IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 1, January 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0901012 |

An Analysis of Customer Satisfaction from airline Tweets Using Machine Learning Approach

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

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

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

ABSTRACT: Today feedback is the foremost required thing to improve any kind of service in every field. Social media, Hashtags, tags, etc. are the best way to do any kind of feedback suggestions or anything because it creates a social impact on the social life of people. Today if any of us want to use any new services which anyone haven't used yet, in that case, we simply go to Google, Facebook, and stuff to check the review of the people over that thing. In between all of these social media, the most important source of feedback is Twitter where tweets play the main role over the social image of that page. Positive, Negative, and neutral tweets over something plays a very important role in someone's social image especially the big companies. Now when we talk about using tweets to find their polarity one thing comes into action that is sentimental analysis. Sentimental analysis is a way to find the polarity of the text in terms of emotion, feeling, or any other factor. About tweets, the sentimental analysis is used to find whether the tweet is positive, negative, or neutral. By this, we are exactly extracting opinions of different people over that kind of service. In this, we are designing a model for sentimental analysis specifically in terms of airline services. The proposed model will exactly extract the opinion from the feedback of different airline services. There are many datasets available of different collections of tweets. We will be choosing the official dataset from Kaggle to design and test the model of sentimental analysis. Today the airline industry is trending and a rapidly growing field. So by implementing such a model we can even help both parties the public and the large companies to resolve and make things efficient. Even these companies use a feedback system with the customer, the filtration of reviews in different classes is a big task there. Even in that case, we can use the proposed model to separate the feedbacks depending upon the opinion extraction of the feedback. Before Machine learning models this process is very lengthy and time consuming. The analysis is carried out with many different kinds of classifiers available and picked the best one out. The proposed model is capable of giving more than 90% accuracy in analysing the sentiment of the feedback or tweets.

KEYWORDS: Machine learning, Tweets, Airlines, Customer satisfaction, Feedback, Logistic regression, Sentiment analysis.

I. INTRODUCTION

Customer feedback is very crucial to Airline companies as this helps them in improving the quality of services and facilities provided to the customers. Sentiment Analysis in Airline industry is methodically done using traditional feedback methods that involve customer satisfaction questionnaires and forms. These procedures might seem quite simple on an overview but are very time consuming and require a lot of manpower that comes with a cost in analyzing them. Moreover, the information collected from the questionnaires is often inaccurate and inconsistent. This may be because not all customers take these feedbacks seriously and may fill in irrelevant details which result in noisy data for sentiment analysis. Whereas on the other hand, Twitter is a gold mine of data with over 1/60th of the world's population using it which nearly amounts to 100 million people, more than half a billion tweets are tweeted daily and the number keeps growing with every passing day. With the rising demand and advancements of Big Data technologies in the past decade, it has become easier to collect tweets and apply data analysis techniques on them. Twitter is a much more reliable source of data as the users tweet their genuine feelings and feedbacks thus making it more suitable for investigation. For example, with the iPhone X market release, the company can perform a sentiment analysis on the tweets related to the product as a part of their market research to improvise their product. Once the airline tweets are collected, they undergo pre-processing to remove unnecessary details in them. Sentiment classification techniques are then applied to the cleaned tweets. This gives data scientists and Airline companies a broader perspective about the feelings and opinions of their customers. The main motive of this paper is to provide the airline industry a more comprehensive view about the sentiments of their customers and provide to their needs in all good ways possible. In

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0901012 |

this paper, we go through several tweet pre-processing techniques followed by the application of seven different machine learning classification algorithms that are used to determine the sentiment within the tweets. The classifiers are then compared against each other for their accuracies.[2]

