A Comparative study on Airline Recommendation System Using Sentimental Analysis on Customer Tweets

Divisha khaturia, Aditi saxena, Syed Muzamil Basha,
Iyengar N.Ch.S.N[1] and Ronnie D. Caytiles[2]

VIT, Vellore-632014
[1]Sreenidhi Institute of Science and Technology, Ghatkesar, Hyderabad, India.
[2]Multimedia Engineering Department, Hannam University, Daejeon, Korea.
srimannarayanach@sreenidhi.edu.in, rdcaytiles@gmail.com

Abstract

*Recommendation compels in many fields and has become a great immersion in the zone of research. Important factor for recommendation system is to recognize user's personalized historic behaviors through analyzing different tweets made by them. Recommender systems have become common in the research, where many opinions are replenished on the grounds of algorithms. The paper is on comparative study of the techniques used for airlines recommendation systems. The basic motives behind this paper are **1)**To build a model for airlines this performs sentiment analysis on customer reviews in order to have an attend concise feedback about the airlines. **2)**Encouraging for the most important facet of advancement so they can improve the given customers' complaints. Here, we perform Sentimental Analysis on the Twitter Airlines dataset from Kaggle. Vital accuracy is achieved, which shows the reliability of the project for future prediction.*

1. Introduction

Information over loaded on the searching and recommender systems has led to increase in challenges for users to obtain a more appropriate searched and recommended outcomes, all at a time and under one system. Various recommender systems enlisting dissimilar algorithms and methodologies are used to depict some challenges. This research paper highlights on graphical model that offers a standard data validations and support varied recommendation algorithms. More algorithms and models required for Flight Recommender Systems are explored that depicts how sentimental analysis can be appealed to airlines recommendation systems to make ideal suggestions. A summoning task for recommendation is to enhance the efficiency of the recommender system for new users, where usually tweet is scrutinized about particular flight and interactions between users are found as an outcome. This paper centers on algorithms used by researchers to detect the definitiveness of recommender systems. However, due to the nature of the problem, the accuracy of sentiment analysis on single sentence like movie reviews never reaches above 80% for the past 7 years [1]. Looking at last year's project on twitter [2],their accuracy was 59.32% to 63.71%, depending on different models. We achieved near 20% more than their result, which is a significant improvement. Since tweets texts are usually short and verbal, the same problem presents in our dataset as well. However, even though the tweets are short, there are strong indicative words. Specific words can be used as indicators for spam and achieve good test accuracy.

