

# A SURVEY ON MACHINE LEARNING TECHNIQUES BASED SENTIMENT ANALYSIS FOR AIRLINE TWITTER DATA

<sup>1</sup>Ajai Rai, <sup>2</sup>Dr. Daya shankar pandey

<sup>1</sup>Department of Information Technology, <sup>2</sup>Department of Computer science  
RKDF-Institute of Science &Technology  
SRK University  
Bhopal, India

Abstract— Web technology's innovation & expansion have resulted in an enormous amount of data being made available to internet consumers, as well as a large amount of data being created. The Internet has evolved into a tool for online education, idea exchange, & opinion sharing. Due to the vast number of opinion-rich online resources like discussion forums, review sites, & blogs accessible in digital form, existing research concentrates on Opinion Mining, also known as sentiment analysis. Several individuals like expressing their views on the internet regarding a wide range of topics, including social events, certain items, and the services provided by various industries. It is possible to do sentiment analysis on opinion-rich data sources like Twitter to get insight into people's attitudes & sentiments. In addition, comments and views are useful in making decisions. Therefore, the application analyzing data about interesting services such as airline business on Twitter will be an actual useful assistance to airline service customers and providers. A great deal of work has been done on Twitter sentiment analysis. Sentiment analysis of the Airline Twitter dataset was described in this research using a variety of machine learning methods. Sentiment analysis in Twitter data is an important part of this study because it allows researchers to look at the information in tweets that is very unstructured & varied, and where views may be either positive or negative.

Keywords—Classification techniques, Machine Learning (ML), Sentiment analysis, Opinion Mining, US Airline Twitter Data.

## 1. INTRODUCTION

There is no doubt that the rapid & impressive growth of social media has surprised the globe. When new conditions arise, whether social, political, or recent affairs-related, individuals all around the globe communicate their thoughts via their actions, offering them good candidates for sentiment mining. The sentimental analysis turns out to be a valuable tool for any firm looking to better understand & serve its customers. To perform a sentiment analysis on their consumers, airlines rely on sophisticated data sources like Twitter to collect input from their users. Users who wish to know who's who and what's what in daily life can't obstruct the advantages of Twitter sentiment analysis [1].

Opinion mining, or SA, is a combination of NLP, computational linguistics, & text analytics that identifies polarity in sentiment. According to summaries of stated views or attitudes on documents, in phrases or at entities, opinion mining aims to categories the polarity of a particular document, sentence, phrase, or entity feature level. Because of this, a sentiment categorization method is used to determine whether or not the viewpoint represented in a text is positive or negative. Techniques for classifying opinions or feelings fall into two broad types [2,3]: the use of machine learning techniques and the adoption of a semantic orientation. For machine learning, a collection of representative data labeled manually with sentiment values is necessary for the training classification model. The semantic orientation technique, on the other hand, does not need any previous training since it uses sentiment scores directly to calculate positive & negative sentiments. To classify the text classification problem in sentiment analysis, machine learning is used. In this, to train a model, training data records are used which are later used to identify the predicted model without level. Every record is labeled into different classes. When we give a new unlabeled record to the model, then the model will label that dataset into different classes. There are three types of different classes such as positive, negative, and neutral. Generally neutral class is mixed opinion. Rarely do we consider the neutral class [4,5].

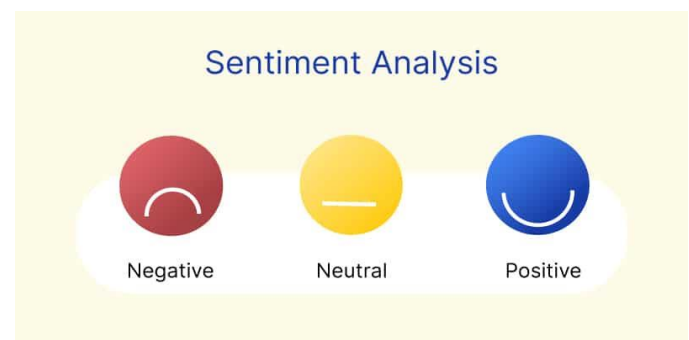


Figure 1: Sentiment Analysis

To extract emotions from text data like online articles, product reviews, movie reviews, Twitter data, and so on., Sentiment analysis is an essential technique to sentiment analysis. Twitter is a well-known social networking site with a global audience. Many people are using Twitter to express their opinions & ideas on an extensive range of topics,

resulting in an ever-increasing volume of data every day. Using machine learning algorithms to collect and interpret such data, which contains a wealth of information, may lead to useful recommender systems that help people handle their life in a much more comfortable manner. Most of the time, Twitter data contains information on a person's views on a wide range of subjects. Airline customers may express their thoughts on their travel experiences through Twitter. The quantity of information & data about airline services that can be found on Twitter is staggering. These tweets are gathered & analyzed to find out how clients feel about the airline's services & where they are located [6].