The world is getting more and more interconnected both economically and socially thanks to the recent advancements in the Internet and networking technologies. There has been a remarkable rise in the last few years in the number of Internet users. Around 4.1 billion people have access to the Internet representing about half of the global population and this number has been growing at an enormous rate since the year 2000, primarily due to the smartphone revolution in the telecom sector . People create their own content, share videos, images or repost other user's content. This content is related to the user's social activities and personal experiences about different services, products and events. Social networks are helping businesses grow and make profits around the globe. It helps them advertise their upcoming services and products at much lower cost compared to traditional marketing models and maintain a close relationship with customers. Upcoming festivals, music albums, movies, sporting events and new product releases create a buzz among the general public much ahead of the release and this can be gauged through social network responses related to them. Thus, social networks are finding their way in business models as they are helpful in recognizing new opportunities and threats.[1] At present, large scale companies are investing plenty of time, resources and energy to enhance the consumer's loyalty. It may explore more opportunities for the interaction between companies and consumers to get their feedback and suggestion about the products and services with an aspect of customer satisfaction and product quality improvement. This may increase the both the economic and social development of the company. A crucial but challenging step is to automatically analyze the customer feedback by extracting useful information from the huge data of customer feedbacks. Customer feedback data is very important in addressing several issues and sentiment and opinion analysis is one of the important issue among them. Extracted patterns from the data may be utilized by company experts to understand the polarity of the opinion towards different products and services. In general the polarity of opinion may be positive, negative or neutral. Companies may use this polarity of opinions in order to improve their quality of products and/or services. Sentiment analysis/opinion mining assists in answering different question about products and services by understanding the emotions in the feedbacks. Present world is utilizing the natural language processing (NLP) and text classification techniques to map the sentiments within the text into positive, negative and neutral classes. The sentiments can be seen as an indirect publicity of a company's products and services in the world that provide a direct impact on other customer's. For travelers, the most popular and convenient platform for sharing their opinion is Twitter. Each travel journey on different carriers may bring different comfort levels i.e. good, average or poor level of comfort. These comfort levels are conveyed to the social media i.e. Twitter etc. by the travelers in terms of tweets. If a traveler enjoyed the trip, the respective tweet would demonstrate the happiness or positive emotions towards the travel carrier otherwise negative emotions may be conveyed. It is a furious tweet by a passenger on British Airways flight. As a result, the company considered it very urgent and important and settled the issue at the earliest. In another, a sarcasm tweet for Indigo Airways was fired because the baggage of passenger was transferred to a different location (Hyderabad) other than the traveler (Calcutta). The tweet seems to be negative from a human perspective whereas it is difficult to put this into negative class for the machine because of the complex words used in the tweet. Also, tweets on tweets may not contain more than 140 characters at once. Therefore, it is useless to expect the detailed information inside the tweet. However, a general understanding about polarity of emotions can be developed using machine learning methods. Further, tweets in the categories may be analyzed to get insights or possible reasons for these sentiments. Every day more than a million of people are travelling around the world and tweeting their views with respect to the journey. It results in a huge amount of data available for analysis every day. Hence, machine learning techniques can be considered as a solution for such analysis. Machine learning techniques are efficient to handle huge data with large dimensions. [3]

II. SENTIMENT ANALYSIS

Sentimental Analysis [4] is a method of opinion mining to extract information about peoples views, opinions, sentiments towards an everyday happening things. And an each individual have a different opinions on same topic. The sentiment analysis task is technically more challenging but practically more useful. For ex- ample, Businessmen always want to know about the public opinion regarding that products and feedback from different customers. The customers also wants to know the rating of that product which has given by other customers who had pur- chased earlier, and marketers also prefer sentiment analysis because they wanted to know the targeted customers. With the major development in social networking (i.e., Facebook, Twitter, LinkedIn, Stumble upon etc.,) on the Web, individuals and large associations are concentrating on publics opinion for their decision making. The task of min- ing opinion information on web sites is not easy; one because of vast number of websites currently present and still populating and second because of lack of standardized methodology to do the same. Moreover, the text corpora present on websites constitute both useless and useful data that will be required for our analysis. There is always a thin line between these

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0901012 |

kinds of data which always add unnecessary overhead in analysis. The normal human reader will experience issues distinguishing relevant sites and also summarizing information and opinions in them. Additionally, it is likewise realized that human analysis of content data is liable to significant preferences, e.g., peoples regularly give more priority to opinions that are reliable with their own preferences. There are other factors as well like human mental capacity and physical limitation that make humans inept to analyze large amount of data. Thus an automated opinion mining is required which will eventually help humans in sentiment analysis.[4]