2. Literature Survey

A large amount of work is done in sentiment analysis, mainly in the areas of product reviews, movie reviews, and blogs. [1]Pang and Lee researched the performance of various machine learning techniques in the specific domain of movie reviews but these approaches were not able to perform as well as traditional sentiment analysis methods. [6] AlecGo, Richa Bhayani and Lei Huang used a approach for automatically classifying the sentiment of Twitter messages by using NaiveBayes, maximum entropy classification, and support vector machines where they were getting the almost same result for all the methods. [12] Andrew McCallum, Kamal Nigamdida comparison between two different first-order probabilistic classifiers and it was found out that both the methods reduce the error by an average of 27% and 50% respectively . [7] Alexander Pak, Patrick Parou used twitter data to perform a opinion mining where they used Tree Tagger for POS-tagging and observed the difference in distributions among positive, negative and neutral sets. [10] Albert Bifet and Eibe Frank used the twitter data for knowledge discovery for this they used Kappa statistic for evaluation in time-changing data streams. [2] Mohit Tuteja used Flight Recommendation client (FRC) to recommend flights to customers on the basis of the feedback provided by the customers. The data is used from each user to provide more personalized recommendation. In [16] the author, graphical model based, supervised, semi-supervised and unsupervised algorithms was used and a comparison between these methods is done, in order to provide efficient flight recommendation for the users. In [18] the author, suggested methodology isused on a variety of data mining techniques such as web usage mining, decision tree induction, association rule mining and the product taxonomy in order to recommend user products based on their past shopping pattern. In [4] the author, Association Rule in order to find the patterns between various products purchased by the customer and based upon that are commendations are generated for each and every customer. Sentimental Analysis has been a widely used area over the years and still it leaves a lot to be researched. [Fried, Surdeanu, Kobourov, Hingle, Bell] investigated the predictive power behind the language of food on social media [14]. They collected a corpus of over three million food-related posts from Twitter and demonstrate that many latent population characteristics can be directly predicted from this data: overweight rate, diabetes rate, political leaning, and homegeographical location of authors. For all tasks, their language-based models significantly outperform the majority class baselines [17]. [Logunov, Panchenko] generated Twitter sentiment indices by analyzing a stream of Twitter messages and categorizing messages in terms of emoticons, pictorial representations of facial expressions in messages. Based on emoticons they generated daily indices. Then they explored the time-series properties of these indices by focusing on seasonal and cyclical patterns, persistence and conditional. They developed a new method that combines existing approaches, providing the best coverage results and competitive agreement. [Gomide,Veloso, MeiraJ r] analysed how Dengue epidemic is reflected on Twitter and to what extent that information can be used for the sake of surveillance. Dengue is a mosquito-borne infectious disease that is a leading cause of illness and death in tropical and subtropical regions, including Brazil. They proposed an active surveillance methodology that is based on four dimensions: volume, location, time and public perception. First they explored the public perception dimension by performing sentiment analysis. This analysis enables their to filter out content that is not relevant for the sake of Dengue surveillance. [Hong and Davison] proposed several schemes to train a standard topic model and compare their quality and effectiveness through a set of carefully designed experiments from both qualitative and quantitative perspectives. They showed that by training a topic model on aggregated messages they could obtain a higher quality of learned model which results in significantly better performance in two real world classification

3. Motivation

This field of research has made a wide progress in technology and around. Applying data mining algorithms on textual data is yet another concern as there is plenty of information which is not stored in facts and figures, rather is not statistical. [11] Whether it comes to recommending a particular airline system to the customers based on tweets or any other fact that is related to textual data, converting it to facts or a language that the machine understands is no less than a hectic task to do. [12] There are many recommendations or decisions that are dependent on the responses one gets from the individuals using that one particular application. Thus in order to carry on a further study and provide a conclusion that is more relatable and easy to understand, a comparative study is made on the most commonly used methods and evaluating the score as outcomes of the comparisons is presented the paper.

4. Problem Statement

Airlines are vast industries spreading across large number of continents, countries and city. Daily a large number of flights operates between various locations covering a large amount of distance and provide services to many people. [9] The major problem in the airline industry is the lack of proper recommendation system for the user based on the past experience of the customers. We aim to provide a proper mechanism which helps the customers to opt for an airline in accordance to their comfortable experience. When it comes to evaluating data given in textual format, sentimental analysis is serves the best to its purpose. Collecting tweets on a large scale analyzing them and categorizing them as positive, negative and neutral helps to provide a better rating and reviews which then helps the customers to opt for an option as per their convenience. [8] Naïve Bayes here provides a comparative platform to the research done through sentimental analysis is so as to recognize how accurate there commendations done on airlines using sentimental analysis is are. Here, we use the tweets made on six major U.S. Airlines and splitting them as positive, negative and neutral, we calculate the score for each airline in order to rank each with respect to one another [15]. A comparative study is made with another aspect of this field of opinion mining known as Naïve Bayes using which there commendations are made to the customer, both statistically as well as graphically.

5. Workflow

Internet plays a vital role in this work as the dataset is collected from www.Kaggle.com which contains the tweets for the above mentioned six airlines popularly used in the United States. Customer Tweets, refers to the statements given by various customers who have travelled through these airlines so far. Referring to the words used by them, the tweets are classified as positive, negative and neutral and then in carried towards the next step. Applying NaiveBayes and sentimental analysis, After the categorization of the tweets in to positive, negative and neutral the two algorithms, *i.e.* NaiveBayes and sentimental analysis is applied in to which the total number of positive, negative and neutral tweets are counted. Calculation of Scores and comparisons between the methods.