One of the most talked-about subjects on Twitter right now is air travel. Twitter is a common medium for airline customers to discuss their experiences. If this data is processed utilizing machine learning methods, it may yield insights that assist determine a passenger's degree of comfort throughout the journey. In the area of sentiment analysis, there is an enormous amount of research available. In terms of long-distance travel, air travel is one of the most convenient options, both domestically and globally. Globally, there is a big no. of ASPs (airline service providers). The airline industry is driven to compete for consumers by the competitive environment. Travelers, on the other hand, think about numerous factors before settling on a certain airline. These factors may include costs like fares, journey duration & luggage allowance, as well as reviews from previous customers and so on. As a result, all ASPs are attempting to enhance their facilities and in-flight comforts to attract more consumers. [7].

The rest of the study is structured accordingly: The deep description of sentiment analysis and OM in Section II. The SA using US airline twitter data is given in Section III. A detailed description of ML and its classification methods are given in section IV. Section V summarizes the associated studies on the sentiment analysis using airline twitter data. Section VI also gives our conclusion and suggestions for future study.

## 2. SENTIMENT ANALYSIS (SA) OR OPINION MINING (OM)

Several uses may be derived from opinion mining. For instance, in marketing, it keeps track of and assesses the rate of success of a marketing campaign or the introduction of a new product, determines the popularity of goods & services using their variants & informs us about populations that favor or hate specific characteristics. For instance, a digital camera review may be mostly good but negatively focused on the camera's weight. Via systematic identification, the vendor has a better understanding of public sentiment than through traditional methods like surveys or focus groups. Opinion mining & sentiment analysis are techniques for finding & locating subjective data in text materials. Attitude analysis may be used to gauge a writer's total contextual polarity or sentiment toward a certain issue or

topic. The fundamental problem in this field is the sentiment classifications, where sentiment can be a judgment, mood, or appraisal of an item, such as a film, book, or product, and can take the form of a document, statement, or characteristic which could be categorized as positive or negative. [8].

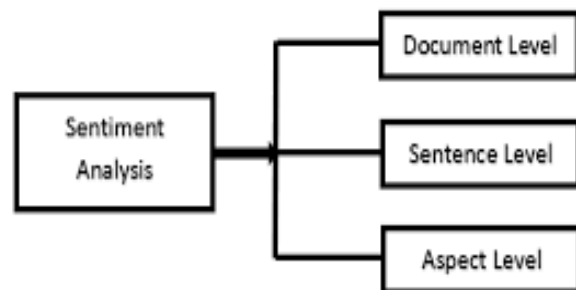


Figure 2: Key three levels of sentiment analysis

People's feelings, attitudes, & views concerning a certain entity may be studied using Sentiment Analysis (SA) or Opinion Mining (OM). People, events, or concepts may all be used to describe the entity. SA & OM have the same meaning when used together. Some researchers, on the other hand, believe that OM & SA have somewhat distinct beliefs. To extract & evaluate people's opinions on an entity, Opinion Mining is used, whereas Sentiment Analysis is used to identify and analyze the sentiment represented in text or documents. As a result, the goal of SA is to discover people's perspectives, detect the feelings they communicate, and then classify their agreement or disagreement. Categorization in SA may be divided into three categories: document, sentence, & SA aspect level [9]:

### 1) Document-level

The major problem in document-level sentiment analysis is to extract relevant material for inferring the overall sentiment. Because objective claims are generated by subjective statements, learning techniques might be muddled, which makes document classification more difficult because of contradicting sentiments. SA evaluates text material to determine if it expresses a positive or negative viewpoint or sentiment. The entire text is seen as a single entity of fundamental information.

### 2) Sentence-level

To classify emotions, we must proceed to a finer granularity than document level, where the sentence's polarity may be assigned to one of three categories: either positive, negative, or neutral. Classifying sentences based on their emotion presents a difficulty because of the identifying elements that indicate whether or not phrases are on-topic, an issue known as a co-reference. With the help of SA, each phrase is analyzed for the mood it expresses. To begin, determine whether the statement is subjective or objective. Sentence level SA determines whether or whether a sentence offers positive or negative views if it is subjective.

