



# Airline Reservation Using Sentiment Analysis with Naïve Bayes Classifier

Ruth Atubonengi<sup>1</sup>, V.I.E Anireh<sup>2</sup>, Daniel Matthias<sup>3</sup>

Department of Computer Science, Rivers State University, Port Harcourt, Nigeria<sup>1,2,3</sup>

**Abstract-** Customer reviews are a powerful strategic tool for industry experts and for the Airline industry it deserves great attention. Before now, traditional methods used for this review provide inconsistent information about data in the airline industry. Sentiment analysis from online comments posted on review or micro-blogging sites has the potential to assist customers decide best airline of choice by analyzing other customer opinions. This paper presents information about airline reservation system using sentimental analysis with Naïve bayes classifier. The system uses airline information dataset, containing 16 columns of which few columns were selected by means of feature extraction. The selected columns are the text column with the airline sentiment column. Stop words were used in eliminating words that are so commonly used that they carry little information. LabelEncoder function was used in converting the y variable, which arranges the sentiment analysis columns into numbers ranging from -1, 0 to 1, where -1 represents a neutral comment, 1 represents positive and 0 represents negative. Further 70% of the dataset was used in training the Naïve bayes classifier while 30% of data was used for testing. Result show training accuracy of about 95% and a testing accuracy of about 79%. The Naïve bayes classifier was then saved and deployed to web using python flask.

**Keywords-** Airline Reservation, Sentiment Analysis, Naïve Bayes Classifier, Stopwords

## 1. INTRODUCTION

Social media such as Twitter created since 2006 has increasing patronage day by day with a wide-spread of customers expressing their feedbacks about goods and services they deal with. Many organizations attract customers with the use of these social media [1]. Hence, it is therefore easier for the companies to automatically find customer views of a product or services by mining their opinion through social media platforms. The Airline industry is critical about customer reviews. Sentiment analysis introduced by Twitter in 2009 was used to classify the sentiment tweets. It is easier to collect tweets from twitter because it is reliable and can be used for data analysis [2]. Sentiment analysis has the potential to help customers decide which airline is best by analyzing other customer opinions from online comments posted on review or micro-blogging sites.

Sentiment analysis (or sentimental analysis or opinion mining) is a term frequently used to turn up at a dual verdict: users like or dislike something, or the product is good or bad, or someone who is either with or against something. It has been applied across a variety of different domains, such as entertainment, education, automobiles, and also used in NLP, text analysis, and statistics to recognize the 'emotional attitude' pertinent to text into affirmative (positive), unenthusiastic (negative), or impartial (don't-care/neutral) classes [3]. Natural Language Processing is a way for computers to analyze, understand, and derive meaning from human language in a smart and useful way. It is used to analyze text, allowing machines to understand how humans speak. Sentiment analysis and opinion analysis denote the same field of study (which in itself can be considered a sub-area of subjectivity analysis) [4]. They only differ in the assessment of the data that has been mined since the opinion mining focuses on the quality of evaluating the texts and sentiment analysis is more on the polarity of the text, whether it is positive or negative.

Sentiment Analysis field has three different levels: document, sentence, and aspect level. The document level system analysis aims to classify the whole document as positive or negative sentiment. The sentence level system analysis classifies the sentiment expressed in each sentence, while, the aspect level sentiment analysis identifies the sentiment of the specific aspects or features of the entities. This research study is in the sentence level.

This human-computer interaction allows real-life applications like automatic text summarization, sentiment analysis, topic extraction, named entity recognition, parts-of-speech tagging, relationship, extraction, stemming, and more. NLP is commonly used for text mining, machine translation, and automated question answering. Machine learning is an artificial intelligence procedure that replicates/simulates the manner in which the brain of human being functions, aiming to furnish computers with intelligence. It is widely exploited in systems seeking knowledge discovery and application of



data mining referred to as knowledge-based systems and expert systems. Comprehensively revised techniques for machine learning often incorporate artificial neural network (ANN). The support vector machine (SVM), is a new statistical machine learning and data mining tool [5].