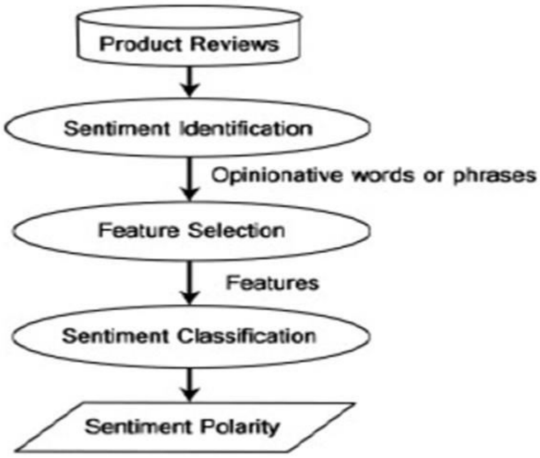

Fig.1 Sentiment Analysis

III. RELATED WORK

T. Hemakala1 and S. Santhoshkumar (2018) formulated a framework for emotional analysis by asking ideas for aviation service response. Many of the available datasets for hotel-led hotel reviews offer a wide range of researchers' work in relation to pre-data processing work. Twitter is an SNS with great data on user submissions, with this amount of data, has research capabilities related to text mining and can be emotionally analyzed. Airlines are turning to traditional customer feedback forms which are also very tedious and time consuming. In this work, a database containing the tweets of six major Indian Airlines airlines was used and various emotional analyzes were performed. This approach starts with the pre-processing techniques used to refine the tweets and then represents these tweets as vectors using a deeper learning concept to perform sentence level analysis. The analysis was performed using 7 different phase strategies: Decision Decision, Random Forest, SVM, K-Nearest Neighbors, Logistic Regression, Gaussian Naïve Bayes and AdaBoost. The effect of the tweet sensor test set [21] **Guoning Hu et al. (2017)** analyzed the views of 19M Twitter users on 62 popular industries, including 12,898 businesses and consumer products, as well as related news headlines, with a survey of 330M tweets in a month's time. Good in building and producing more service industries In addition, they tend to be better at communicating with products than general on Twitter. They also found that the sentiments against branded within in the industry are very different and indicate that we use two industries as consumer cases. Moreover, you find that there is no strong correlation between the views of the various industries, indicating that the sentiments of the subject depend largely on the context of the field in which they are mentioned. They highlighted the importance of such analysis in assessing the impact of products on social media. Hopefully this initial research will prove useful to researchers and companies in understanding consumer understanding of related industries, products and topics and encourage further research in this field.[22]

Prayag Tiwari et al. (2019) proposed a machine-readable method of distinguishing passenger tweets in relation to aviation services to understand the emotional pattern. We welcome Random Forest (RF) and Logistic Regression (LR) to divide each tweet into positive, negative and neutral impressions. Analysis of the actual data collected shows that these two methods are able to achieve ≈80% accuracy. [24] **Janet R. McColl-Kennedy et al. (2018)** applied to after-sales sales in business-to-business settings, the authors contribute to the teaching and practice of customer experience (CX) in three important ways. First, by providing a CX novel conceptual framework that combines previous CX

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0901012 |

research to better understand, manage, and improve CXs - which contain quantitative elements (resources, tasks, context, interactions, and customer roles), psychological responses, and different emotions in places touches across the customer journey. Second, by demonstrating the use of longitudinal CX analytic in terms of a conceptual framework that integrates measurement and measurement methods. Thirdly, by providing a step-by-step guide to how to use the text in a non-functional way, it has thus shown that CX analytics using big data strategies in CX can provide important information that is important. The authors emphasize six physicians who need understanding to manage their clients 'travels, (1) by taking a client by observation, (2) identifying causes, (3) identifying dangerous components, (4) capturing clients' emotional and psychological responses, (6) and prioritization of actions to improve CX. The article concludes with a withdrawal from future research. [28]

IV. MATERIALS AND METHOD

4.1 Dataset Description

In this work, the Kaggle dataset is used, which comprises tweets for airlines of the United States (US). The "Twitter US Airline Sentiment" is name of dataset which contains 14K+ tweets done to give feedback to airline services of the US in 2015. The dataset has Rows count-14641, Column count-15 and the column names: tweet_id, airline_sentiment, airline, sentiment_confidence, negativereason, negativereason_confidence, airline, airline_sentiment_gold, name , negativereason_gold, retweet_count, text, tweet_coord, tweet_created, tweet_location, user_timezone,.
After pre Processing of dataset, we are using: airline_sentiment, negativereason and text.
Only these three columns are used. Two of them are used as features and one is used as the label in the Machine learning Problem.

4.2 Methodology

Logistic Regression is a mathematical model used in statistics to estimate (guess) the probability of an event occurring using some previous data. Logistic regression is a classification algorithm used to assign observations to a discrete set of classes. Unlike linear regression which outputs continuous number values, logistic regression transforms its output using the logistic sigmoid function to return a probability value which can then be mapped to two or more discrete classes.[109]

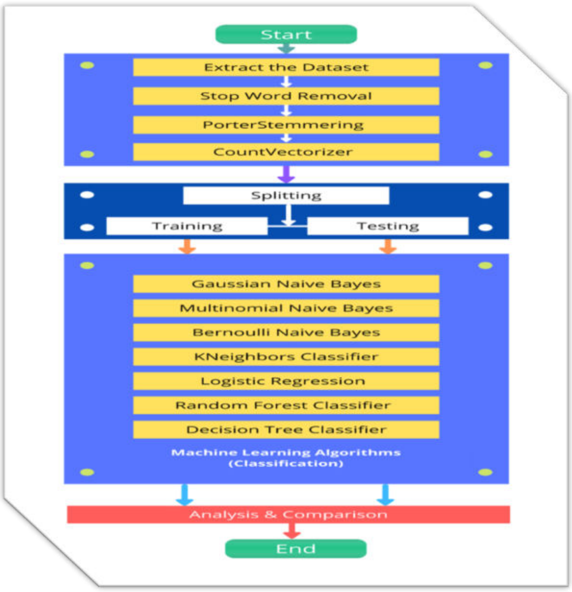

Fig. 4.1: Overall architecture of Proposed work

The above is the detailed architecture of the process the basic architecture of the layout can be seen as:

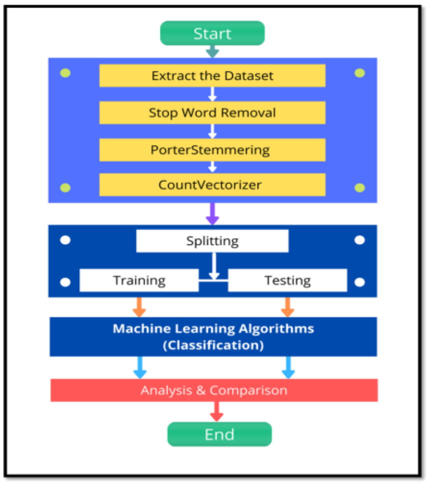

Fig. 4.2: Process architecture of proposed work

The Architecture is divided into following steps

1. Web searching of Dataset

The very first step of any kind of machine learning problem is to look for a raw dataset to work on. Sentimental Analysis is a technique that needs two segments of any data one is the text and the other is the opinion or the required opinion. For the proposed system we need a lot of tweets in the dataset whose polarity may vary in Negative/Neutral/Positive. The dataset which we will go to use to implement the solution must be authentic or collected by any official institute. We looked the web for database and Kaggle the official store of databases had a database similar to our requirements so we picked that dataset up and now we have our raw dataset to work on. Now the next step will be extraction or pre-processing of the dataset.

Output of the step:

- ➔ We have “Tweets.csv” as our raw dataset.

2. Extracting of Dataset

As the raw dataset has 15 columns, the dataset is detailed as the institute which designed the dataset. The Proposed model has its own set of requirements, so the dataset needs to be filtered out as per the requirements. Sentimental analysis usually needs two different segments of the dataset one is the feature or the text which needs to be analysed another one is the label or the opinion, these two segments will be used to train the machine learning model. The raw dataset has two columns that have the text of the tweets so we will pick those two columns as the features. The column which has the opinion will be picked as the label. So we need only those three columns from the raw dataset.

3. Stop Word Removal

Stop word removal is the process of removing the words which don't have any role in opinion mining or which don't have any opinion. Like the, i, an, in, etc. Such words are known as stop words which don't have any kind of opinion. So we simply remove all the stop words from all the sentences from the whole database.

The list of stop word in English is like:

‘ourselves’, ‘hers’, ‘between’, ‘yourself’, ‘but’, ‘again’, ‘there’, ‘about’, ‘once’, ‘during’, ‘out’, ‘very’, ‘having’, ‘with’, ‘they’, ‘own’, ‘an’, ‘be’, ‘some’, ‘for’, ‘do’, ‘its’, ‘yours’, ‘such’, ‘into’, ‘of’, ‘most’, ‘itself’, ‘other’, ‘off’, ‘is’, ‘s’, ‘am’, ‘or’, ‘who’, ‘as’, ‘from’, ‘him’, ‘each’, ‘the’, ‘themselves’, ‘until’, ‘below’, ‘are’, ‘we’, ‘these’, ‘your’, ‘his’, ‘through’, ‘don’, ‘nor’, ‘me’, ‘were’, ‘her’, ‘more’, ‘himself’, ‘this’, ‘down’, ‘should’, ‘our’, ‘their’, ‘while’, ‘above’, ‘both’, ‘up’, ‘to’, ‘ours’, ‘had’, ‘she’, ‘all’, ‘no’, ‘when’, ‘at’, ‘any’, ‘before’, ‘them’, ‘same’, ‘and’, ‘been’, ‘have’, ‘in’, ‘will’, ‘on’, ‘does’, ‘yourselves’, ‘then’, ‘that’, ‘because’, ‘what’, ‘over’, ‘why’, ‘so’, ‘can’, ‘did’, ‘not’, ‘now’, ‘under’, ‘he’, ‘you’, ‘herself’, ‘has’, ‘just’, ‘where’, ‘too’, ‘only’, ‘myself’, ‘which’, ‘those’, ‘i’, ‘after’, ‘few’, ‘whom’, ‘t’, ‘being’, ‘if’, ‘theirs’, ‘my’, ‘against’, ‘a’, ‘by’, ‘doing’, ‘it’, ‘how’, ‘further’, ‘was’, ‘here’, ‘than’.

Input:

- ➔ I am a Powerful Man.
- ➔ Power is good.

