive Model	Surveys	98.00%
Jose Berengueres, Dmitry Efimov	2014, [31]	Airline new customer tier level forecasting for real-time resource allocation of a miles program	Surveys	87.00%
S.T.M. van Velthoven	2014, [33]	Sentiment analysis on social media to predict Net Promoter Score	Social Media	85.67%
Hari Bhaskar Sankaranarayanan, B V Vishwanath, Viral Rathod	2017, [32]	An exploratory analysis for predicting passenger satisfaction at global hub airports using logistic model trees	Airport Reports	80.00%

TABLE V. CONSUMER LOYALTY PREDICTION

Model	Accuracy	Misclassification
Random Forest	99.05	0.95
Decision Tree	98.97	1.03
Logistic Regression	91.10	8.90

V. CONCLUSION AND FUTURE WORK

In recent years, the tremendous growth of social media is impacting various sectors including businesses. It is vital for any brand today to have a presence on the Internet, one that is memorable for the consumers. In this research, data from social media such as Twitter is gathered for airline industry. We collected airline tweets from four regions namely India, Europe, Australia, America and performed sentiment analysis.

We identified the compliments and complaints of customers, variations in sentiment over a period of time and depicted mean sentiments scores using visualization techniques. This analysis provides a general opinion of passengers towards airlines and its variation over time.

Furthermore, we searched tweets with the terms such as "loyal to airline", "loyal flyer" and "left airline" and collected 10000 tweets. Consumer loyalty analysis is performed on these tweets. We proposed a new method to measure consumer loyalty based on Twitter information such as positive, negative sentiment scores, mean likes, mean retweets and number of followers. Then, consumer loyalty prediction is performed using three classifiers, which are Random Forest, Decision Tree, and Logistic Regression. These classifiers are trained using features collected from Twitter information on 10,000 tweets. All three classifiers are tested using 10-fold cross validation and classification accuracies are collected. A maximum accuracy of 99.05% is observed for Random Forest classifier on 10-fold cross validation. The consumer loyalty analysis helps airline companies to retain consumers and bring in new loyal customers. Moreover, consumer loyalty measure and prediction can be performed for different business sectors.

REFERENCES

[1] M. Lovelin, P. Felciah, R. Anbuselvi, "A study on sentiment analysis of social media reviews", 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-3, Coimbatore, 2015.

[2] P. Samiei, A. K. Tripathi, "Effect of Social Networks on Online Reviews", 2014 47th Hawaii International Conference on System Sciences, Waikoloa, HI, 2014, pp. 1444-1453.

[3] F. Zhu and X. Zhang, "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics", Journal of Marketing, Vol. 74, no. 2, pp. 133-148, 2010.

[4] Amrita, S.M., Mohan, R., "Application of social media as a marketing promotion tool — A review" in Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on 15-17 Dec. 2016, Chennai, India

[5] Jamilah, Putu Wuri Handayani, "Analysis on effects of brand community on brand loyalty in the social media: A case study of an online transportation (UBER)", in International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2016.

[6] Munnukka J., Karjaluoto H., Mahlamäki T., Hokkanen V. "Effects of Social Media on Consumers' Sports Brand Experiences and Loyalty." In: Stieler M. (eds) Creating Marketing Magic and Innovative Future Marketing Trends. Developments in Marketing Science: Proceedings of the Academy of Marketing Science. Springer, Cham

[7] J. b. Shao, Z. z. Li and M. y. Hu, "The impact of online reviews on consumers' purchase decisions in online shopping", 2014 International Conference on Management Science & Engineering 21th Annual Conference Proceedings, Helsinki, 2014, pp. 287-293.

[8] Jansen, Bernard J., M. Zhang, K. Sobel, A. Chowdury, "Twitter power: Tweets as electronic word of mouth", Journal of the American society for information science and technology 60, no. 11, 2009, Pages 2169-2188.

[9] Baumöl, U., Hollebeek, L. & Jung, R. "Dynamics of Customer Interaction on social media platforms." Electron Markets (2016) August 2016, Volume 26, Issue 3, pp 199–202

[10] R. Dowling Grahame, Uncles Mark, "Do Customer Loyalty Programs Really Work? Sloan Management Review", vol. 38, no. 4, pp. 71-82, Sep. 1997

[11] Fan, Yingren, "Research on improving the customer loyalty of retail enterprises, in Artificial Intelligence", Management Science and

Electronic Commerce (AIMSEC), 2011 2nd International Conference on 8-10 Aug. 2011

[12] Jacob K. Eskildsen , Kai Kristensen, "The accuracy of the Net Promoter Score under different distributional assumptions" in Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2011 International Conference on 17-19 June 2011

[13] Choi Sang Long, Raha Khalafinezhad, "Customer Satisfaction and Loyalty: A Literature Review" in the Perspective of Customer Relationship Management. J. Appl. Bus. Fin. Res. 1 (1): 06-13.