## 2. RELATED WORKS

The five competitors for movements in price have hypothesizing effect on customer behavior when presented with a set of itineraries. Daniel et.al. [6] combined data from multiple sources, including competitor data, pricing, social media, safety for airline reviews. Using real airline data for ten different OD-pairs and by means of Extreme Gradient Boosting (EGB), they show that customer behavior can be categorized into price-sensitive, schedule-sensitive and comfort ODs. Through a simulation study, they showed the model produces forecasts that results in higher revenue than traditional time series forecasts.

Nam and Cheol [7] employed service encounters, in which service experiences occur and thus most customer perceptions are created as a set of attributes composed of latent dirichlet allocation (LDA) topics that contain direct perceptions of service experiences. They also investigated statistical correlations between text analytics-based importance performance analysis (TAIPA). Calculations and typical benchmarks of firm's performance in the air-transport industry were used to verify how effective their proposed TAIPA is with respect to the degree that customer satisfaction is represented. As a primary result, TAIPA is more effective than comparison targets in that it shows stronger correlations with firm performance. TAIPA is specialized in determining which service step (i.e., a one-to-one relationship with a service encounter) needs to be improved. Moreover, TAIPA is flexible in considering multiple competitors.

Monika et.al. [8] explored word embedding models (Word2Vec, Glove) in tweets using deep learning methods in detecting sentiment polarity. Here, they investigated sentiment analysis using the Recurrent Neural Network (RNN) model along with Long-Short Term Memory networks (LSTMs) units can deal with long term dependencies by introducing memory in a network model for prediction and visualization. The results showed better significant classification accuracy trained 80% for training set and 20% for testing set which shows that our models are reliable for future prediction.

Kwon et.al. [9] conducted topic modeling and sentiment analysis on the posts of Skytrax (airlinequality.com), where there are many interests and participation of the people who have used or are willing to use it for airlines. The data collected were online reviews from 27 airlines, with more than 14,000 reviews. Topic modeling and sentiment analysis were used with the collected data to figure out what kinds of important words are in the online reviews. As a result of the topic modeling, 'seat', 'service', and 'meal' were significant issues in the flight through frequency analysis. Additionally, the result revealed that delay was the main issue, which can affect customer dissatisfaction while 'staff service' can make customers satisfied through sentiment analysis as the result shows the 'staff service' with meal and food in the topic modeling.

Rana [10] conducted experiments on the US Airline dataset with six classification methods to predict customer sentiment prediction on tweets. They adopted two machine learning algorithms, which are Logistic Regression (LR), Naïve Bayes (NB), and four Deep Learning algorithms such as Convolutional Neural Networks (CNN), BERT, XLNET, and. The Machine learning algorithms were applied using different feature extraction approaches such as Bag-of Word (Bigram, Trigram), TF-IDF with (Bigram, Trigram), Spicy with (Bigram, Trigram), Word2Vec. The best results of both models were with Spacy and Trigram. In addition, with Deep learning algorithms BERT and ALBERT are applied with binary classes and multiple classes. They then used different evaluation measurements such as accuracy, precision, recall, F1 score, and log loss to demonstrate the effectiveness of the model.

Jane et.al. [11] classified a tweet's polarity through annotation using classifiers Naïve Bayesian, Support Vector Machine and Random Forest to develop a model. The researchers gathered tweets of local airlines concerning their experience in the services provided by local airlines in the Philippines. The researchers also determined the sentiment of a tweet if it is positive, neutral or negative and provided the quantitative and qualitative analyses, as well as opinion analysis, to better understand the result of the experiment.

Sahu et.al. [12] enhanced the pre-processing by applying spell checking algorithm. Instead of classifying tweets as positive and negative, they also rank the tweets by the degree of positivity or negativity thus, a scoring system was introduced. The classifiers they used are SVM and Maximum Entropy. They gathered 60,000 tweets which in the pre-processing were filtered by removing external website links, username, RT for retweets, stopwords and also, they replaced emoticons with words. At the end of their study, the 60,000 tweets achieved an accuracy of 74.2% and they concluded that pre-processing is important in short opinionated text like tweets as it increased the classifying accuracy.