The individual score for each airline is calculated using both naivebayes and sentimental analysis, based on which a comparative study is made leading to chose a better algorithm out of the two. The steps are as follows: The tweets recorded are imported from the dataset and separated for every in particular airline, Sentimental analysis and NaiveBayes algorithm are applied, the positive negative and neutral tweets

are calculated and scores are calculated using both methods. Comparisons are performed and accuracy is judged accordingly.
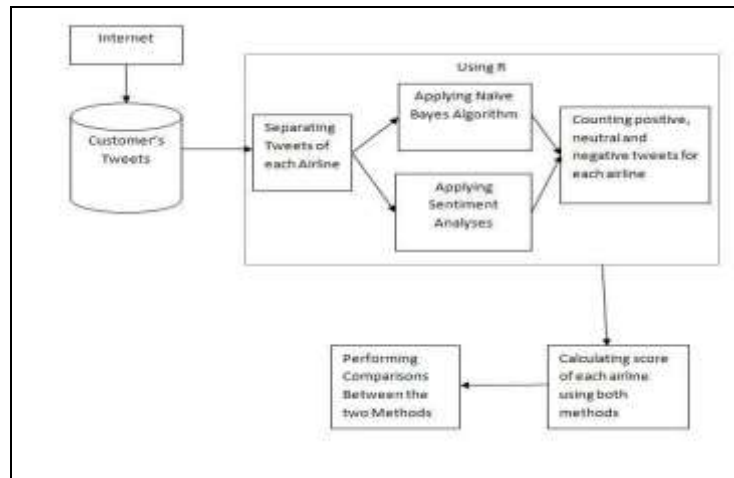
Figure 1. Series of Activities Undergone

The second and following pages should begin 1.0 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be 1-3/16 inches (2.86 cm) from the bottom edge of the page for 8.5 x 11-inch paper; for A4 paper, approximately 1-5/8 inches (4.13 cm) from the bottom edge of the page.

6. Methodology

The sentiment analysis tweets are positive, negative, and neutral and are collected from kaggle. The negative reason tweets are bad flights and cancelled flights, customer services issues, damaged luggage, flight attendant complaints, flight booking problems, late flights, long lines, and lost luggage.

Figure. 2 Dataset for Analysis

In the preprocessing step, non-English word, symbols and website links are eliminated. Then the whole data set is randomly separated into training set and test set. The dictionary is made based on the training data and all sentences are broken down in to list of words: Delete common words such as a, an, to, of, on *etc*. With

high frequency but little semantic usage**.** Stem words, such as "thanks" and "thank" as one word. Delete low frequency words that appear once to reduce the size of dictionary for calculation efficiency. A feature matrix is built to convert the textual information in to numerical information. In the feature matrix, the number of rows indicates the number of samples, the number of columns the length of the dictionary, and each element indicates whether the specific word has appeared in the current review, 1 for existence and0 for absence. The methodology used is sentimental analysis and with the help of Rtool the outcomes are evaluated. Recommendations will be made on the basis of the comparisons between various airlines in order to help the customers chose best to their concern.

abort	abrubt	accelerate	accid	accident	account	acknowledged	across	activ	actual	adjust	advis
0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1
5	1	1	3	1	1	0	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	1
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

Figure. 3 Frequency Matrix of the Tweets Collected

Sentiment analysis also known as opinion mining or emotion AI concerns to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely put to raise the customer materials like reviews and survey responses, online and social media, and health care materials for applications ranging from marketing to customer service to clinical medicine and so on. Sentiment analysis aims to depict the view point of a speaker, writer, or anyone else with respect to some topic or the overall context to a document, interaction, or event. The attitude may be a judgment or evaluation, affective state, or the intended emotional communication.

