International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-04, Issue-02, May 2018

An Exploration of Airline Sentimental Tweets with Different Classification Model

M.Vadivukarassi[1], N.Puviarasan[2], P.Aruna[3]

[1,3]Department of Computer Science and Engineering, Annamalai University

[2]Department of Computer and Information Science, Annamalai University

[1]vadivume28@gmail.com, [2]npuvi2410@yahoo.in, [3]arunapuvi@yahoo.co.in

Abstract- In Twitter, the customer of airline services can tweet their opinions about their travelled experiences in flight. So Twitter contains massive amount of data and information regarding airline services. These tweets are collected and explored the sentiments about the airline services to track customer satisfaction reports and to discover location of the customer. The main aim of this paper is to analyze the twitter airline dataset for finding the best and the worst airlines and also to predict the most common issues occurred during the airline services. Then the word clouds of negative tweets are created and also the location of the negatively tweeted customer is predicted and visualized using geographical analysis. Finally training and testing was done on the dataset and also compared with seven different classifiers such as Logistic Regression classifier, KNeighbors classifier, SVC, Decision Tree classifier, Random Forest classifier, AdaBoost classifier and GaussianNB. The results of our experiments demonstrate that the Random forest approach works best in real-world practice on sentiment classification of tweet data.

Keywords —airline services, classifier, geographical analysis, twitter, word cloud.

I. INTRODUCTION

The internet provides the vast amount of information about almost everything. Individuals have perceived the internet as an important source in which a large number of opinions and experiences are readily available. People's evaluations significantly influence others beliefs, perceptions, and particularly their buying decisions. Nowadays, the information flow has gradually transformed to an online-based aggregation of experiences, insights, and views. The steep increase of online data creates a significant opportunity for companies to understand better what customers are saying about a product, topic, or other entity [1]. In recent years, Twitter Sentiment analysis has been getting very popular for automatic customer satisfaction analysis of online services. The customer's feedback about their services for airline companies is very essential.

In this paper, Twitter airline dataset are analyzed and most common problems occurs during services are predicted and location of the predicted tweets are visualized. Then several sentiment classification algorithms including LogisticRegression, KNeighbors, SVC, DecisionTree, RandomForest, AdaBoost and GaussianNB are tested and compared. In the test results, the best sentiment classification algorithms for airline services companies are selected based on good accuracy. In the next section, some related works on sentiment analysis and machine learning classifier are discussed. Then the proposed system is described and classification approaches are explained. The experiment section describes the experimental results from different sentiment classification algorithms tested on airline services datasets. Finally, the best sentiment analysis algorithm for airline services is presented and several directions of future work are also suggested.

II. RELATED WORKS

Sreenivasan et al studied tweets from three airline brands such as Malaysia Airlines, Jet Blue Airlines and Southwest Airlines and they used twitter as the data to analyze consumers' communications about airline services [2][9]. Breen et al illustrates classifying tweets sentiment by applying sentimental lexicons and suggests retrieving real time tweets from Twitter API with queries containing airline companies' names. In this study, there is no data training process or testing process. But in our work, this method was applied and tested with pre-labeled data. It yielded inaccurate testing results because sentiment classifications are highly domain specific [3]. Adeborna et al adopted Naive Bayesian method in sentiment detection process by comparing with SVM and Entropy. The result of this case study reached 86.4% accuracy in subjectivity classification and displayed specific topics describing the nature of the sentiment. In this research, the author only used unigrams as sentiment classification features in Naive Bayes algorithm, which can cause problems because phrases and negation terms can change sentiment orientation of those terms in sentences [4]. In my work, seven classifier are compared and random forest classifier

 DOI : 10.18231/2454-9150.2018.0124

yields higher accuracy. A discussion of the relationship between sentiment classification and airline service domain studied. The author Baker collected the data from the Department of Transportation's Air Travel Consumer Report on the measures such as percentage of on-time arrival; passengers denied boarding, mishandled baggage and customer complaints which are similar to the airline service [5].

III. CLASSIFICATION APPROACHES

The classification and prediction are two type of data exploration that can be used to extract models describing important data classes or to predict future data trends. Classification is a machine learning technique used to predict group membership for data instances. Machine learning refers to a system that has the capability to automatically learn knowledge from experience and other ways. This classification approaches predicts categorical labels whereas prediction models continuous valued functions and also it is one of the task of generalizing known structure to apply to new data. The different classification approaches are implemented for the Twitter Airline dataset to find the sentiment of the tweets. They are described as below.