Output:

- ➔ Powerful Man.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |
|| Volume 9, Issue 1, January 2021 ||
| DOI: 10.15680/IJIRCCE.2021.0901012 |

➔ Power good.

4. Porter Stemming

It a process of checking the base words for all kinds of words. All the words in a sentence must not be in higher variants of words if it is so the word needs to be replaced by the root word of the same. As in sentimental analysis words like "Like" and "Likes" have the same opinion so the words used must be root words only.

Input:

➔ Powerful Man.

Output:

➔ Power Man.

5. Count Vectorizer

Now we have the final text data don't we need to transform it into some other data form to make it a suitable input for a classifier? We need to encode the text data into integer data which process is known as vectorization or it is also known as feature extraction. This process is carried out to transform the text data into data that can give a good score in machine learning problems. It works like the counter of the word, you can see in the example below.

Input Sentence:

➔ Man Power.

➔ Power Good.

Output

Man	Power	Good
1	1	0
0	1	1

6. Splitting of Dataset

As the raw dataset is filtered into an efficient dataset which is as per the requirements of the system. Now when the Machine learning model is trained we need two different segments of the dataset Training and testing. When supervised machine learning is taken into action we need a dataset to train the model and have to use a part of it to test the model too. So here we are using 10% of the same dataset as the testing segment and the rest 90% as the training dataset. So in this step, we have Training And testing dataset.

7. Train and Test different classifiers

The detailed model of this step is like, here different classifiers is used

Fig. 4.3: Machine learning model architecture

Basically, in this step, we have used the training partition of the dataset to train different models and then tested them with the help of the test partition of the dataset. As the Filtered dataset is divided into two segments the Training and the testing. In both of them, they have two internal segments of the dataset as the features and the labels. Features are the input of any model and the label is the expected output. While training Both the features and label is given as input at the time of testing only testing features should be given and the Expected label matched with the actual labels.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0901012 |

There following are the Classifiers used in this model for train and testing purpose.

A. Gaussian NB
B. Multinomial NB
C. Bernoulli NB
D. K Neighbours Classifier
E. Logistic Regression
F. Random Forest Classifier
G. Decision Tree Classifier

8. Analysis and Comparison of All Classifiers

As in the implementation part, the entire models which are trained in the previous step will be compared on the basis of different factors; the following are the factors which are calculated for each model then all those factors are used to compare the models. The comparison is easy when it is carried out in pictorial representation. So, we use the graphical representation of all those five factors to compare all the models. The five factors used to compare the model are :

A. Score
B. Accuracy
C. Precision
D. F1 Score
E. Recall Score

V. RESULTS AND ANALYSIS

In this section of the paper, we perform the result analysis on different measuring parameters like score, accuracy, precision, recall, f1-measure, and comparison is done between the proposed methodology (logistic regression), Gaussian NB, Multinomial NB, Bernoulli NB, K-neighbors Classifier, Random Forest Classifier and Decision Tree Classifier.

5.1 Score Analysis

The score parameter is used to prove the qualification on various machine learning techniques and the comparative analysis of this parameter is done among different machine learning such as random forest, Gaussian NB, Multinomial NB, Bernoulli NB, Kneighbors Classifier, and Decision Tree Classifier and our proposed method (logistic regression). The simulation results of our proposed method and existing method is shown in table 5.1 and it is 94% which is much more about the other exiting approach. The analysis is done using the comparison graph shown in figure 5.1 and it is found that our proposed method has higher value than the others. It means that the proposed method is more success in the prediction of movie success or hit.

Table 5.1: Comparative analysis of score parameter between Logistic regression and existing method

S. No.	Model Name	Score
1	Gaussian NB	0.84
2	Multinomial NB	0.86
3	Bernoulli NB	0.93
4	Kneighbors Classifier	0.87
5	Logistic Regression	0.94
6	Random Forest Classifier	0.92
7	Decision Tree Classifier	0.91