### 3) Aspect-level

The Aspect level SA is used to find out sentiment on Aspect of those entities. "My car has good handling but it is a little heavy" let's take this example. In this example, there is an opinion on a car that the handling of a car is positive but the weight of the car is negative. Aspect level sentiment analysis includes competitive statements. SA examines the feeling expressed above in light of the particular characteristics of entities. It uses fine-grained analysis & examines the opinion straight away. With this level of analysis, you're looking for attitudes toward different components of objects you're looking at.

### 4) Phrase level

When opinion words are used in a sentence, the phrase-level categorization is completed. These have advantages and disadvantages both because the advantage is, they're where the exact opinion about an entity is there. But in disadvantage, there is contextual polarity matter so the result may not be accurate.

### 5) Feature Level

A product feature is identified as a product attribute. In document Analyzing of these features for identifying sentiments called as feature-level sentiment analysis. Identifying if something is positive, negative, or neutral based on extracted attributes.

## A. Opinion Mining Applications

The natural language processing community is very interested in Sentiment Analysis & Opinion Mining systems these days since opinion-based or feedback-based applications are increasingly trendy. Researchers were able to collect user-generated material more readily because of the advent of the internet, which has transformed people's lifestyles by making them more vocal about their thoughts & beliefs. A broad variety of uses may be found for sentiment analysis & opinion mining [10,11].

### 1) Purchasing Product or Service

Choosing the proper product or service is no longer a tough chore when making a purchase. A person may readily analyze & contrast rival brands using this approach, as well as other people's opinions & experiences with a certain product or service in general. People no longer want to depend on outside consultants. Opinion mining & sentiment analysis take people's views from the internet's massive collection of unstructured information, evaluate them, and then display them in a highly organized & understandable way.

### 2) Quality Improvement in Product or service

OM and SA allow manufacturers to gather both negative and positive feedback on their product or service, allowing them to make improvements that will benefit their customers. They may make advantage of product reviews found on websites like Amazon, C|Net, RottenTomatoes.com, & IMDb.

### 3) Marketing research

Sentiment analysis methods may be used in marketing research to get useful results. User attitudes about a particular product or service may be assessed using sentiment analysis techniques. As with current popular attitudes regarding new government policies, it's simple to figure out which way they lean. A collective effort at intelligent inquiry might explain all of these findings.

### 4) Recommendation Systems

The method can determine which product should be suggested and which should not be suggested by sorting people's views into positive and negative ones.

### 5) Detection of "flame"

SA makes it simple to keep track of newsgroups & forums, blogs, & social media accounts. Opinion mining & sentiment analysis could automatically detect arrogant terms [6,] overheated words or hate language in emails, forum comments, & tweets from a variety of different online sources [7, 8].

### 6) Opinion spam detection

Since everyone has access to the internet and anybody may post anything, the likelihood of spam appearing online has grown. Creating spam material to deceive the public is a possibility. The internet material may be divided into two categories: spam & content that isn't spammed using opinion mining & sentiment analysis.

### 7) Policy Making

Sentiment analysis allows policymakers to see how citizens feel about a policy and use that knowledge to develop new policies that are more geared toward them.

### 8) Decision Making

An important part of decision-making is the input of other people's thoughts & experiences. Decision-making may benefit from analyzing people's opinions utilizing Opinion Mining & Sentiment Analysis.

## B. Approaches for Sentiment Analysis

There are no. methods available for analyzing and classifying sentiments to understand the opinions posted by individuals. There are several techniques available for sentiment analysis [12]:

### 1) Lexicon based approach

A dictionary containing both positive and negative terms used by Lexicon is applied to assess the polarity of opinion. The count of optimistic and pessimistic words is discussed in the text. If the text is more positive, a positive score will be assigned to the text. The text is awarded a negative score if it has a high amount of negative or pessimistic words. If the text contains the same number of good and bad terms, a neutral score is given. A lexicon of opinion (positive and negative opinions) is developed to finalize the word is positive or negative. There are numerous

ways to build and compile a dictionary.

### 2) Dictionary-based approach

A smaller number of words of opinion with established guidelines are gathered manually. In corpora like WordNet or thesaurus, synonyms and opposite from these words are then searched and appended to the group. The collection decreases slowly until there are no new terms. This method has the inconvenience of depending on the dictionary scale, the intensity of the sentiment classification. As the dictionary size increases, this approach is wrong

### 3) Corpus-based approach

They rely on massive corporations for syntactic & semantic opinion patterns. The created words are context-specific and need a large dataset labeled.

### 4) Machine learning-based approach

The usage of well-known ML technologies on text data is essential when using ML techniques for sentiment categorization. In terms of sentiment categorization using machine learning, there are essentially 2 kinds of techniques: supervised & unsupervised.