A. *Logistic Regression*

The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest and a set of independent variables [6]. It is a statistical method for analyzing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable.

B. *K nearest neighbors*

It is a simple algorithm that stores all available cases and classifies new cases by a majority vote of its k neighbors. The case being assigned to the class is most common amongst its K nearest neighbors measured by a distance function. These distance functions can be Euclidean, Manhattan, Minkowski and Hamming distance. First three functions are used for continuous function and fourth one for categorical variables. If K = 1, then the case is simply assigned to the class of its nearest neighbor. At times, choosing K turns out to be a challenge while performing KNN modeling [6].

C. *Support Vector Machine*

Support vector machine classifiers are supervised machine learning models used for binary classification and regression analysis. However, in our work, we aim to build classifiers, which can classify tweets into three sentiment categories.. Based on the study done by Hsu and Lin, the pairwise classification method outperforms the one-against-all classification method in multiclass support vector machine classification. In the pairwise classification method, each pair of classes will have one SVM classifier trained to separate the classes. The accuracy of the classification will be the overall accuracy of every SVM classification included[6].

D. Decision Tree

It is a flowchart-like tree structure, in which each internal node represents a test on an attribute and each branch represents an outcome of the test, and each leaf node represents a class. Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. The topmost decision node in a tree which corresponds to the best predictor called root node[11].

E. Random forests

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set[11].

F. AdaBoost

AdaBoost classifier is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases [10].

G. Gaussian NB

Gaussian NB classifier is a classification technique based on Bayes Theorem with an assumption of independence among predictors. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods. When dealing with Gaussian Naïve Bayes classifier, the outcome model will have a high-performance with high training speed with the capabilities to predict the probability of the feature that belongs to Zk class[8].

IV. PROPOSED SYSTEM

In this paper, tweets about people's flight experiences are gathered from Twitter and airline companies can make suggestion about the people comments based their trip. Fig 1 shows the block diagram of the proposed system. The dataset contains about 14,640 tweets which are collected on February 2015 from various airline reviews. Every review is labelled as positive, negative or neutral. First, a reason to each negative response were analysed as late flight, lost luggage, etc and also seven different classifier model are

 DOI : 10.18231/2454-9150.2018.0124

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-04, Issue-02, May 2018

build to perform sentiment analysis on the data set. In our data set, about 80% of the negative reviews have a negative reason label, yet the rests are labelled as”can’t tell”. By knowing every review’s negative reason, we can give specific suggestions to different airline companies on how to improve their service.

Fig.1. Block Diagram of the proposed system

In existing work, the dataset was analyzed and predict only the negative tweets and visualized the negative tweets using word cloud. So it displays the frequently tweeted words from the user about the airline service. But in this paper, the proposed work is to collect the negative reviews separately and form a word cloud and also the locations of the negatively tweeted user are visualized using Google map. This visualization displays the location of the user easily who is negatively tweeted. Then seven different classification methods are implemented for training and testing. They are Logistic Regression classifier, KNeighbors classifier, SVC, Decision Tree classifier, Random Forest classifier, AdaBoost classifier and GaussianNB. These classifiers were called as Supervised Learning algorithm. The training process continues until the model achieves a desired level of accuracy on the training data. For most of the sentiment classification research, the accuracy of the classification result is evaluated by calculating the ratio between correctly classified tweet and incorrectly classified tweets. The calculation of precision, recall and f measure are known to be common accuracy evaluation method for sentiment classification result.

V. RESULTS AND DISCUSSIONS

In this proposed work, the dataset contains various tweets based on different airline company services. The “Twitter Airline Sentiment” dataset was downloaded from Kaggle which contains tweets covering six U.S. airline companies with a total number of (14,640) tweets, each of which is labeled according to sentiment polarity as: positive, negative, and neutral.