Mamgain et.al. [13] applied Naïve Bayes, SVM and an Artificial Neural Network Model in their experiment. Pre-processing removed URLs and stopwords was used in removing words which are common words and do not contribute to the sentiment of the sentence. The result of the study show that the All-India Institute of Medical Sciences had the highest average positive sentiment with 4.56 unlike IIT with 2.93 and NIT with 2.94. Also, it showed that the multilayer perceptron matches or surpasses the performance of Naïve Bayes and SVM. The following accuracy for multilayer



perceptron are 92.6%, 89.6% and 87.6%, for Naïve Bayes are 90%, 90% and 83% and lastly, SVM with 91%, 83% and 81%.

Hakh et. al., [14] applied the SMOTE method to solve the imbalanced challenge of the datasets and analyze a collection of tweets about six airline companies found in the US using machine learning techniques. They found that the feature selection and over-sampling techniques are equally essential to achieve refined results. Then they applied the sentiment classification (i.e., AdaBoost, Decision Tree, Linear SVM, Naïve Bayes, Random Forest, K- NN, and Kernel SVM). Das et.al. [15] used 200 tweets directed at Emirates and Jet Airways and analyzed airline Twitter data using the Naïve Bayes algorithm for sentiment analysis. They used R and Rapid Miner tools to improve the classification model and map the tweets into the positive, negative, and neutral categories. They posited that the outcomes achieved using Naïve Bayes classifiers were promising for a more significant number of tweets in the analysis.