### 5) Hybrid based approach

The hybrid technique combines machine learning with a lexicon-based categorization system. To improve sentiment categorization, some research strategies advocate using both manual & automatic learning methods. This hybrid method is primarily advantageous as it can achieve the best of both. The combination of Lexicon and Learning has demonstrated increased accuracy.

### C. Advantages of sentiment analysis [13]:

- Lower cost than customer insight support.
- It is the faster way to collect customer insight data.
- It will be easy to act on the customer suggestion using sentiment analysis
- It will become very easy to identify a strength or weaknesses of other organizations or companies.
- The customer opinion will be more accurate.

## 3. SENTIMENT ANALYSIS OF US AIRLINE TWITTER DATA

Messages are posted & interactions with other individuals known as "tweets" are conducted through Twitter is tweeting and social networking service popular in the United States. Evan Williams and Jack Dorsey started Twitter in March 2006, and the service went online in July of that year. Twitter sentiment analysis is the technique of identifying whether a tweet expresses a good or negative attitude. It's a procedure of figuring out whether or not a piece of online writing has a positive, negative, or neutral attitude.

Thousands of clients every day are served by the airline industry, one of the world's most important & biggest sectors. An estimated 2,246,000 people use airplanes in the

USA each day, according to data published by the Federal Aviation Administration Air-Traffic [FAA]. It concentrates on the top 10 US airline carriers, especially America Air and Alaska Air and Delta and Jetblue and Hawaiian and Skywest, Southwest Airlines, United, Spirit, and Us Airways.

These airlines cover the same geographic region when flying throughout the United States, so it's no surprise that they're the most popular option. Additionally, there are low-cost airlines in the United States, as well as comparable flight fares. Furthermore, they are in a fierce rivalry with one another, which forces them to develop a strong competitive edge. There haven't been many further discoveries in the study of the airline industry depending on sentiment analysis. As a significant milestone, this study concentrated on bridging the gaps among users' perspectives & airline companies. In addition, the recommended study is implemented in various fields like entertainment, education, and transportation, among others.

User sentiment analysis on review or micro-blogging sites can aid consumers in their choice of the airline by examining other customer sentiments from different industries, including the airline business. This kind of sentiment analysis has been used in an extensive range of industries, like entertainment, education, and autos. Automatic sentiment analysis of US airline consumer opinion is the focus of this work, which makes extensive use of Twitter data to do so. Researchers have access to a variety of information about their user's thanks to Twitter data. Twitter is the largest common social networking site & app for real-time microblogging, and it is where breaking news first appears. The Tweet platform allows users to communicate by sending and receiving 140-character text messages called tweets. To help with information retrieval & decision-making, it's beneficial to categorize tweets. This characteristic of data mining allows for the extraction of meaningful information from textual data using text mining techniques. Scholars have begun to recognize the unique problems & possible uses of using sentiment analysis on Twitter, which is a growing trend in text mining [15].

A wide variety of airline services are available across the globe, each providing a unique set of airline amenities for its passengers to use. There's no way to know whether or not such airline services will please its consumers. Clients may not be able to offer feedback right away, thus airlines provide a twitter blog for them to do so. The number of people utilizing Twitter has risen to enhance the overall QoS. This study uses NLP to do sentiment analysis on a dataset built from tweets about US airlines. Positive, negative, & neutral sentiments were found in the dataset's text field. In today's world, airline companies use sentiment analysis & opinion analysis like machine learning tools to better understand their consumers and the public's perception of their services gleaned from online social media posts. Employees in the airline services department spend their time analyzing social media content like posts in online forums, blog comments,

tweets, & customer feedback. This evaluation is being exploited to form people's opinions or to track the improvement in the quality of their services. The input data to a categorization method must be closed as training data for classification algorithms. Based on the new training data, these methods estimate the types of class labels [15].

#### 4. OVERVIEW OF MACHINE LEARNING

By studying arithmetical & non-arithmetical methodologies, machine learning (ML) seeks to improve results and efficiency by systematizing information gained through comprehension & experiences. After many years of struggle, Machine Learning has progressed from a limited number of computer devotees using a choice of computer learning for fundamental statistics/mathematics with only occasional reflections of computational techniques, playing games, to a liberated discipline of studies that have provided the necessary platform for mathematical analysis, computational principles of learning methods but has also developed some many techniques that are frequently used for trend recognizing, text analysis, explanation/interpretation, and a variety of other money-making drives he has also developed a distinct study curiosity in the data mining field to explore hidden consistencies or anomalies in social data that is generated on a minute-to-minute basis. The theory, effectiveness, & attributes of learning systems & methods are the subject of ML, a branch of AI study. It is a very multidisciplinary area that draws on concepts from a wide range of scientific, engineering, & mathematical disciplines, including AI, information theory, statistics, optimization theory, cognitive science, and perfect controls. ML has had a significant influence on science & society because of its widespread usage in a variety of fields & contexts, including almost every scientific field. It has been used to a wide range of issues, such as suggestion engines, recognition systems, informatics & DM, and self-driving cars. [16,17].