[14] Twitter (2017) Twitter Developer Documentation. [online] Twitter Developer Documentation. Available at: https://dev.twitter.com/rest/public/rate-limiting [Accessed 10th March 2018].

[15] Neethu, M. S., Rajasree, R, "Sentiment analysis in twitter using machine learning techniques", in Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on 4-6 July 2013, Tiruchengode, India, 2013.

[16] Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., Badhani, P., "Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python" in International Journal of Computer Applications, Volume 165 – No.9 on May 2017

[17] Kerns C. "Understanding Brands on Twitter." In: Trendology. Palgrave Macmillan, New York, 2014, Pages 39-85.

[18] Tweepy (2017) Streaming With Tweepy — tweepy 3.5.0 documentation. [online] Tweepy.readthedocs.io. Available at: http://tweepy.readthedocs.io/en/v3.5.0/streaming_how_to.html [Accessed 10th March 2018].

[19] Mäntylä, M. V., Graziotin, D., Kuutila, M., "The Evolution of Sentiment Analysis in Computer Science Review", Volume 27, February 2018, Pages 16-32

[20] Loria, S., Keen, P., Honnibal, M., Yankovsky, R.; Karesh, D., Dempsey, E. Textblob: Simplified Text Processing. Available online: https://textblob.readthedocs.org/en/dev/ (accessed on 10th Match 2018).

[21] Bird, S., Klein, E., Loper, E. "Natural Language Processing with Python", O'Reilly Media, 2009.

[22] Sarlan, A., Nadam, C. and S. Basri. "Twitter sentiment analysis" in Information Technology and Multimedia (ICIMU), International Conference on 18-20 Nov. 2014, Putrajaya, Malaysia, 2014.

[23] Jo Mackiewicz, Dave Yeats, Thomas Thornton, "The Impact of Review Environment on Review Credibilit", Professional Communication IEEE Transactions on, vol. 59, pp. 71-88, 2016, ISSN 0361-1434.

[24] Chen LS., Yang TY.K. "Increasing Customer Loyalty in Internet Marketing." In: Pan JS., Snasel V., Corchado E., Abraham A., Wang SL. (eds) Intelligent Data analysis and its Applications, Volume II. Advances in Intelligent Systems and Computing, vol 298. Springer, Cham.

[25] Isah, H., Trundle, P., Neagu, D. "Social media analysis for product safety using text mining and sentiment analysis" in Computational Intelligence (UKCI), 2014 14th UK Workshop on 8-10 Sept. 2014, Bradford, UK.

[26] Cuthbertson, R. & Laine, A. J Target Meas Anal Mark "The Role of CRM within retail loyalty marketing." October 2003, Volume 12, Issue 3, pp 290–304.

[27] Bao Y., Quan C., Wang L., Ren F. "The Role of Pre-processing in Twitter Sentiment Analysis." In: Huang DS., Jo KH., Wang L. (eds) Intelligent Computing Methodologies. ICIC 2014. Lecture Notes in Computer Science, vol 8589. Springer, Cham.

[28] Hu, G., Bhargava, P., Fuhrmann, F., Ellinger, S., Spasojevic, N. "Analyzing Users' Sentiment Towards Popular Consumer Industries and Brands on Twitter" in Data Mining Workshops (ICDMW), 2017 IEEE International Conference on 18-21 Nov. 2017, New Orleans, LA, USA.

[29] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881-892, Jul 2002.

[30] M. Zaki, D. Kandeil, A. Neely, J. R. McColl-Kennedy, "The Fallacy of the Net Promoter Score: Customer Loyalty Predictive Model", Cambridge Service Alliance 2016, October 2016.

[31] Berengueres, J. & Efimov, "Airline new customer tier level forecasting for real-time resource allocation of a miles program", D. Journal of Big Data, December, vol. 1, issue 3, 2014.

[32] H. B. Sankaranarayanan, B. V. Vishwanath and V. Rathod, "An exploratory analysis for predicting passenger satisfaction at global hub airports using logistic model trees", 2016 Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, 2016, pp. 285-290.

[33] S.T.M. van Velthoven, "Sentiment analysis on social media to predict Net Promoter Score", in partial fulfillment of the requirements for the degree of Master of Science in Operations Management and Logistics Eindhoven, October 2014.

[34] M. S. Alam and S. T. Vuong, "Random Forest Classification for Detecting Android Malware", 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, Beijing, 2013, pp. 663-669.

[35] Linna Li and Xuemin Zhang, "Study of data mining algorithm based on decision tree", 2010 International Conference On Computer Design and Applications, Qinhuangdao, 2010, pp. V1-155-V1-158, 2010.

[36] S. T. Indra, L. Wikarsa and R. Turang, "Using logistic regression method to classify tweets into the selected topics", 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Malang, 2016, pp. 385-390.

[37] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification", 2016 IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, 2016, pp. 78-83, 2016.