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0901012 |

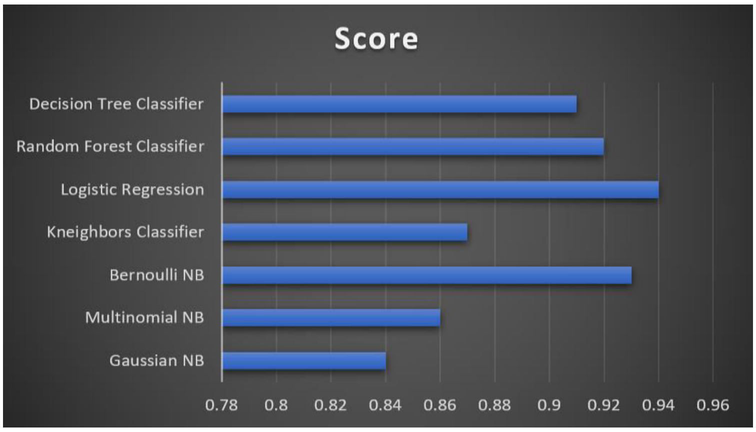

Fig. 5.1 : Analysis of Score parameters

5.2 Accuracy Analysis

This section presents the comparison of accuracy parameter to show the accuracy of customer satisfaction for airline tweets using sentiment analysis of this parameter is done among different machine learning such as random forest, Gaussian NB, Multinomial NB, Bernoulli NB, Kneighbors Classifier, and Decision Tree Classifier and our proposed method (Logistic Regression). The simulation results of our proposed method and existing method is shown in table 5.2 and it is 94% which is much more about the other exiting approach. The analysis is done using the comparison graph shown in figure 5.2 and it is found that our proposed method has higher accuracy value than the others. In this the value of accuracy is equivalent to score parameter. If score of the customer satisfaction will high accuracy of the movie prediction will high. And it is analyzed that the proposed method is more success in the accuracy analysis of customer satisfaction.

Table 5.2: Comparative analysis of accuracy parameter between Logistic Regression and existing method

S. No.	Model Name	Accuracy
1	Gaussian NB	0.84
2	Multinomial NB	0.86
3	Bernoulli NB	0.93
4	Kneighbors Classifier	0.87
5	Logistic Regression	0.94
6	Random Forest Classifier	0.92
7	Decision Tree Classifier	0.91

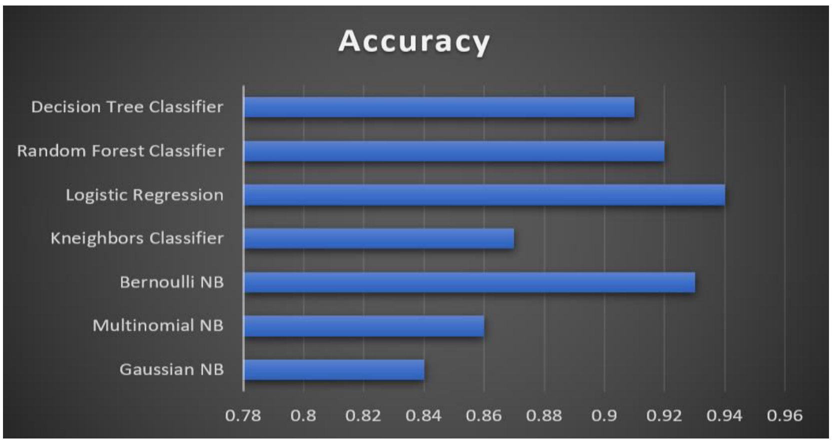

Fig. 5.2 : Analysis of accuracy parameters

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |
|| Volume 9, Issue 1, January 2021 ||
| DOI: 10.15680/IJIRCCE.2021.0901012 |

5.3 Analysis of Precision Score

This section presents the comparison of precision score parameter to show the customer satisfaction for airline tweets using sentiment analysis and the comparative analysis of this parameter is done among different machine learning such as random forest, Gaussian NB, Multinomial NB, Bernoulli NB, Kneighbors Classifier, and Decision Tree Classifier and our proposed method (Logistic Regression). The simulation results of our proposed method and existing method is shown in table 5.3 and it is 94% which is much more about the other exiting approach. The analysis of precision parameter is done using the comparison graph shown in figure 5.3 and it is found that our proposed method has higher value than the others. Due to the higher precision value it is analyzed that the proposed method is more success in the analysis of customer satisfaction.

Table 5.3: Comparative analysis of Precision parameter between Logistic regression and existing method

S. No.	Model Name	Precision
1	Gaussian NB	0.84
2	Multinomial NB	0.86
3	Bernoulli NB	0.93
4	Kneighbors Classifier	0.87
5	Logistic Regression	0.94
6	Random Forest Classifier	0.92
7	Decision Tree Classifier	0.91