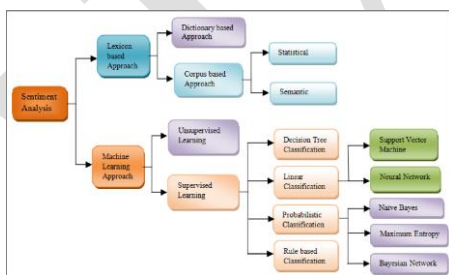


Figure 3: Sentiment Analysis using machine learning techniques

##### A. Machine learning-based Approaches

ML techniques are a common strategy for sentiment categorization. A machine learning technique employs virtually entirely & substantially machine learning (ML) methods to carry out statistical analysis (SA). Linguistic & syntactic aspects are combined utilizing ML techniques. Unsupervised learning, semi-supervised learning, &

supervised learning are three types of ML techniques for sentiment analysis tasks.

##### 1) Supervised Learning

The most well-known machine learning method. This strategy involves feeding a framework labeled source data to train a model. Using additional unlabeled input data, the trained model may then predict an output. Unsupervised & semi-supervised learning systems frequently exceed supervised learning in most circumstances, however, the reliance on labeled training data necessitates a lot of human work and is therefore wasteful in certain cases. Because these are training instances that human experts have labeled, supervised learning creates a function that maps to intended outputs, often known as labels. Any supervised learning approach, such as Naive Bayes classifications & SVM, may be used since it is a text categorization issue [18].

##### 2) Unsupervised Learning

Unsupervised learning approaches organize unlabeled data into clusters that are comparable to one another. The method, for instance, might deem data comparable depending on the document's frequent words or word pairs. Clustering, for example, is a kind of unsupervised learning in which the labels of inputs are not recognized during training. Some predefined syntactic structures are utilized for classifications, like those used to convey views. [19].

##### 3) Semi-supervised learning

Unlabeled & labeled data are used in equal measure throughout the training phase using semi-supervised learning. In addition to unlabeled data, certain samples of labeled data (frequently restricted) are incorporated in the construction of a classification model. If used accurately, this method may provide good results with lower effort than traditional supervised learning. Cross-domain & cross-language identification uses unlabeled data to extract area or language invariant characteristics while labeled target data is used to fine-tune a classifier. For Twitter sentiment analysis, semi-supervised learning is highly popular since enormous amounts of unlabeled data are accessible. [20,21].

#### B. Machine Learning Classification algorithms

Machine learning techniques are used to analyze the dataset. Classification is done by the classifier algorithms. The reviews data is used for performing sentimental analysis. These techniques give an idea about how the analysis is done on Twitter data by using various algorithms and machine learning concepts [.

##### 1) Naive Bayes (NB)

NB is dependent on the Bayes theorem using speculation amongst predictions. The Naive Bayes approach makes it possible to rapidly build frameworks that can forecast the future while also offering a new way to explore & comprehend the data. When using Naive Bayes to develop a predictive model, this method may be used for predictive

analysis. Utilizing naive Bayes as a predictor is not a bad idea.

## 2) Decision tree

The decision tree approach is perhaps the most often used in DM. One of the most effective classifiers is the decision tree, which can be implemented quickly and easily. A decision tree is a prediction model for data mining that makes use of a decision tree. A decision tree and a categorization algorithm are both utilized in this study to estimate illness based on patient data.

## 3) Regression

Data mining functions like regression are quite common. Categorization is comparable to regression. Age, weight, temperature, and illness are all predicted using regression analysis. Regression algorithms may be used to predict all of these things. Several medical area issues may be solved using regression tasks. Logistic & linear regressive approaches are the most often used regression techniques.

## 4) Artificial Neural Network

This information processing unit, inspired by human brains, is known as an Artificial Neural Network (ANN). Usually, neural networks are arranged in layers, with every layer consisting of several linked nodes, each with an activation function. It is the input layer that presents patterns to the network & interacts with one or more hidden layers, where the processing is carried out using a system of connection weights. To get the detecting outcome, the hidden layers are connected to an output layer.

## 5) Random forest

RF is a classifier of tree-defined collection, in which individually assigned random vectors are identically distributed and input  $x$  is the unit for each tree. Most of the time, random forest yields good results. Increasing its efficiency is challenging, and it can also deal with many sorts of data like numerical, binary, & nominal).