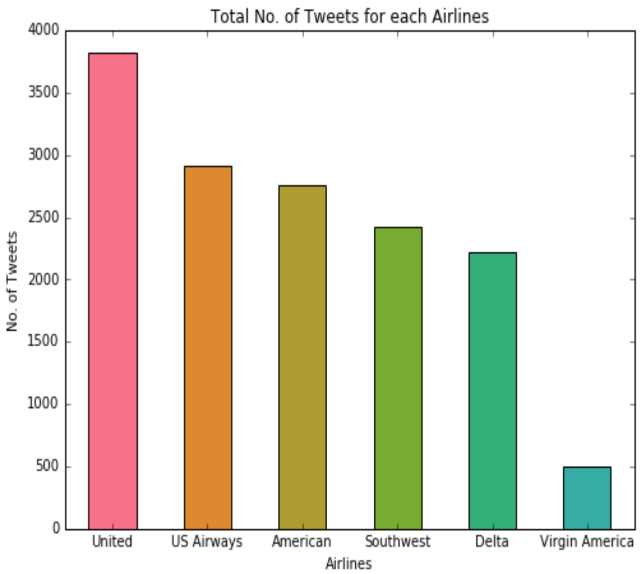

Fig 2. Total number of tweets collected for each airline

The six U.S airline companies and the total number tweets collected on each airlines are United (3822), US airways (2913), American (2759), Southwest (2420), Delta (2222) and Virgin America(504) respectively. Then these tweets are analysed and predict the top most negative reasons (such as "late flight" or "rude service"). Before training machine-learning models on the data, some exploratory data analysis was conducted on the dataset to get a better analysis. Fig 2 represents the total number of tweets collected for each airline. Fig 3(a) shows the total number of tweets for each sentiment. The numbers of negative, neutral and positive tweets are 9178, 3099 and 2363 respectively. Fig 3(b) represents the pie-chart representation of airline sentiment. It shows 62.69% of the tweets contain negative comments, 21.17% of the tweets contain neutral comments and 16.14% of the tweets contain positive comments from the customer services. It shows that customers had tweeted more negative comments and it is further analysed to predict the issues on the airline services. It is also investigated how the sentiments vary across airlines.