3. METHODOLOGY

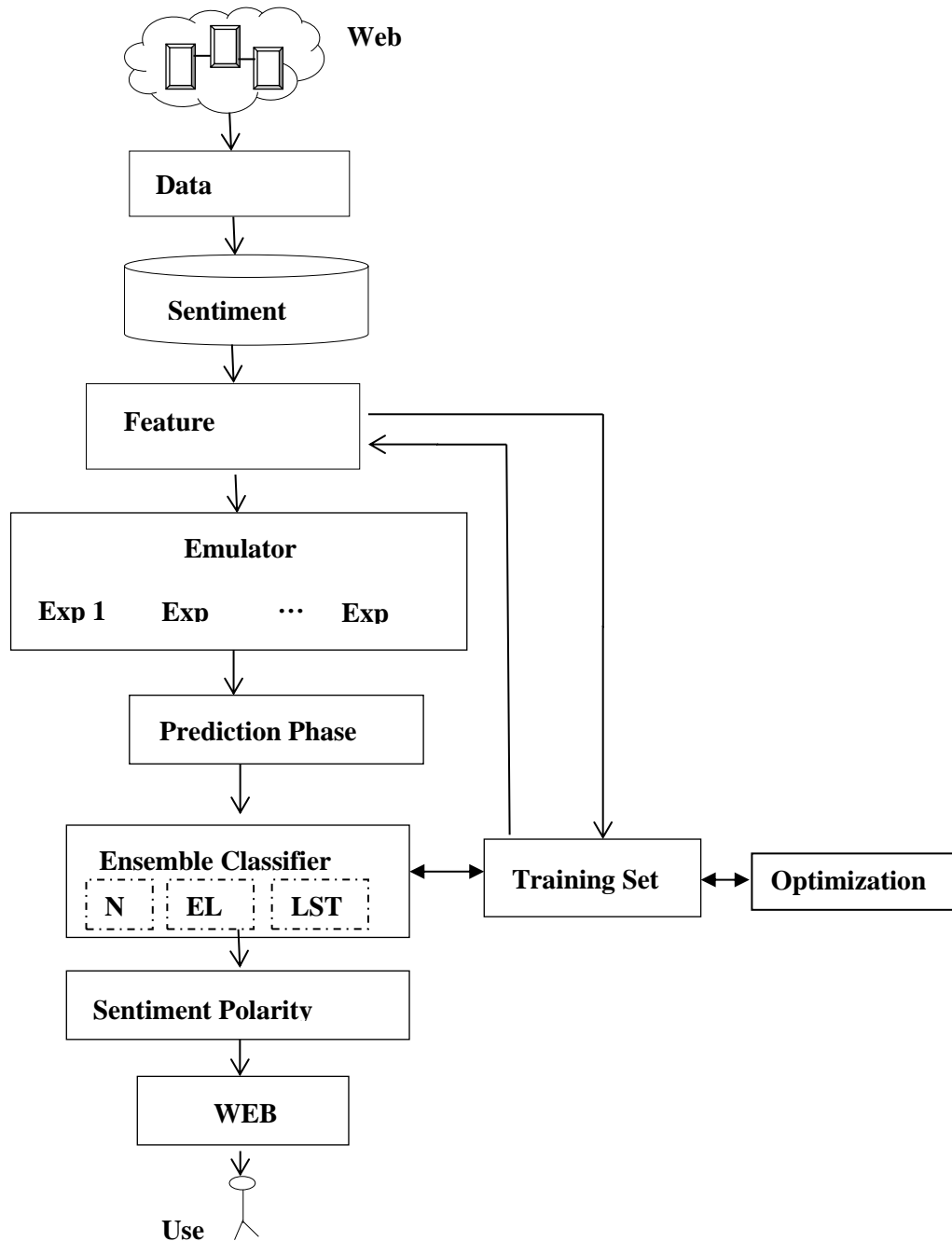


Figure 1: Architecture of the Proposed System



The data collected is unstructured format and several techniques will be employed to transform it to structured data within the feature generation/preprocessing part. The feature identification has the feature vector list that allows each feature (attribute) to be weighted. Each sentence is separated, words separated, special characters and symbols if presents are removed. This is the stage called data preprocessing. It involves the cleaning of data by performing spell correction, punctuation handling, stemming etc. This means that all irrelevant features which might interfere with the performance of the classifier to be removed and removal of unnecessary repetitions so that the classifier is trained on clean data thus ensuring optimum performance.

The feature generation uses many approaches one of which is the N-grams. N-gram uses combination of words as features. For instance, the combination “beautiful” might be a unigram for positive opinions, “I like” would be a bigram for positive opinions, and “I don't like” could be considered a trigram that hides a negative opinion. Bigrams and trigrams enabled us to capture information about adjectives and adverbs and requires substantial amount of data. Feature extraction is the process where properties are extracted from the data. These properties, called features are characteristics from the data, because in practice the whole input data is too large to use in classification. The features should be discriminative, to describe the original data and the features should reduce the space to prevent redundancies and a high dimensionality of the data. This is then closely followed by feature selection where all useful attributes are carefully selected after proper extraction is done. It eliminates feature with little or no predictive information so as to significantly improve the comprehensibility of resulting classifier models. It aims to find a feature subset that produces higher classification accuracy. It is necessary so as to avoid noisy classifier or low rates of accuracy. The emulator uses a mixture of experts' model inside a network to infer (reasoning process) or emulate the decision-making ability of a human expert in sentiment analysis. This is necessary for software driven by user interface requirements. Furthermore, machine learning algorithms such as Extreme Learning Machine, Naïve Bayes algorithm and Long Short-Term Memory are employed for the training phase. The Naïve Bayes classifier is trained to come up with a model that can be used to classify the sentiments from sentences/documents. The Extreme Learning Machine is a feedforward network. Its learning process needs only a single iteration as against the support vector machine that requires multiple iterations. It is a network learning algorithm that overcomes the slow training speed and over-fitting problems. The algorithm works for the generalized single-hidden-layer feed-forward and enables better generalization ability, robustness, and controllability and fast learning rate.

#### 4. DISCUSSION OF RESULT

This paper presents an airline reservation system using sentimental analysis. The system starts by acquiring an airline dataset. The dataset is made up of 16 columns ranging from unit id to the user time zone. The dataset was pre-processed by removing missing and duplicate values. After the pre-processing, we applied feature extraction in selecting relevant features. We reduced the 16 features of the airline dataset to two features. The reduced features can be seen in figure 2. The new features are the text features and the sentiment analysis features. The text features comprise of some comments made by users, while the sentiment analysis features consist of three classes, which are, negative, neutral and positive. This can be seen in figure 3. After selecting the relevant features that will be used for training, we applied data cleaning to the text features in other repeated words. We used the stopwords in eliminating words that are so commonly used that they carry little information. After applying stopwords to the text columns, we then split the dataset into x and y variables. The x variable holds the cleaned text features and the y variable holds the sentiment\_analysis feature. We then applied ConutVectorizer function to the x variable that holds the text features in converting the written comments made by the users into an array of zeros (0s) and ones (1s) so as to make it suitable for training. We also used LabelEncoder function in converting the y variable, which holds the sentiment analysis columns into numbers ranging from 0 to 2, where -1 represents a neutral comment, 1 represents positive and 0 represents negative. After these processes, we then assigned 70% of our data for training and 30% of our data for testing. We then assigned the training data to our Naïve bayes classifier. After training, we obtain a training accuracy of about 95% and a testing accuracy of about 79%. The classification report of the training accuracy can be seen in figure 4 and the confusion matrix evaluation can also be seen in figure 5. The model was then saved and deployed to web using python flask for easy execution and testing. This can be seen in figure 6 and 7, 8 and 9.