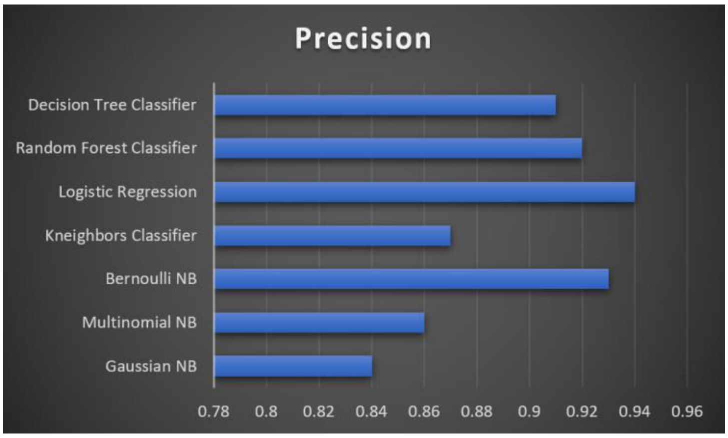

Fig. 5.3: Analysis of Precision Score

5.4 Analysis of F1 Measure

This section presents the comparison of F1 measure parameter to show the customer satisfaction for airline tweets using sentiment analysis and the comparative analysis of this parameter is done among different machine learning such as random forest, Gaussian NB, Multinomial NB, Bernoulli NB, Kneighbors Classifier, and Decision Tree Classifier and our proposed method (Logistic regression). The simulation results of our proposed method and existing method is shown in table 5.4 and it is 94% which is much more about the other exiting approach. The analysis of F1 measure parameter is done using the comparison graph shown in figure 5.10 and it is found that our proposed method has higher value than the others. Due to the higher F1 measure value it is analyzed that the proposed method is more success in the sentiment analysis of customer satisfaction.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |
|| Volume 9, Issue 1, January 2021 ||
| DOI: 10.15680/IJIRCCE.2021.0901012 |

Table 5.4: Comparative analysis of F1 measure parameter between Logistic regression and existing method

S. No.	Model Name	F1 Measure
1	Gaussian NB	0.84
2	Multinomial NB	0.86
3	Bernoulli NB	0.93
4	Kneighbors Classifier	0.87
5	Logistic Regression	0.94
6	Random Forest Classifier	0.92
7	Decision Tree Classifier	0.91

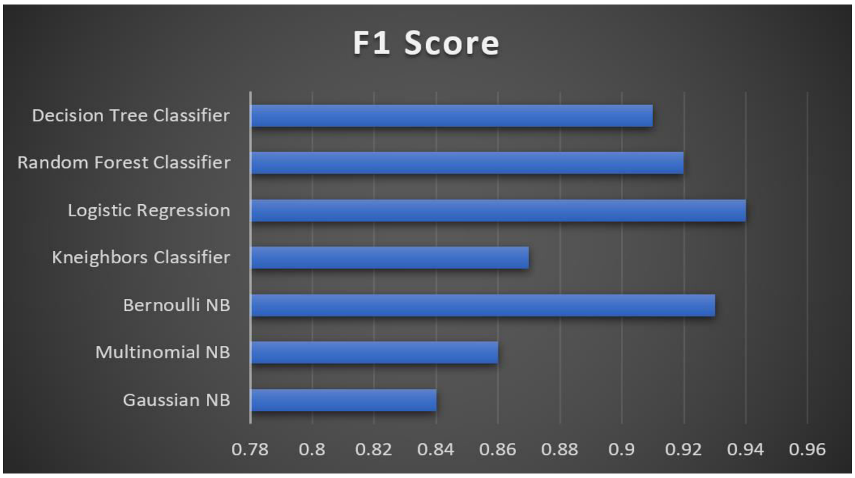

Fig. 5.4: Analysis of F1 measure parameters

5.5 Analysis of Recall parameter

This section presents the comparison of recall parameter to show the customer satisfaction for airline tweets using sentiment analysis and the comparative analysis of this parameter is done among different machine learning such as random forest, Gaussian NB, Multinomial NB, Bernoulli NB, Kneighbors Classifier, and Decision Tree Classifier and our proposed method (Logistic regression). The simulation results of our proposed method and existing method is shown in table 5.4 and it is 94% which is much more about the other exiting approach. The analysis of F1 measure parameter is done using the comparison graph shown in figure 5.5 and it is found that our proposed method has higher value than the others. Due to the higher recall value it is analyzed that the proposed method is more success in the sentiment analysis of customer satisfaction for airline tweets.

Table 5.5: Comparative analysis of F1 measure parameter between Logistic regression and existing method

S. No.	Model Name	Recall Score
1	Gaussian NB	0.84
2	Multinomial NB	0.86
3	Bernoulli NB	0.93
4	Kneighbors Classifier	0.87
5	Logistic Regression	0.94
6	Random Forest Classifier	0.92
7	Decision Tree Classifier	0.91