Algorithm
Sentiment Analysis is implemented to analyze the sentiment of social media content, like tweets and status updates. Sentimental analys is follows up the procedure as follows: Select a string of characters and rate the as positive, negative and neutral respectively. NaïveBayes classifiers are referred to as the simple probabilistic which works when applied Bayes' theorem with an effective independence assumptions between its features (called naive). NaiveBayes throughout has been an extensive study since the early 20th century. When made an apt pre-processing, it became competitive in this domain with advanced methods that included support vector machines. NaiveBayes classifiers are highly scalable as they require a number of parameters linear in the number of variables (features/predictors) in a learning problem. In the statistics and computer science literature, NaiveBayes models are known under a variety of names, including simple Bayes and independence Bayes.[4] All these names reference the use of Bayes' theorem in the classifier's decision rule, but naiveBayes is not (necessarily) a Bayesian method. Score calculation helps in identifying which airlines to chose for any customers and is calculated using the Equation 1 & 2.

$$Score = \frac{positive + negative + neutral}{Total\,number\,of\,Tweets} \times 100 \qquad (1)$$

$$p(c \mid x) = \frac{p(x \mid c)\,p(c)}{p(x)} \qquad (2)$$

p(c|x) is posterior probability, p(x) is predictor prior probability, p(c) is class Prior probability and p(x|c) is Likelihood.

7. Results and Discussions

The algorithms are applied to each airline individually and the then predicted scores are compared as in Figure 4, 5 &6.
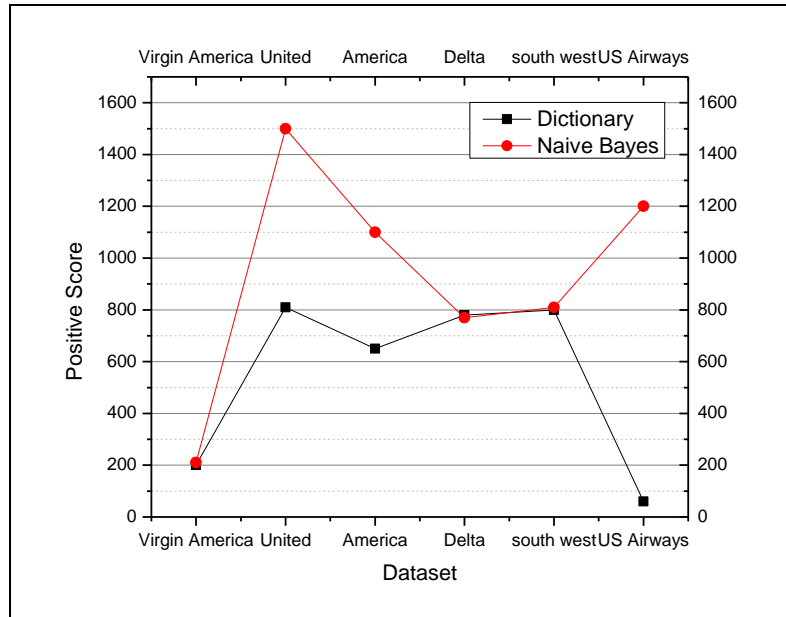