	airline_sentiment	text
0	neutral	@JetBlue's new CEO seeks the right balance to ...
1	negative	@JetBlue is REALLY getting on my nerves !! _Ù+...
2	negative	@united yes. We waited in line for almost an h...
3	negative	@united the we got into the gate at IAH on tim...
4	negative	@SouthwestAir its cool that my bags take a bit...

Figure 2: Refined Dataset

This shows the new dataset that we derived from the original dataset by means of feature extraction

[Text(0, 0, 'neutral'), Text(1, 0, 'negative'), Text(2, 0, 'positive')]

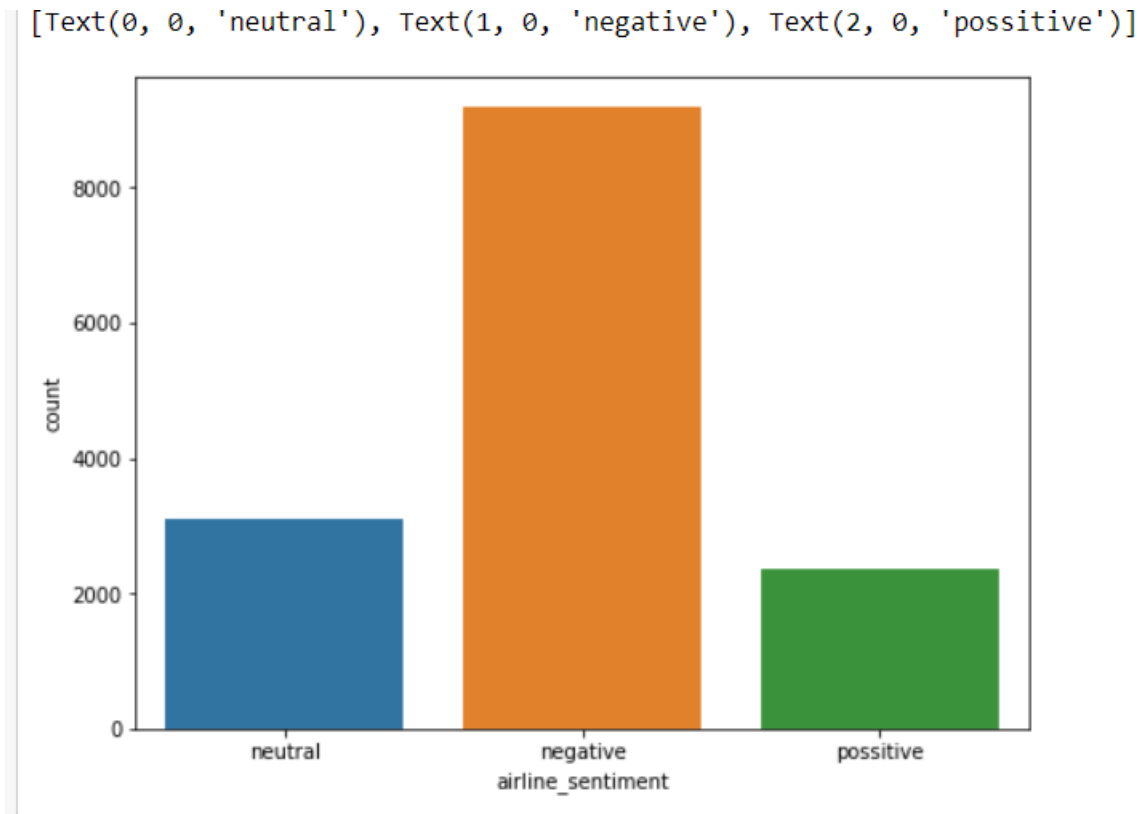


Figure 3: A histogram showing Airline\_sentiment feature

This shows a count plot of the sentiment feature, shows the total number of negative, positive and neutral responds gotten from various individuals



	precision	recall	f1-score	support
-1	0.96	0.98	0.97	6679
0	0.92	0.87	0.90	2231
1	0.95	0.92	0.93	1781
accuracy			0.95	10691
macro avg	0.94	0.93	0.93	10691
weighted avg	0.95	0.95	0.95	10691

Figure 4: Classification Report

This shows a classification report of the trained Naïve Bayes model. Here we obtain an accuracy of about 95%.