## 6) Logistic Regression (LR)

It's made up of logistic functions & looks like a sigmoid curve with an output of 0 and 1. In the form of an S curve, LR demonstrates growth & increases between 0 and 1 in the range.

## 7) XGboost (XGB)

(XGB) depending on the gradient boosting method, it has been built in an improved version to increase efficiency and speed. There are three primary parameters to the methodologies: boosters, learners, & general. In regression & tree, boosters' parameters are in charge of making the booster work. whereas learning variables are accountable for optimization & general variables are in charge of how well a method works as a complete.

## 8) Support Vector Machines (SVMs)

SVMs have supervised learning techniques for classifying, regression, & detecting outliers, among other things.

Employing SVM has a lot of advantages, including the following: i) It works well in high-dimensional spaces, ii) Because it only uses a subset of decision function's training points (known as support vectors), it uses less memory, iii) It is flexible, since alternative kernel functions may be given for decision function [22-26].

## 5. LITERATURE REVIEW

Opinion mining & sentiment analysis are both used in this study under the general name "sentiment analysis." Several associated studies covered in this article explore various sentiment analysis algorithms. Tweets from various airlines have been categorized using supervised Machine Learning algorithms & lexicon-based approaches. An overwhelming majority of studies on sentiment analysis have relied on finding emotional content in tweets' text.

**K. M. Hasib et al. (2021)** To analyze a dataset made up of tweets from 6 main US airlines & multi-class sentiment analysis, this study developed a unique deep learning method that successfully blends diverse word embedding using deep learning approaches. Pre-processing approaches for cleaning up tweets & extracting raw DNN data are used in this methodology. Using a three-class data set or a precision evaluation, the testing set can determine if a tweet is good, negative, or neutral. Our approach outperforms earlier models in terms of reliability, as shown by our understanding of the data from models offered by other researchers [27].

**M. M. Hrazi et al. (2021)** A sentiment analyzer based on the Twitter API will categorize actual tweets into one of the groups above. employed sentiment analysis depending on machine learning using numerous supervised learning techniques, like Logistics Regression (LR), Naive Bayes, SVM, & Decision Trees (DT). For validating & testing data, Logistic regression utilizing tri-gram characteristics with stop words and the TF-IDF feature extraction approach delivers the highest level of accuracy among the classifications strategies they used. [28].

**A. I. Saad et al. (2020)** A machine learning technique was developed in this study to sort tweets into positive, negative, & neutral groups dependent on sentiment. They used a dataset of tweets from six different US airlines to test their model's predictions. After that, they used six different machine learning approaches to sort tweets into different categories, including SVM, LR, NB, RF, XgBoost (XGB) & DT. Lastly, they employed the K-Fold Cross-Validation approach to test & validate the data after dividing it into 70 percent training and 30 percent testing during the validation phase. Next, they computed every classifier's Accuracy, Precision, Recall, & F1 score. The SVM classifier obtained the greatest precision of 83.31 percent when they compared the data from every classifier. [29].

**S. Malik et al. (2020)** For purpose of analyzing &

categorizing tweets as positive or negative, they use eight ML techniques: Multinomial NB, DT, RF, K-Nearest Neighbours, AdaBoosts, Support Vector Machines, Multilayer Perceptrons, and Stochastic Gradient Descents. SVM, Multinomial NB, Stochastic Gradient Descent, & RF exceed the other methods in experiments using the Twitter US Airline Sentiment dataset. They hope that our research will help companies & organizations enhance their offerings via the careful analysis of Twitter data. [30].

**R. Monika et al. (2019)** LSTMs can cope with long-term dependency by incorporating memory into a network prediction model & visualization in the researched sentiment analysis utilizing the Recurrent Neural Network (RNN) framework and the RNN prototype. In terms of categorization accuracy, the findings indicated an 80/20% split between the training set and the testing set, indicating that their methods could be trusted to accurately forecast the future. Further research studies use the Bidirectional LSTM Models (Bi-LSTM) to enhance this efficiency. [31].

**N. K. Sharma et al. (2019)** The new aspect of this project is the use of a suggested method to sort through confusing Tweets & eliminate them. To acquire the most precise outcomes, they mine & procedures the tweets of US airlines using various levels of mining & processes on the Twitter dataset. To categorize tweets according to feelings & convert tweets from ambiguous to positively or negatively, a new enhanced sentiment analytical framework depending on the NB classifier has been presented. [32].