 DOI : 10.18231/2454-9150.2018.0124

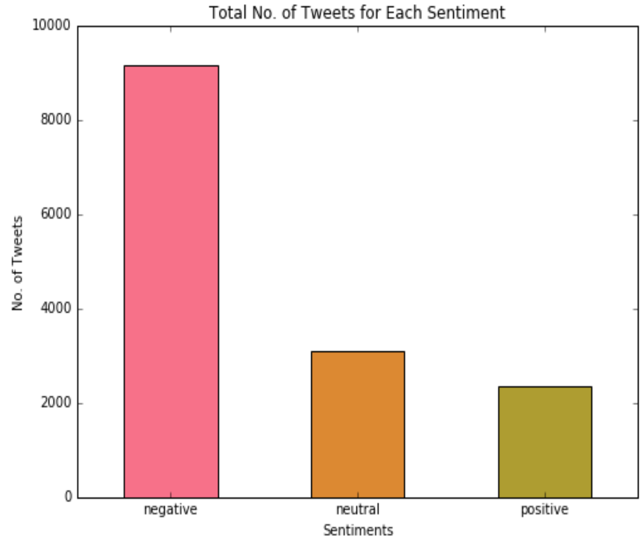

Fig.3. (a) Total number of tweets for each sentiment

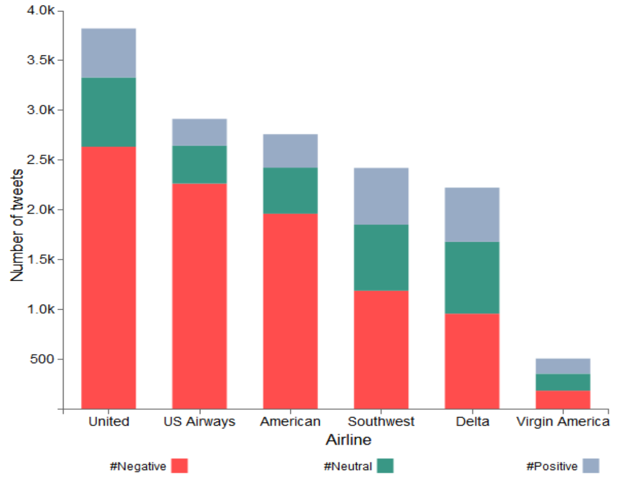

Fig.4. Sentiments of different airlines

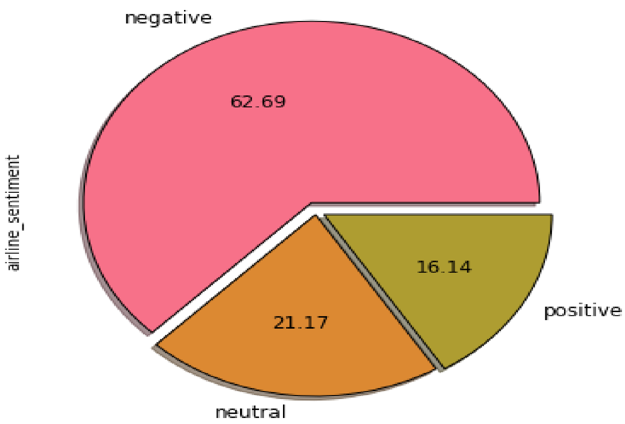

Fig.3. (b) Pie-chart representation of airline sentiment

Fig.4 represents how sentiments vary across different airlines services. United had the most tweets with negative sentiment and it has the maximum number of tweets. The number of tweets about an airline may be correlated to the number of planes the airline operates. Then normalize the number of individual sentiments by the total number of tweets to make relative comparisons. Fig. 5 shows the reason for negative comment reported in the tweets. The excluded data where the reason was not specified or reason was given as 'can't tell' are reported more in the dataset. This reduced the number of negative comments of the airline services. This plot shows that the most common reason for negative sentiment was customer service issue, followed by late fight and cancelled flights.

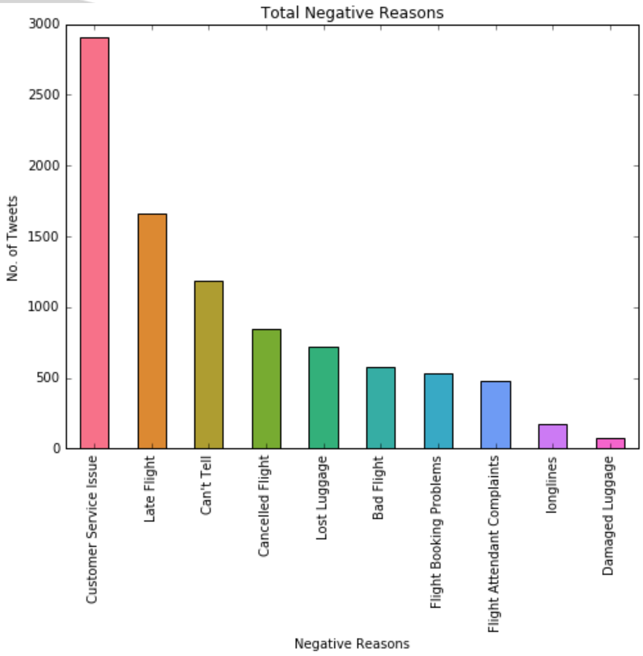

Fig 5. Top 10 negative reasons of the airline services

Fig 6 represents the pie chart of top 5 negative reasons. It shows that customer service issue has occurred 39.63% compared to other issues. Fig 7 grouped the reason for negative comments by airline, and plots them in stacked bar graphs. United had most number of negative tweets, however, the relative distribution of negative comments was different for different airlines. Southwest had the most number of negative comments due to customer service issues. Then normalize the reason for negative comment by total number of negative tweets for each airline.

 DOI : 10.18231/2454-9150.2018.0124

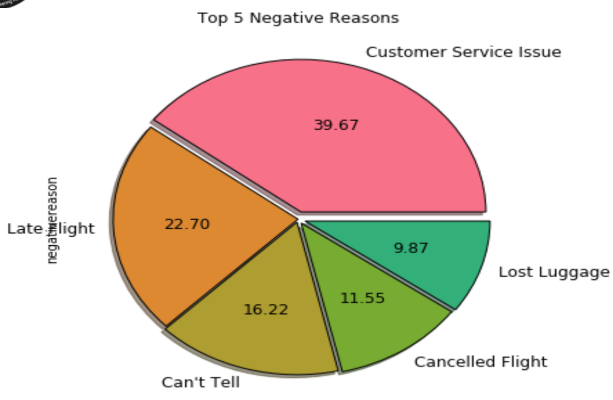

Fig.6. Reason for top 5 negative comments

After normalizing, the contribution of negative tweets due to poorer customer service is higher (more than 50%) for Virgin America, Delta and Southwest. US Airways had the least fractions of negative tweets due to customer service issues. US airways and American Airlines have as much complaints due to customer service as due to lost luggage.