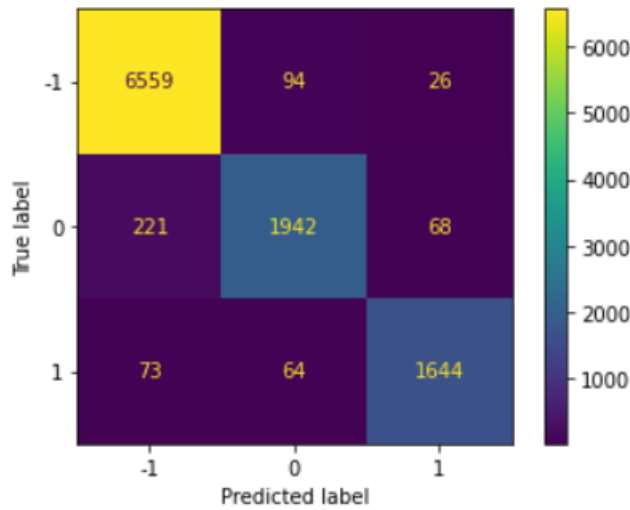


Figure 5 shows the Confusion matrix of our trained model



Figure 6: Home page of the proposed system

Here, users can analyze various comments of other users to see their conclusion about the airline service





SENTIMENT ANALYSIS ON AIRLINE SRESERVATION



USERS'S COMMENTS

This airline, sometimes the booking is fast and sometimes the booking is slow

Analyze

Result

The returned result is a neutral one

Figure 7: Showing a Neutral Comment

Here the analyzed comment happens to be neutral one. It is more like an average

SENTIMENT ANALYSIS ON AIRLINE SRESERVATION



USERS'S COMMENTS

This airline is something elsej

Analyze

Result

The returned result is a negative one

Figure 8: Showing a Negative Comment

Here the analyzed comment happens to be negative one. This means it gives a bad view of airline reservation in that company's airline service



## SENTIMENT ANALYSIS ON AIRLINE SRESERVATION

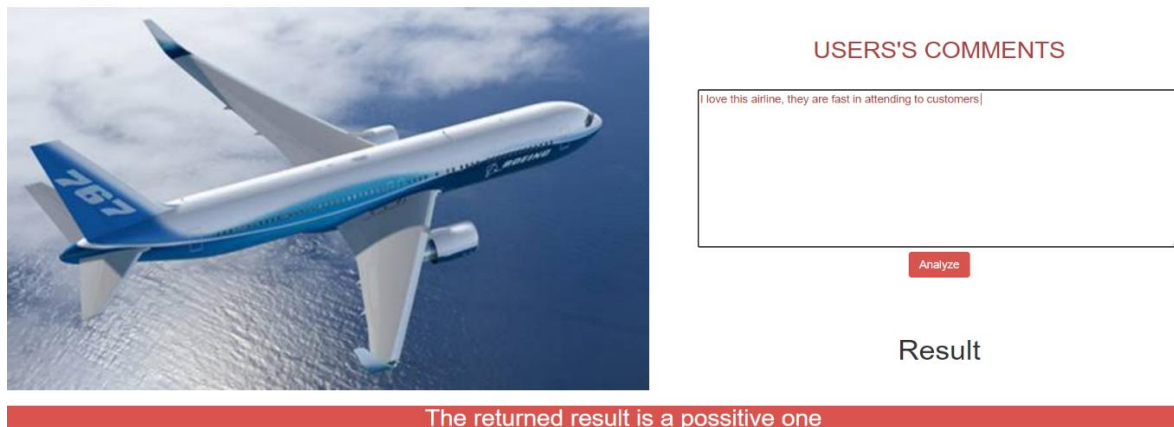


Figure 9: Showing a positive Comment

Here the analyzed comment happens to be positive one. This means it gives a good view of airline reservation in that company's airline service