**A. Rane and A. Kumar (2018)** A multi-class sentiments analysis was done on a dataset of tweets from six major US Airlines for this study. Classification methods used include DT, RF, SVM, K-Nearest Neighbors (KNN), LR, Gaussian NB, & AdaBoost. Eighty percent of data was used to train classifiers, & 20 percent of data was used to test them. There are three possible outcomes from this experiment: positive, negative, & neutral. Findings were used to compare categorization approaches and the total sentiment count was shown for all 6 airlines using the accuracies that were generated. [33].

**Y. Wan and Q. Gao (2015)** In this work, an ensemble sentiment categorization technique was used depending on the majority vote principle of numerous classifications techniques, like NB, SVM, Bayesian Network, C4.5 DT & RF methods. Six various classification algorithms, including the suggested ensemble methods, were trained & validated on the same dataset of 12864 tweets, with a 10-fold assessment used to validate the classifiers. In this airline service Twitter dataset, the suggested ensemble technique outperforms the individual classifiers. According to their findings, using an ensemble technique to classify Twitter sentiment might help other services as well. [34].

## 6. CONCLUSION

As a rule, opinion mining has been associated with the analysis of text strings to determine whether or not they include negative or positive emotions. It has been a while since opinion mining has been used to deal with challenges like distinguishing between an objective & subjective idea, which may be expressed in text informative sets including tweets or message boards or blogs or movie reviews or news articles. Sentiment intensity & opinion point recognition may be used by companies to get a deeper understanding and a broader range of estimates. An experience like this may help advance a company's understanding of its target market, improve the value it provides to clients, & raise its overall competitiveness. However, in the aerospace business where customer feedback could be easily gathered via surveys, Twitter serves as a reliable source of information for the investigation of customer opinions. OM Twitter data for airline services is a valuable application that assists users in collecting messages expressing a positive opinion or a negative comment. The objective of the application is to analyze messages about airline services on Twitter data by sentiment analysis.

## REFERENCES

- [1] Dutta Das, D., Sharma, S., Natani, S., Khare, N., & Singh, B. (2017). Sentimental Analysis for Airline Twitter data. IOP Conference Series: Materials Science and Engineering, 263, 042067. doi:10.1088/1757-899x/263/4/042067
- [2] Pakawan Pugsee, "Opinion Mining on Twitter Data for Airline Services", 2015 The 5th International Workshop on Computer Science and Engineering, 2015
- [3] B. Liu: Sentiment Analysis and Subjectivity in Handbook of Natural Language Processing 2nd Edition, edited by N. Indurkha and F.J. Damerau, Chapman & Hall/CRC Press, Taylor & Francis Group (2010).
- [4] B. Liu: Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data 2nd Edition, Springer (2011)
- [5] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede: Lexicon-Based Methods for Sentiment Analysis, Computational Linguistics Vol. 37 No. 2 (2011), p. 267-307.
- [6] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J., 2014, doi: 10.1016/j.asej.2014.04.011.
- [7] Kumar, S., & Zymbler, M. (2019). A machine