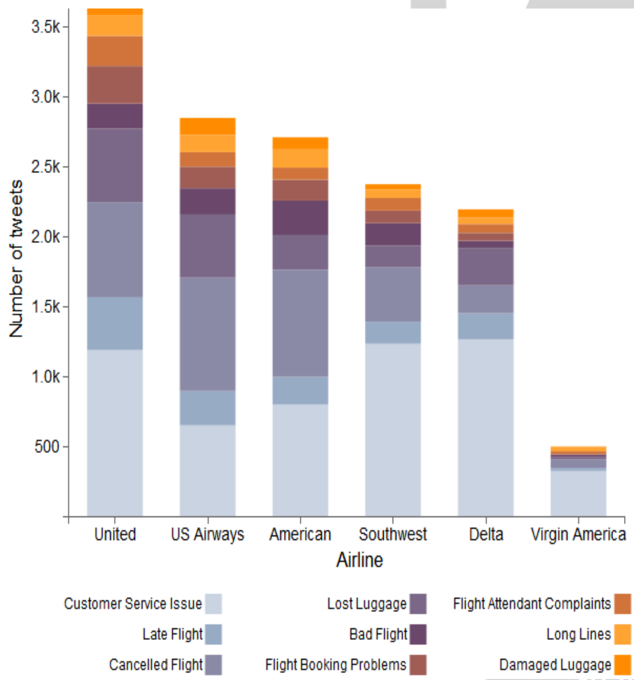

Fig.7. Negative comments by airline services

Fig. 8 represents the word cloud for the negative tweets. These are the frequently repeated words for the negative comments of the airline services. Some of the negative words are flight cancelled, delayed, call hold, weather problem service issue etc. The tweets by location are plotted as shown in Fig.9. Of the 14640 tweets, only 841 had data on location. The locations of tweets across USA are plotted. Tweets appear to be clustered around big airports, like New York, Chicago, Los Angeles, etc. Therefore got data for the 30 busiest airports locations from Wikipedia, and assigned each tweet to the airport nearest to it. Each tweet was made closest to the corresponding airport was assumed. This works well in most cases except when airports are very close to each other. This visualization makes the airline services companies to know at which location, at which airline and at which customer had tweeted negative comments can be analysed and that faults can be resolved by the airline services.

Fig.8. Word Cloud for Negative Mood

Fig.9. Mapping of tweets by location

The experiments with the seven classification models are conducted to evaluate the machine learning classification approaches includes Logistic Regression, KNeighbors, SVC, DecisionTree, RandomForest, AdaBoost and GaussianNB classifier. Accuracy is a ratio of correctly predicted observation to the total observations. Accuracy is the most intuitive performance measure. This performance measure is used to find the accuracy is given in the equation (1)

$$Accuracy = \frac{(True\ Positive + True\ Negative)}{Total\ Population} \qquad (1)$$

where True Positive is the number of correct predictions that the occurrence is positive and True Negative is the number of correct predictions that the occurrence is negative respectively.

Fig 10 represents the comparison of the classifier model. The GaussianNB classifier gets the lowest accuracy, which is 57.24%. The KNeighborsClassifier obtain the accuracy of 58.91%. The Logistic Regression and SVC classifier gets the same accuracy of 64.51%. The Decision Tree Classifier and AdaBoostClassifier contains the accuracy of 75.88% and 78.55% respectively. The accuracy of the

 DOI : 10.18231/2454-9150.2018.0124

Random Forest Classifier model classification reached 81.35% which is the highest accuracy. In our experiment, the accuracy of the Random Forest Classifier is high compared to other classifier and also it will be the most suitable sentiment classification methods for tweets about airline services.