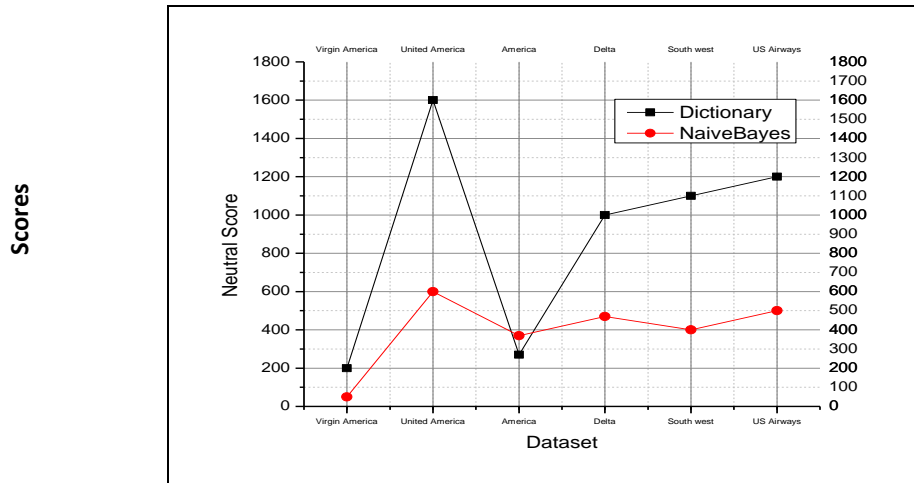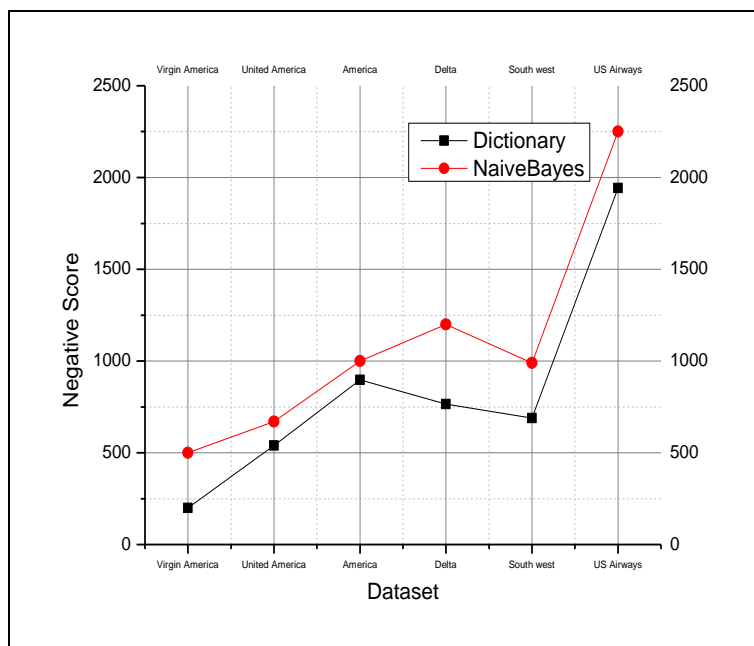**Figure 4. Plot of Positive Scores**

Figure 5. Plot of Neutral Scores

Figure. 6 Plot of Negative Scores

Here, x-axis represents the name of each airline and they-axis represents the scores calculated using sentiment analysis and Naïvebayes. Sentiment analysis is represented using diamonds while Naïvebayes is represented using squares.

8. Conclusion

The study illustrates how the recommendation systems work based on the customer tweets. Textual data collected from Kaggle is streamed and opinion mining techniques are carried out which then helps one to choose an option in accordance with their convenience. The comparisons made between the two methods are presented in the results graphically are generated. Calculating the scores separately relates to the way one can predict the efficiency of the airlines based on which one can take the decision that which air line should be chosen as per the preferences of each passenger individually. A large amount of work is done in sentiment analysis, mainly in the areas of product reviews, movie reviews, and blogs *etc*. In near future, the demand for such techniques are expected to only increase with advances in techniques for applying data mining and also in the research areas.

References

[1] B. Pang and L. Lee. "Opinion Mining and Sentiment Analysis (Foundations and Trends (R) in Information Retrieval)", **(2008)**.

[2] A. Go, R. Bhayani and L. Huang, "Twitter sentiment classification using distant supervision", CS224N Project Report, Stanford, vol. 1, no. 2009, **(2009)**, p. 12.

[3] A. Go, L. Huang and R. Bhayani, "Twitter sentiment analysis", Entropy, vol. 17, **(2009)**, p. 252.

[4] A. Sarlan, C. Nadam and S. Basri, "Twitter sentiment analysis", In Information Technology and Multimedia (ICIMU), 2014 International Conference on, IEEE, **(2014)**, pp. 212-216.

[5] E. Kouloumpis, T. Wilson and J. D. Moore, "Twitter sentiment analysis: The good the bad and the omg!", Icwsm, vol. 11, no. 538-541, **(2011)**, p. 164.

[6] H. Wang, D. Can, A. Kazemzadeh, F. Bar and S. Narayanan, "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle", In Proceedings of the ACL 2012 System Demonstrations, Association for Computational Linguistics, **(2012)**, pp. 115-120.