- learning approach to analyze customer satisfaction from airline tweets. *Journal of Big Data*, 6(1). doi:10.1186/s40537-019-0224-1
- [8] Arti Buche, "Opinion Mining and Analysis: A Survey", *International Journal on Natural Language Computing (IJNLC)* Vol. 2, No.3, June 2013
- [9] Archana Kalia, "A Brief Survey Paper on Sentiment Analysis", *International Journal of Innovative Science and Research Technology*, ISSN No:-2456 -2165, Volume 3, Issue 2, February – 2018
- [10] Ms. Sneha A. Sahare, "A Survey Paper on Opinion Mining and Sentiment Analysis", *JARIE-ISSN(O)-2395-4396*, Vol-3 Issue-2 2017
- [11] David Osimo and Francesco Mureddu, "Research Challenge on Opinion Mining and Sentiment Analysis", <https://www.w3.org/2012/06/pmod/opinionmining.pdf>
- [12] Anvar Shathik J. , "A Literature Review on Application of Sentiment Analysis Using Machine Learning Techniques", *International Journal of Applied Engineering and Management Letters (IJAEML)*, ISSN: 2581-7000, Vol. 4, No. 2, August 2020
- [13] Nirag T. Bhatt," Sentiment Analysis using Machine Learning Technique: A Literature Survey", *International Research Journal of Engineering and Technology (IRJET)*, Volume: 07 Issue: 12 | Dec 2020
- [14] Rana Al-Qahtani, "Predict Sentiment of Airline Tweets Using ML Models", March 29, 2021, EasyChair Preprint
- [15] M.VeeraKumaria , Prof.B.Prajna, "Collaborative Classification Approach for Airline Tweets Using Sentiment Analysis", *Turkish Journal of Computer and Mathematics Education*, Vol.12 No.3(2021), 3597-3603
- [16] Meenakshi, "Machine Learning Algorithms and their Real-life Applications: A Survey", *International Conference on Innovative Computing and Communication (ICICC 2020)*, Research scholar, U.I.E.T (M.D. University), Rohtak,,124001,India
- [17] Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, 2016(1). doi:10.1186/s13634-016-0355-x
- [18] Hemmatian F, Sohrabi MK (2017) A survey on classification techniques for opinion mining and sentiment analysis. *Artif Intell Rev* 52(3):1495–1545.
- [19] Li G, Liu F (2014) Sentiment analysis based on clustering: a framework in improving accuracy and recognizing neutral opinions. *Appl Intell* 40(3):441–452.
- [20] Pan Y, Liang M (2020) Chinese text sentiment analysis based on BI-GRU and self-attention. In: 2020 IEEE 4th information technology, networking, electronic and automation control conference (ITNEC) vol. 1. IEEE, pp 1983–1988
- [21] Da Silva NFF, Coletta LFS, Hruschka ER, Hruschka ER Jr (2016a) Using unsupervised information to improve semi-supervised tweet sentiment classification. *Inf Sci* 355:348–365.
- [22] N. Snehal and Tarun Gangil," Analysis of diabetes mellitus for early prediction using optimal features selection",*Springer open*, 2019, page 1-19
- [23] K.Priyadarshini," A Survey on Prediction of Diabetes Using Data Mining Technique", *International Journal of Innovative Research in Science, Engineering and Technology*, Vol. 6, Special Issue 11, September 2017
- [24] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," *IEEE Commun. Surv. Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016, doi: 10.1109/COMST.2015.2494502.
- [25] I. Muhammad and Z. Yan, "Supervised Machine Learning Approaches: a Survey," *ICTACT J. Soft Comput.*, vol. 05, no. 03, pp. 946–952, 2015, doi: 10.21917/ijsc.2015.0133.
- [26] N. F. Haq, A. R. Onik, A. K. Hridoy, M. Rafni, F. M. Shah, and D. Farid, "Application of Machine Learning Approaches in Intrusion Detection System: A Survey," *Int. J. Adv. Res. Artif. Intell.*, vol. 4, no. 3, pp. 9–18, 2015.
- [27] K. M. Hasib, M. A. Habib, N. A. Towhid and M. I. H. Showrov, "A Novel Deep Learning based Sentiment Analysis of Twitter Data for US Airline Service," 2021 International Conference on Information and Communication



- Technology for Sustainable Development (ICICT4SD), 2021, pp. 450-455, doi: 10.1109/ICICT4SD50815.2021.9396879.
- [28] M. M. Hrazi, A. M. Althagafi, A. T. Aljuhani, J. Rahman, M. M. Rahman and M. Shorfuzzaman, "Sentiment Analysis of Tweets from Airlines in the Gulf Region Using Machine Learning," 2021 International Conference of Women in Data Science at Taif University (WiDSTaif ), 2021, pp. 1-6, doi: 10.1109/WiDSTaif52235.2021.9430231.
- [29] A. I. Saad, "Opinion Mining on US Airline Twitter Data Using Machine Learning Techniques," 2020 16th International Computer Engineering Conference (ICENCO), 2020, pp. 59-63, doi: 10.1109/ICENCO49778.2020.9357390.
- [30] S. Malik, H. El-Sayed, M. A. Khan and H. Alexander, "Application of Containerized Microservice Approach to Airline Sentiment Analysis," 2020 14th International Conference on Innovations in Information Technology (IIT), 2020, pp. 215-220, doi: 10.1109/IIT50501.2020.9299043.
- [31] R. Monika, S. Deivalakshmi and B. Janet, "Sentiment Analysis of US Airlines Tweets Using LSTM/RNN," 2019 IEEE 9th International Conference on Advanced Computing (IACC), 2019, pp. 92-95, doi: 10.1109/IACC48062.2019.8971592.
- [32] N. K. Sharma, S. Rahamatkar and S. Sharma, "Classification of Airline Tweet Using Naïve-Bayes Classifier for Sentiment Analysis," 2019 International Conference on Information Technology (ICIT), 2019, pp. 70-75, doi: 10.1109/ICIT48102.2019.00019.
- [33] A. Rane and A. Kumar, "Sentiment Classification System of Twitter Data for US Airline Service Analysis," 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 2018, pp. 769-773, doi: 10.1109/COMPSAC.2018.00114.
- [34] Y. Wan and Q. Gao, "An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis," 2015 IEEE International Conference on Data Mining Workshop (ICDMW), 2015, pp. 1318-1325, doi: 10.1109/ICDMW.2015.7.