A LITERATURE REVIEW: CUSTOMER SATISFACTION ON AIRLINE TWEETS USING MACHINE LEARNING

Pramod Kumar Panda[1], Dr. Jitendra Sheetlani[2]

[1] *Research Scholar, Sri Satya Sai University of Technology and Medical Sciences, Sehore, (M.P.), India*
[2] *Associate Professor, Sri Satya Sai University of Technology and Medical Sciences, Sehore, (M.P.), India*

ABSTRACT

Nowadays social media gains lots of popularity for expressing their views and for other services. Consumer sentiment analysis is a recent fad for social media related applications such as healthcare, crime, finance, travel, and academics. Twitter is one of the most prominent social networking platforms so far. Millions of users utilize Twitter to share their thoughts and views on various topics of interest every day resulting a huge amount of data. Sentiment Analysis is new way of machine learning to extract opinion orientation (positive, negative, neutral) from a text segment written for any product, organization and all other entities. In this research work, we present the review of literature related to the customer satisfaction on airline tweets using machine learning and also explaining the different machine learning approach which gives customer satisfaction for the social media (Tweets).

Keyword :- *Social Media, Customer satisfaction, Machine Learning, Sentiment Analysis, Tweets*

---

1. INTRODUCTION

Advancement in technology has boosted the availability and use of smart mobile phones. At present, the number of smart phone users is 2.71 billion across the world [21]. The major online social media (SM) platforms i.e. Facebook, Twitter and Instagram are available as mobile applications in the smart phones. Therefore, there is no need to visit cyber cafes to access them, as everything is available in the smart phones. Every piece of information shared on SM carries an emotion, sentiment or feeling. These emotions can be positive, negative and neutral. All these emotions may come from a travel trip, restaurant trip, exhibitions, movies, elections, hospital visits etc. These emotions carries some hidden information related to comfort/discomfort in related areas. Hence, there is a good scope of analyzing this information to detect the patterns of the emotions. This analysis can help us to understand the emotions of the people in respective domain and the reasons behind it.[4]
Air travel is one of the most convenient modes for long distance travel at both national and international level [22]. There are many airline service providers (ASPs) around the world. The competitive world motivates the airlines company to attract the customers can be airfare, travel time, number of stoppages, number of baggage allowed, and existing customer feedback etc. Therefore, all ASPs are working in all these customer service areas to improve their facility and in-flight comfort in order to attract the customers. It is very important to understand the needs and comfort level of customers i.e. customer satisfaction during the flight. Therefore, customer feedback is very important for any airline industry. There could be several possible ways to collect the customer feedback. The most easiest and traditional way is the customer feedback form available during the journey. However, most of the passengers do not show any interest in filling feedback forms. Another shortcoming of this approach is that it may or may not have appropriate questionnaire and may be biased on certain parameters i.e. the feedback form may only have certain specific questions. Other approaches for customer feedback collection could be through online website or online mobile applications of the airlines. After the journey, an email with a link can be sent to the passenger to request for a feedback. However, there is no guarantee of its success. Another approach is to send a message on passenger's mobile phone and ask them to rate your service (1for poor and 5 for excellent) on certain parameters. All these traditional methods opted by the industry are restricted to certain parameters only. The more convenient

way for a passenger is to express their feedback, as they want. Therefore, the most convenient way for the passengers to share their opinions is the social media instead of feedback form. Social media provides a platform where a user can freely express his feedbacks on any issues they observed during flight. Twitter [23] is one of the popular platforms worldwide. The information from Twitter can be utilized to develop a recommender system [24]. In addition, travelers are more comfortable in sharing their views about travel experiences on Twitter.
A variety of major issues affects the emotions of a passenger in air travel. These issues can be cabin crew behavior, food quality, loss of baggage, seat comfort, flight delay, airfare etc. All these issues may give rise to both positive and negative emotions. In addition, if there is a continuous trend of negative tweets for an airline, then it may put a negative impact to the economic growth of the airline company. Therefore, it is important to understand the issues that give rise to negative tweets so that the respective airline company can take appropriate action on time. There are a large number of airlines operating every day to connect different geographical locations [25]. Therefore, we may expect a large number of people travelling every day in these flights. In addition, the number of tweets by passengers for airlines would be very large. Therefore, it is a challenging task to extract the hidden emotion behind a tweet. Therefore, we required some tools and techniques that are able to handle such a large number of tweet database and can provide insights to help airline industry. Machine learning [26] and big data technologies [27] made it possible to analyze huge database and to develop highly accurate prediction or classification models.