[7] A. K. Jose, N. Bhatia and S. Krishna, "Twitter sentiment analysis", In Seminar Report, National Institute of Technology Calicut, **(2010)**.

[8]   A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining", In LREc, vol. 10, no. 2010, **(2010)**.

[9]   J. Faret and J. Reitan, "Twitter sentiment analysis", **(2011)**.

[10] A. Go, R. Bhayani and L. Huang, "Twitter sentiment classification using distant supervision", CS224N Project Report, Stanford, vol. 1, no. 2009, **(2009)**, p. 12.

[11] J. Tatum and J. Travis Sanchez, "Twitter Sentiment Analysis", CS29 Machine Learning course at Stanford University, **(2013)**.

[12] S. Rosenthal, N. Farra and P. Nakov, "SemEval-2017 task 4: Sentiment analysis in Twitter", In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), **(2017)**, pp. 502-518.

[13] D. Zimbra, M. Ghiassi and S. Lee, "Brand-related Twitter sentiment analysis using feature engineering and the dynamic architecture for artificial neural networks", In System Sciences (HICSS), 2016 49th Hawaii International Conference on, IEEE, **(2016)**, pp. 1930-1938.

[14] M. Ghiassi, D. Zimbra and S. Lee, "Targeted Twitter Sentiment Analysis for Brands Using Supervised Feature Engineering and the Dynamic Architecture for Artificial Neural Networks", Journal of Management Information Systems, vol. 33, no. 4 **(2016)**, pp. 1034-1058.

[15] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani and V. Stoyanov, "SemEval-2016 Task 4: Sentiment Analysis in Twitter," In SemEval@ NAACL-HLT, **(2016)**, pp. 1-18.

[16] P. Nakov, S. Rosenthal, S. Kiritchenko, S. M. Mohammad, Z. Kozareva, A. Ritter, V. Stoyanov and X. Zhu, "Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts", Language Resources and Evaluation, vol. 50, no. 1, **(2016)**, pp. 35-65.

[17] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani and V. Stoyanov, "Evaluation Measures for the Semeval-2016 task 4: Sentiment Analysis in Twitter (Draft: Version 1.12)", In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016), San Diego, California, June. Association for Computational Linguistics, **(2016)**.

Authors

Divisha khaturia and **Aditi saxena**, they are both are M. Tech. graduate studens with specialization in Big Data Analytics from Vellore Institute of Technology. Their interest is in Machine Learning and Data Science. Currently they are working in the field of Machine Learning, designing Recommendation engine and prediction models.

Syed Muzamil Basha, he had his Bachelor of Science in Information Technology at SITAMS, MTech in Information Technology (Networking) at VIT University and currently doing his research at VIT University. His research area are Wireless Sensor Networks, Text Mining and Big Data Predictive Analytics.

N. Ch. S. N. Iyengar (b 1961), he currently Professor at the Sreenidhi Institute of Science and Technology (SNIST) Yamnapet, Ghatkesr, Hyderabad, Telengana, India. His research interests include Agent-Based Distributed Computing, Intelligent Computing, Network Security, Secured Cloud Computing and Fluid Mechanics. He had 32+ years of experience in teaching and research, guided many scholars, has authored several textbooks and had nearly 200+ research publications in reputed peer reviewed international journals. He served as PCM/reviewer/keynote speaker/ Invited speaker. He received 2017 achievement award for his contributions in teaching and research by IOSRD.

Ronnie D. Caytiles, he had his Bachelor of Science in Computer Engineering- Western Institute of Technology, Iloilo City, Philippines, and Master of Science in Computer Science– Central Philippine University, Iloilo City, Philippines. He finished his Ph.D. in Multimedia Engineering, Hannam University, Daejeon, Korea. Currently, he serves as an Assistant Professor at Multimedia Engineering department, Hannam University, Daejeon, Korea. His research interests include Mobile Computing, Multimedia Communication, Information Technology Security, Ubiquitous Computing, Control and Automation