## 5. CONCLUSION

This paper presents an airline reservation system using sentimental analysis with Naïve bayes classifier. The system uses an airline dataset, which contains 16 columns, but we selected just few columns of the features by means of feature extraction. The selected columns are the text column and the airline sentiment column. stopwords was used in eliminating words that are so commonly used that they carry little information. We also used LabelEncoder function in converting the y variable, which holds the sentiment analysis columns into numbers ranging from 0 to 2, where -1 represents a neutral comment, 1 represents positive and 0 represents negative. After these processes, we then assigned 70% of our data for training and 30% of our data for testing. We then assigned the training data to our Naïve bayes classifier. After training, we obtain a training accuracy of about 95% and a testing accuracy of about 79%. The Naïve bayes classifier was then saved and deployed to web using python flask. This paper can further be extended by creating an android application in analyzing user's behaviour on airline booking or reservation.

## REFERENCES

- [1]. H. P. Patil and M. Atique, "Sentiment analysis for social media: A survey," 2015 IEEE 2nd Int. Conf. Information Science Security. ICISS 2015, 2016
- [2]. T. K. Shivaprasad and J. Shetty, "Sentiment analysis of product reviews: A review," Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2017, no. Iccict, pp. 298–303, 2017
- [3]. S. Kitaoka and T. Hasuike, "Where is safe: Analyzing the relationship between the area and emotion using Twitter data," 2017 IEEE Symp. Ser. Comput. Intell. SSCI 2017 - Proc., vol. 2018–January, pp. 1–8(2018).
- [4]. S. P. Algur and R. H. Patil, "Sentiment analysis by identifying the speaker's polarity in Twitter data," Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT 2017, vol. 2018–January, no. 2010, pp. 898–902(2018).
- [5]. Chen, L.; Li, Y.Q.; Liu, C.H. How airline service quality determines the quantity of repurchase intention-Mediate and moderate effects of brand quality and perceived value. J. Air Trans. Manag. 2019, 75, 185–197.
- [6]. H. Daniel, G. Koole, R. van, "A machine learning approach to itinerary-level booking prediction in competitive airline markets", arXiv:2103.08405v1 [math.OA] 15 Mar 2021.
- [7]. S. Nam, H. Cheol, "A Text Analytics-Based Importance Performance Analysis and Its Application to Airline Service", Journal of Sustainability, vol. 11, 1–24, 2019.
- [8]. R. Monika, S. Deivalakshmi, B. Janet, "Sentiment Analysis of US Airlines Tweets using LSTM/RNN", 9th International Conference on Advanced Computing (IACC), 92–95.
- [9]. H. Kwon, H. Ban, J. Jun, H. Kim, "Topic Modeling and Sentiment Analysis of Online Review for Airlines", Journal of information 12(21), 1–14, 2021.
- [10]. A. Rana, P. Nourah "Predict Sentiment of Airline Tweets Using ML Models", EasyChair, 2021.
- [11]. M. Jane, J. Michael, V. Jade, S. Celine, "Sentiment and Opinion Analysis on Twitter about Local Airlines", Proceeding from the 3<sup>rd</sup> International Conference on Communication and Information Processing, 415–422, 2017.
- [12]. S. Sahu, S. Rout, D. Mohanty. Twitter Sentiment Analysis -- A More Enhanced Way of Classification and Scoring. In Proceedings of the 2015 IEEE International Symposium on Nanoelectronic and Information Systems.
- [13]. N., Mangain, E. Mehta, A. Mittal, G. Bhatt. Sentiment analysis of top colleges in India using Twitter data. In Proceedings of the 2016 International Conference on Computational Techniques in Information and Communication Technologies, 2016.
- [14]. Hakh, H.; Aljarah, I.; Al-Shboul, B. Online social media-based sentiment analysis for us airline companies. In Proceedings of the New Trends in Information Technology, Amman, Jordan, 25–27 April 2017; pp. 176–181.
- [15]. D. D. Das, S. Sharma, S. Natani, N. Khare, and B. Singh, "Sentimental Analysis for Airline Twitter data," in IOP Conf. Series: Materials Science and Engineering, 2017.