2. RELATED WORK

This section research paper presents the earlier work done by the various researchers in the field of customer satisfaction of airline tweets using machine-learning techniques such as logistic regression, SVM, KNN, random forest, Naïve Bayes classifier, Adaboost etc.

Authors	Description	Publishing Year
T. Hemakala and S. Santhoshkumar	In this research, design a framework for sentiment analysis with opinion mining for the case of airlines service feedback. Most available datasets of hotel reviews are not labelled which presents many works for researchers as far as text data pre-processing task is concerned. Twitter is a SNS that has a huge data with user posting, with this significant amount of data, it has the potential of research related to text mining and could be subjected to sentiment analysis. The airline industry is a very competitive market, which has grown rapidly in the past 2 decades. Airline companies resort to traditional customer feedback forms which in turn are very tedious and time consuming. In this work, worked on a dataset comprising of tweets for 6 major Indian Airlines and performed a multi-class sentiment analysis. This approach starts with pre-processing techniques used to clean the tweets and then representing these tweets as vectors using a deep learning concept to do a phrase-level analysis. The analysis was carried out using 7 different classification strategies: Decision Tree, Random Forest, SVM, K-Nearest Neighbors, Logistic Regression, Gaussian Naïve Bayes and AdaBoost. The outcome of the test set is the tweet sentiment.[1]	2018
Guoning Hu et al.	Analyze the opinion of19M Twitter users towards 62 popular industries, encompassing 12,898 enterprise and consumer brands, as well as associated subject matter topics, via sentiment analysis of 330M tweets over a period spanning a month. We find that users tend to be most positive towards manufacturing and most negative towards service industries. In addition, they tend to be more positive or negative when interacting with brands than generally on Twitter. We also find that sentiment towards brands within an industry varies greatly and we demonstrate this using two industries as use cases. In addition, we discover that there is no strong correlation between topic sentiments of different industries, demonstrating that topic sentiments are highly dependent on the context of the industry that they are mentioned in. We demonstrate the value of such an analysis in order to assess the impact of	2017

	brands on social media. We hope that this initial study will prove valuable for both researchers and companies in understanding users' perception of industries, brands and associated topics and encourage more research in this field.[2]	
YasminYashodha	This study examines the extensive strategic analysis of Air AsiaBerhad that has enabled it to sustain its competitive advantage as Asia's leading low cost carrier (LCC). The study demonstrates the diverse business-level, corporate level and competitive strategies of Air AsiaBerhad, played crucial roles in the LCC to successfully penetrate the under-served market segment of the airline industry within the ASEAN region. An in-depth analysis using a wide array of academic resources, relevant financial, legal and management resources and authorized websites, including face-to-face interviews were used to provide a more consequential comprehension on the varied business and international strategies that were implemented by AirAsiaBerhad. This research exhibits critical analysis pertaining to the current macro environment of the aviation industry which includes the PESTEL framework and Porter's Industry Analysis. The competitive environment analysis for Air AsiaBerhad is thoroughly scrutinised to examine the driving determinants that attributed to the organization's competitive advantage in the industry.[3]	2012
M.Vadivukarassi et al.	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.[6]	2018
Janet R. McColl-Kennedy et al.	Contextualized in post purchase consumption in business-to-business settings, the authors contribute to customer experience (CX) management theory and practice in three important ways. First, by offering a novel CX conceptual framework that integrates prior CX research to better understand, manage, and improve CXs—comprised of value creation elements (resources, activities, context, interactions, and customer role), cognitive responses, and discrete emotions at touch points across the customer journey. Second, by demonstrating the usefulness of a longitudinal CX analytic based on the conceptual framework that combines quantitative and qualitative measures. Third, by providing a step-by-step guide for implementing the text mining approach in practice, thereby showing that CX analytics that apply big data techniques to the CX can offer significant insights that matter. The authors highlight six key insights practitioners need in order to manage their customers' journey, through (1) taking a customer perspective, (2) identifying root causes, (3) uncovering at-risk segments, (4) capturing customers' emotional and cognitive responses, (5) spotting and preventing decreasing sales, and (6) prioritizing actions to improve CX. The article concludes with directions for future research.[5]	2018
Rida Khan and SiddhalingUrolagin	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.[7]	2018
Ankita Rane and	The airline industry is a very competitive market which has grown rapidly in the past 2 decades. Airline companies resort to traditional customer feedback	2018

anand kumar	forms which in turn are very tedious and time consuming. This is where Twitter data serves as a good source to gather customer feedback tweets and perform a sentiment analysis. In this paper, we worked on a dataset comprising of tweets for 6 major US Airlines and performed a multi-class sentiment analysis. This approach starts off with pre-processing techniques used to clean the tweets and then representing these tweets as vectors using a deep learning concept (Doc2vec) to do a phrase-level analysis. The analysis was carried out using 7 different classification strategies: Decision Tree, Random Forest, SVM, K-Nearest Neighbors, Logistic Regression, Gaussian Naïve Bayes and AdaBoost. The classifiers were trained using 80% of the data and tested using the remaining 20% data. The outcome of the test set is the tweet sentiment (positive/negative/neutral). Based on the results obtained, the accuracies were calculated to draw a comparison between each classification approach and the overall sentiment count was visualized combining all six airlines.[8]	
Yun Wan and QigangGao	In airline service industry, it is difficult to collect data about customers' feedback by questionnaires, but Twitter provides a sound data source for them to do customer sentiment analysis. However, little research has been done in the domain of Twitter sentiment classification about airline services. In this paper, an ensemble sentiment classification strategy was applied based on Majority Vote principle of multiple classification methods, including Naive Bayes, SVM, Bayesian Network, C4.5 Decision Tree and Random Forest algorithms. In our experiments, six individual classification approaches, and the proposed ensemble approach were all trained and tested using the same dataset of 12864 tweets, in which 10 fold evaluation is used to validate the classifiers. The results show that the proposed ensemble approach outperforms these individual classifiers in this airline service Twitter dataset.[12]	2015
Mustafa Altinkök	This research was conducted for the purpose of analyzing the effect of the movement education program through a 12-week-coordination on the development of basic motor movements of pre-school children. A total of 78 students of pre-school period, 38 of whom were in the experimental group and 40 of whom were in the control group, were incorporated into the study in line with their own consent after their families had also been informed.[21]	2016
M.Vadivukarassi et al.	Analyzed 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.[11]	2018
Bee Yee Liau and Pei Pei Tan	The purpose of this paper is to study the consumer opinion towards the low-cost airlines or low-cost carriers (LCCs) (these two terms are used interchangeably) industry in Malaysia to better understand consumers' needs and to provide better services. Sentiment analysis is undertaken in revealing current customers' satisfaction level towards low-cost airlines. About 10,895 tweets (data collected for two and a half months) are analyzed. Text mining techniques are used during data pre-processing and a mixture of statistical techniques are used to segment the customers' opinion. The results with two different sentiment algorithms show that there is more positive than negative polarity across the different algorithms. Clustering results show that both K-	2016

	Means and spherical K-Means algorithms delivered similar results and the four main topics that are discussed by the consumers on Twitter are customer service, LCCs tickets promotions, flight cancellations and delays and post-booking management.[13]	
Xiang Ji et al.	An important task of public health officials is to keep track of health issues, such as spreading epidemics. In this paper, we are addressing the issue of spreading public concern about epidemics. Public concern about a communicable disease can be seen as a problem of its own. Keeping track of trends in concern about public health and identifying peaks of public concern are therefore crucial tasks. However, monitoring public health concerns is not only expensive with traditional surveillance systems, but also suffers from limited coverage and significant delays.[14]	2015

3. Machine Learning Classification

Machine Learning is a perception that consents the machine to acquire from examples and experience, and that moreover deprived of being overtly programmed. Thus, instead of you scripting the code, what you do is you feed data to the generic technique, and the technique/ machine builds the logic based on the given data.[15] It permits the computers system or the machines to construct data-driven decisions rather than being explicitly programmed for carrying out a certain task. These programs or techniques are designed in a way that they learn and improve over time when are exposed to new data.

Machine Learning Technique is trained using a training data set to create a model. When new input data is introduced to the ML technique, it makes a prediction on the basis of the model. The prediction is evaluated for accuracy and if the accuracy is acceptable, the Machine Learning technique is deployed. If the accuracy is not acceptable, the Machine Learning technique is trained again and again with an augmented training data set.[18,19] This is just a very high-level example as there are many factors and other steps involved.

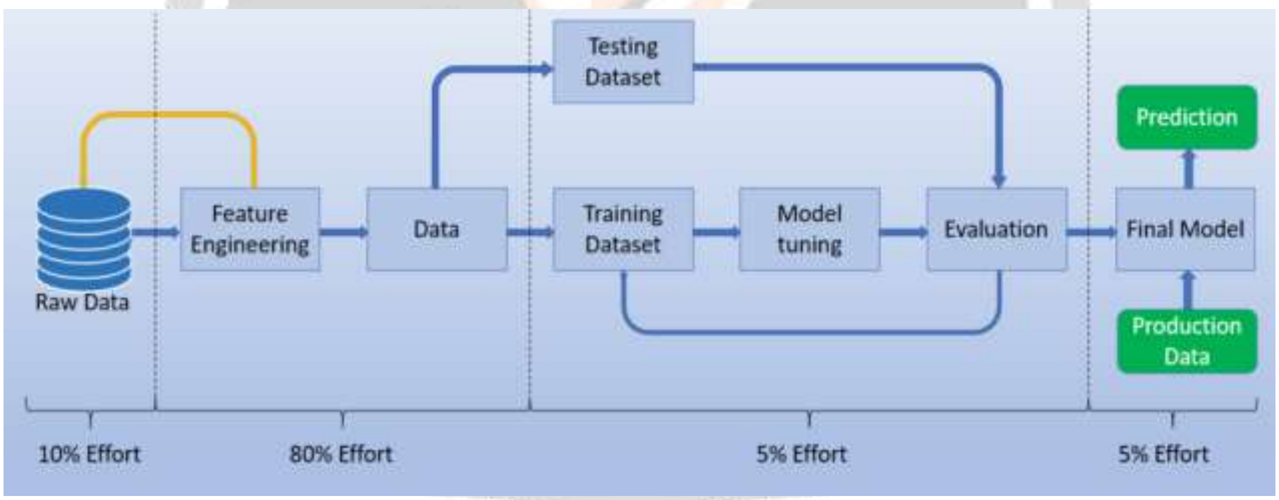

Fig.1: Working steps of Machine learning technique [20]

3.1 Types of Machine Leaning Techniques

There are three important types of Machine Learning Techniques such as supervised learning, unsupervised learning and reinforcement learning which we are discussing in detail:

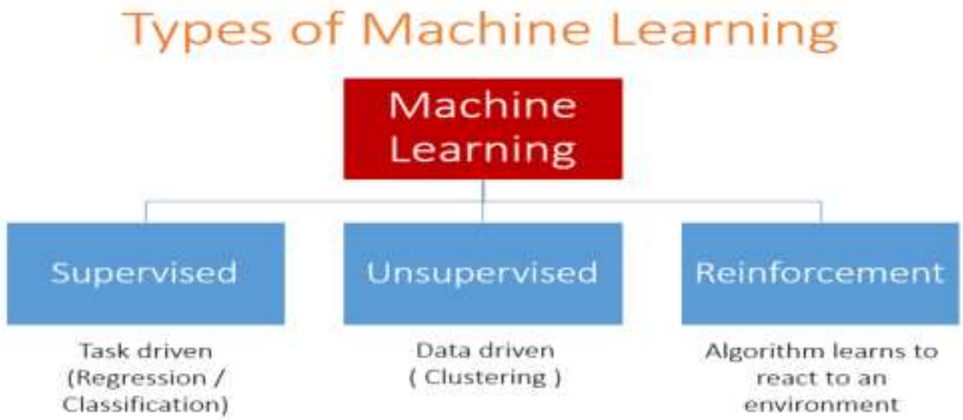

Fig.2: Classification of Machine Learning Techniques

3.1.1 Supervised Learning

Supervised Learning is the most popular paradigm for performing machine learning operations. It is widely used for data where there is a precise mapping between input-output data. The dataset, in this case, is labeled, meaning that the algorithm identifies the features explicitly and carries out predictions or classification accordingly. [9] As the training period progresses, the algorithm is able to identify the relationships between the two variables such that we can predict a new outcome. Resulting Supervised learning algorithms are task-oriented. As we provide it with more and more examples, it is able to learn more properly so that it can undertake the task and yield us the output more accurately. Some of the algorithms that come under supervised learning are as follows: Linear regression, random forest, support vector machine, artificial intelligence [10], etc.
There are two main types of supervised learning problems: they are classification that involves predicting a class label and regression that involves predicting a numerical value.[10]

- Classification: Supervised learning problem that involves predicting a class label.
- Regression: Supervised learning problem that involves predicting a numerical label.

Both classification and regression problems may have one or more input variables and input variables may be any data type, such as numerical or categorical.[45]

3.1.2 Unsupervised Learning

Unsupervised machine learning holds the advantage of being able to work with unlabeled data. This means that human labor is not required to make the dataset machine-readable, allowing much larger datasets to be worked on by the program. The model learns through observation and finds structures in the data. Once the model is given a dataset, it automatically finds patterns and relationships in the dataset by creating clusters in it.[11] In supervised learning, the labels allow the algorithm to find the exact nature of the relationship between any two data points. However, unsupervised learning does not have labels to work off of, resulting in the creation of hidden structures. Relationships between data points are perceived by the algorithm in an abstract manner, with no input required from human beings. The creation of these hidden structures is what makes unsupervised learning algorithms versatile. Instead of a defined and set problem statement, unsupervised learning algorithms can adapt to the data by dynamically changing hidden structures.[11] This offers more post-deployment development than supervised learning algorithms. What it cannot do is add labels to the cluster, like it cannot say this a group of apples or mangoes, but it will separate all the apples from mangoes. Suppose we presented images of apples, bananas and mangoes to the model, so what it does, based on some patterns and relationships it creates clusters and divides the dataset into those clusters. Now if a new data is fed to the model, it adds it to one of the created clusters. The example of unsupervised learning is k-mean clustering, principle component analysis, SVD, FP-growth etc.[16] There are many types of unsupervised learning, although there are two main problems that are often encountered by a practitioner: they are clustering that involves finding groups in the data and density estimation that involves summarizing the distribution of data.[11]

- Clustering: Unsupervised learning problem that involves finding groups in data.
- Density Estimation: Unsupervised learning problem that involves summarizing the distribution of data.

3.1.3 Reinforcement Learning

Reinforcement learning directly takes inspiration from how human beings learn from data in their lives. It features an algorithm that improves upon itself and learns from new situations using a trial-and-error method. Favorable outputs are encouraged or 'reinforced', and non-favorable outputs are discouraged or 'punished'. Based on the psychological concept of conditioning, reinforcement learning works by putting the algorithm in a work environment with an interpreter and a reward system. In every iteration of the algorithm, the output result is given to the interpreter, which decides whether the outcome is favorable or not.[12] In case of the program finding the correct solution, the interpreter reinforces the solution by providing a reward to the algorithm. If the outcome is not favorable, the algorithm is forced to reiterate until it finds a better result. In most cases, the reward system is directly tied to the effectiveness of the result.[12] In typical reinforcement learning use-cases, such as finding the shortest route between two points on a map, the solution is not an absolute value. Instead, it takes on a score of effectiveness, expressed in a percentage value. The higher this percentage value is, the more reward is given to the algorithm. Thus, the program is trained to give the best possible solution for the best possible reward.[17] This simple feedback reward is known as a reinforcement signal.

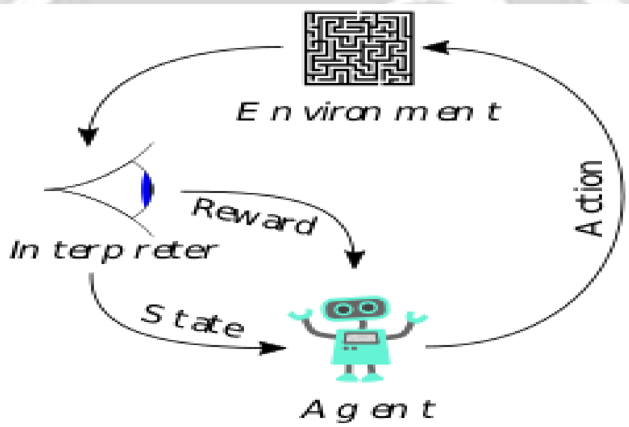

Fig. 3: Example of reinforcement learning

3.2 Classification Approach

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 [28]. 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 [28].

C. 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 [29].

D. 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.

E. 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 [28].

F. 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[30].

G. 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].

4. CONCLUSION

The use of social media is increasing for the purpose of marketing or presenting their views and it is used in various applications such as Health, travel, crime, finance, academic etc. With the advancement of technology, a massive amount of social web-data increasing in terms of volume, subjectivity, and heterogeneity, becomes challenging to process it manually. This paper presents, for airline tweets customer satisfaction various machine learning techniques/ algorithms has been developed and implemented and after reviewing literature about this we found the merely no classification technique is very much efficient in the customer satisfaction of airline tweets. So in future, we need to design an ensemble classifier, which can satisfy the customer or consumer problem efficiently, and accurately.

Reference

[1] T. Hemakala and S. Santhosh kumar, "Advanced Classification Method of Twitter Data using Sentiment Analysis for Airline Service", International Journal of Computer Sciences and Engineering, Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693, pp-331-335.

[2] Guoning Hu, Preeti Bhargava, Saul Fuhrmann, Sarah Ellinger and Nemanja Spasojevic, "Analyzing users' sentiment towards popular consumer industries and brands on Twitter", arXiv:1709.07434v1 [cs.CL] 21 Sep 2017,pp-1-8.Twitter https ://en.wikip edia.org/wiki/Twitter. Accessed 23 Jan 2018.

[3] Yasmin Yashodha, "AirAsia Berhad: Strategic analysis of a leading low cost carrier in the Asian Region", Elixir Mgmt. Arts 51 (2012), pp-11164-11171.

[4] Number of flights in a day: https ://www.quora .com/How-many-airpl anes-fly-each-day-in-the-world . Accessed 25 Jan 2018.

[5] Janet R. McColl-Kennedy , Mohamed Zaki, Katherine N. Lemon, Florian Urmetzer, and Andy Neely, “Gaining Customer Experience Insights That Matter”, Journal of Service Research 2018, Vol. 22(1), DOI: 10.1177/1094670518812182 journals.sagepub.com/home/jsr, pp-8-26.
[6] M.Vadivukarassi, N.Puviarasan, P.Aruna, “An Exploration of Airline Sentimental Tweets with Different Classification Model”, International Journal for Research in Engineering Application & Management (IJREAM), DOI : 10.18231/2454-9150.2018.0124,ISSN : 2454-9150 Vol-04, Issue-02, May 2018, pp-72-77.
[7] Rida Khan, Siddhaling Urolagin, “Airline Sentiment Visualization, Consumer Loyalty Measurement and Prediction using Twitter Data”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 6, 2018, pp-380-388.
[8] Ankita Raneand anand kumar , “Sentiment Classification System of Twitter Data for US Airline Service Analysis”, 2018 42nd IEEE International Conference on Computer Software & Applications, IEEE DOI 10.1109/COMPSAC.2018.00114, pp- 769-773
[9] Kusen E and M Strembech. An analysis of tweeter discussion on the 2016 Austrian presidential election, arXiv preprint arXiv: 1707.09939, 2017.
[10] Stieglitz S and L Dang-Xuan. Emotions and information diffusion in social media - sentiment of microblogs and sharing behavior. Journal of management information systems, 29 (4):217-248, 2013. https://doi.org/10.2753/MIS0742-1222290408.
[11] M.Vadivukarassi et al. An Exploration of Airline Sentimental Tweets with Different Classification Model. International Journal for Research in Engineering Application & Management (IJREAM) ISSN: 2454-9150 Vol-04, Issue-02, May 2018.
[12] Yun Wan and Qigang Gao, “An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis”, 2015 IEEE 15th International Conference on Data Mining Workshops, 978-1-4673-8493-3/15 $31.00 © 2015 IEEE DOI 10.1109/ICDMW.2015.7, pp-1318-1325.
[13] Bee Yee Liau and Pei Pei Tan, “Gaining customer knowledge in low cost airlines through text mining”, Industrial Management & Data Systems Vol. 114 No. 9, 2014 pp. 1344-1359.
[14] Xiang Ji1 Soon Ae Chun2 Zhi Wei1 James Geller, “Twitter sentiment classification for measuring public health Concerns”, Soc. Netw. Anal. Min. (2015) 5:13 DOI 10.1007/s13278-015-0253-5, pp-1-25.
[15] Seunghyun BP, C Ok, B Chae. Using Twitter Data for Cruise Tourism Marketing and Research. Journal of Travel & Tourism Marketing, 33(6):885-898, 2016. https://doi.org/10.1080/10548408.2015.1071688.
[16] 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.
[17] 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.
[18] J. Zhu, H. Zou, S. Rosset, T. Hastie, “Multi-class AdaBoost”, 2009.
[19] Mustafa Altinkök, “The Effects of Coordination and Movement Education on Pre School Children’s Basic Motor Skills Improvement”, Universal Journal of Educational Research 4(5): 1050-1058, 2016 http://www.hrpub.org DOI: 10.13189/ujer.2016.040515, pp-1050-1058.
[20] 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.
[21] Swapn Joshi. Aspect based sentiment analysis for United States of America Airlines. National College of Ireland, 2017
[22] Twitter https ://en.wikip edia.org/wiki/Twitt er. Accessed 23 Jan 2018.
[23] Abel F, Gao Q, Houben G-J, Tao K (2013) Twitter-based user modeling for news recommendations. In: Rossi F (ed) IJCAI 2013, proceedings of the 23rd international joint conference on artificial intelligence, Beijing, China, August 3–9, 2013.IJCAI/AAAI.
[24] Number of flights in a day: https ://www.quora .com/How-many-airpl anes-fly-each-day-in-the-world . Accessed 25 Jan 2018.
[25] Han J, Kamber M. (2006) Data mining: concept and techniques. 3rd ed.
[26] Zomaya AY, Sakr S. Handbook of big data technologies. Berlin: Springer; 2017. https ://doi.org/10.1007/978-3-319-49340 -4.
[27] 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.
[28] J. Zhu, H. Zou, S. Rosset, T. Hastie, “Multi-class AdaBoost”, 2009.
[29] 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.
[30] 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.