TABLE 2
Accuracy of 7 different classifier models

Classifier Model	Accuracy
Logistic Regressionis	64.51
KNeighbors Classifier	58.91
SVC	64.51
Decision Tree Classifier	75.88
Random ForestClassifier	81.35
AdaBoost Classifier	78.55
GaussianNB Clasifier	57.24

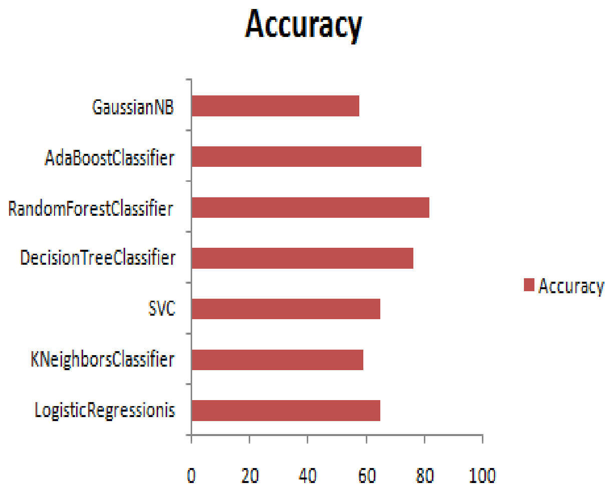

Fig.10. Comparison of classifier model

VI. Conclusion

In recent years, Twitter has become the online customer service platform in the world. In this paper, Twitter Airline Sentiment dataset is extracted and more than 60% as negative comments are only predicted. So the negative tweets are analyzed and the wordcloud of the negative words are created. On analyzing the dataset, United airlines services contains the most number of tweets, followed by US airway and American airlines services. United, US airways and American airlines have more proportions of negative comments whereas, Southwest, Delta and Virgin America had less proportion of negative comments. The customer service issues and late flights are the two main reasons for negative comments of the customer in the airline services are analyzed. US Airways contains the highest proportion of negative tweets due to customer service issues. Here, the customer's problems can be known to the airline services to resolve and the location of the customers are mapped using graphical visualization. Then the seven classification methods are compared for Twitter sentiments of airline services and the best sentiment classifier was selected, which is the Random forest classifier. This classifier can be used for airline services business analysis applications, which will be able to automatically classify customer's satisfaction about airline services.

Acknowledgment

This work was supported by Department of Computer Science and Engineering, Annamalai University and funded by University Grants Commission (UGC) of Indian Union government.

References

[1] Guohuibin Li," Application of Sentiment Analysis: Assessing the Reliability and Validity of the Global Airlines Rating Program", 9th IBA Bachelor Thesis Conference, July 5th, 2017.

[2] Sreenivasan, Nirupama Dharmavaram, Chei Sian Lee, and Dion Hoe-Lian Goh. "Tweeting the friendly skies: Investigating information exchange among Twitter users about airlines." Program: electronic library and information systems 46.1, pp. 21-42, 2012.

[3] Breen, Jeffrey Oliver. "Mining twitters for airline consumer sentiment." Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications , pp. 133, 2012.

[4] Adeborna, Esi, and Keng Siau. "An Approach to Sentiment Analysis–The Case of Airline Quality Rating." *PACIS 2014 Proceedings*, pp.363.

[5] David Mc. A Baker, "Service Quality and Customer Satisfaction in the Airline Industry: A Comparison between Legacy Airlines and Low-Cost Airlines", American Journal of Tourism Reseearch Vo1 2. No. 1,pp.67-77,2013.

[6] Joshua Acosta, Norissa Lamaute," Sentiment Analysis of Twitter Messages Using Word2Vec", Proceedings of Student-Faculty Research Day, CSIS, Pace University, pp. C8-1-7, 2017.

[7] Yun Wan, Dr. Qigang Gao," An Ensemble Sentiment Classification System ofTwitter Data for Airline Services Analysis", IEEE 15th International Conference on Data Mining Workshops,2015.

[8] Yun Wan, Dr. Qigang Gao," An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis", IEEE 15th International Conference on Data Mining Workshops, pp.1318-1325,2015.

[9] Heba Hakh, Ibrahim Aljarah, Bashar Al-Shboul," Online Social Media-based Sentiment Analysis for US Airline companies", Proceedings of the New Trends in Information Technology, pp.25-27,2017.

[10] J. Zhu, H. Zou, S. Rosset, T. Hastie, "Multi-class AdaBoost", 2009.

[11] Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood," Random Forests and Decision Trees", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012.

 DOI : 10.18231/2454-9150.2018.0124