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0901012 |

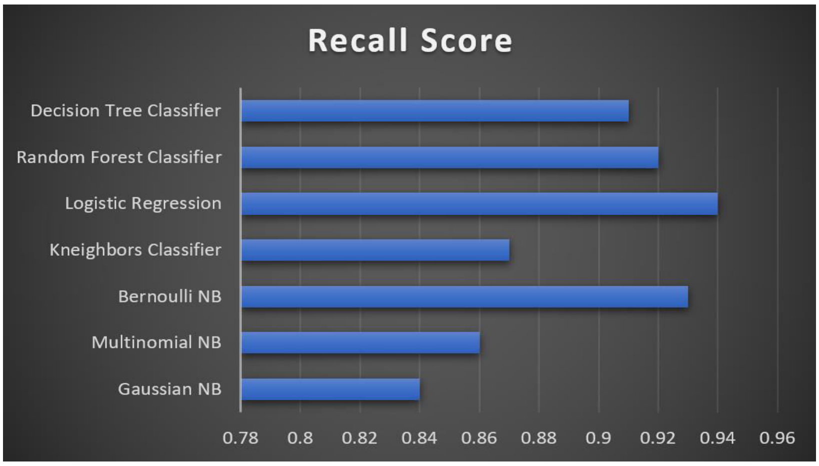

Fig. 5.5: Analysis of Recall parameters

VI. CONCLUSION AND FUTURE WORK

Twitter has become the online customer service platform in the world. The study evaluated twitter data take out with regard to Airline Industry. A number of standard major airline companies are selected across the world based on their followers and number of tweets on the Twitter. Today feedback is the foremost required thing to improve any kind of service in every field. Social media, Hashtags, tags, etc. are the best way to do any kind of feedback suggestions or anything because it creates a social impact on the social life of people. In this work, Twitter Airline Sentiment dataset is extracted and more than 65% as negative comments are only predicted. So the negative tweets are analyzed and the wordcloud of the negative words are created. The customer service issues and late flights are the two main reasons for negative comments of the customer in the airline services are analyzed. From the experimental results, we found that is it difficult to apply any other technique of data mining to the Twitter Airline Sentiment dataset. Before Machine learning models this process is very lengthy and time consuming in the analysis. In addition, most of the data of these dataset is in textual and numerical form which makes the extraction of data difficult. Due to this source data cannot be integrated easily. So, for this we apply machine learning technique which requires web searching, extracting, porter stemming and training and testing on different classifier datasets properly. In this dissertation, we use machine learning technique for our experimentation which have powerful classification like GuassianNB, MultinomialNB, BernoulliNB, KNeighnorsClassifier, Decision Tree, Logistic regression and random forest for classification and regression. The experimental analysis of proposed method and existing method is done using the performance measuring parameters like score, precision, recall, F1 measure and accuracy. For the simulation of proposed method and existing method Python language uses which is easy to implement and consume less computation time than other language. The result generated for proposed method after simulation for the accuracy and score parameter is 94% which is much more than the existing method. Similarly, the analysis of proposed and existing method is done using precision and recall parameter and the value of proposed methods is 94% which is also more than the existing method. Later the analysis of proposed and exiting method is done using F1 score and the value of F1 score of is 94% which is about 2-10%% more than the existing method. Based on these parameters the proposed model is capable of giving more than 90% accuracy in analyzing the sentiment of the feedback or tweets. This classifier can be used for airline amenities, business analysis bids, which will be able to spontaneously classify customer's satisfaction about airline services. Although customer satisfaction analysis can be analyzed using the available Twitter Airline Sentiment dataset. In this dissertation our proposed architecture for the classification and logistic regression technique analyze the sentiments in the tweets and it provides a drastic improvement in the performance of classification model. In future, the study recommends the ensemble classifier for analysis on tweets data to gain insight into tweets to help relevant airlines improve their customer experience.

REFERENCES

[1] Sandeep Ranjan et al., “Twitter Sentiment Analysis of Real-time Customer Experience Feedback for Predicting Growth of Indian Telecom Companies”, https://www.researchgate.net/publication/330391226.

[2] T. Hemakala and S. Santhoshkumar, “ Advanced Classification Method of Twitter Data using Sentiment Analysis for Airline Servic”, International Journal of Computer Sciences and Engineering, Vol.-6, Issue-7, July 2018, E-ISSN: 2347-2693.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0901012 |

[3] Prayag Tiwari et al. "Twitter-based Opinion Mining for Flight Service Utilizing Machine Learning", https://doi.org/10.31449/inf.v43i3.2615 Informatica 43 (2019) 381–386 381.

[4] T. Stein, E. Chen, and K. Mangla. Facebook immune system. In Proceedings of the 4th Workshop on Social Network Systems, SNS, volume 11, page 8, 2011.

[5] T. Hemakala and S. Santhoshkumar, "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.

[6] 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.

[7] Prayag Tiwari, Hari Mohan Pandey, Aditya Khamparia and Sachin Kumar, "Twitter-based Opinion Mining for Flight Service Utilizing Machine Learning", https://doi.org/10.31449/inf.v43i3.2615 Informatica 43 (2019), pp- 381–386.

[8] Janet R. McColl-Kennedy1 , Mohamed Zaki2, Katherine N. Lemon3, Florian Urmetzer2, 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.

[9] https://www.kaggle.com/crowdflower/twitter-airline-sentiment.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING