

SENTIMENT ANALYSIS AND CLASSIFICATION OF AIR INDIA FLIGHT INCIDENT USING YOUTUBE COMMENTS

Raj Kumar Singh¹ and Ani Thomas²

¹Department of Computer Science and Engineering, Chhattisgarh Swami Vivekanand Technical University, India

²Department of Information Technology, Bhilai Institute of Technology, India

Abstract

The usage of social media to interchange ideas and facts has increased exponentially due to technological advancements. Platforms for video sharing, like YouTube, have distinctive environments and architecture that people use for entertainment, education, and to keep themselves updated. YouTube is one of the most frequently used social media platforms, and users can connect to it by viewing, sharing opinions through comments, liking and disliking videos. A viewpoint or judgement formed about anything is referred to as an opinion. It can be collected and used to check knowledge, suggest the author with new video ideas, and analyze user behaviour. In this study, the data extracted from the free video-sharing platform YouTube concerning the 'Air India Flight Urination Case' was observed recently to recognize people's opinions on national and international levels. Based on approximately 10,000 comments about the incident, models are applied to classify and investigate the sentiments. This investigation uses TF-IDF and Bag of Words (BoW) text modelling techniques and observed that BoW performs better than TF-IDF. Moreover, Naive Bayes, Logistic Regression, Decision Tree, Support Vector Machines, and some ensemble algorithms like Random Forests, Gradient Boost, and Voting Classifier combining (Support Vector Machine, Decision Tree, Logistic Regression and Random Forest) with soft and hard voting had been applied and found that Support Vector Machine has the highest classification accuracy of 84%.

Keywords:

Natural Language Processing, Sentiment Analysis, Machine Learning, Ensemble, Flight, TF-IDF, Bag-of-Words

1. INTRODUCTION

Web 2.0 technology's emergence has increased the Internet bandwidth, technology for streaming videos has advanced significantly, and sharing information is no longer limited to still text or still photos. The dynamic video method increasingly became popular in presenting content. Due to this pattern, Internet users now frequently view content online, and various video-sharing services have sprouted up to accommodate the increased demand. The steady expansion of web videos has steadily impacted video viewers' behavioural traits. In today's scenario, where everyone needs all information within a few seconds, some search engines like YouTube and Google provide that information to the users on the spot. YouTube is useful in every industry, whether discussing education, medicine, entertainment, or tourism [1]. It is the second-largest search engine and the third-most visited website [2], [3]. YouTube is a global platform accessible in over 100 countries and 80 languages. Originally designed for video sharing, it has become a multifaceted resource, offering education, entertainment, and even a space for online communities. With over 2 billion monthly users, YouTube keeps people engaged for an average of more than 1 billion hours each day [4]. YouTube is becoming more and more popular due to the

prevalence and accessibility of smartphones and used at least once by an individual to gain some knowledge.

The incident can happen anywhere, and some incidents can be ignored, but some incidents are such that they impact human life. Apart from the news channel, many channels on YouTube discuss various incidents happenings. People keep themselves updated by watching them. Multiple users also use it as a discussion forum where views about the incident are shared as comments. People like to express their feelings more when an incident occurs where common people start visualizing themselves. Due to the broad occurrence, there has been an increase in worldwide social media users sharing their thoughts and feelings. The public's thoughts and feelings cover a wide spectrum. Every individual discusses their views and feelings as a YouTube comment. In this study, the sentiments on an incident are analyzed where a drunk man exposed himself and peed on a female passenger in the business class of Air India flight AI-102 travelling from New York to Delhi on November 26. Different people may have different views about this incident. It is important to declare that we use only YouTube comments for sentiment analysis here. Since the day of that incident, more than 100 YouTube videos have been uploaded by different content creators, whether it is a news channel or by an individual. It has also topped the list of most commented news on the Times of India website [5]. Information circulates digitally between users in this quick-paced world, which might influence how other users feel. Consequently, it's essential to understand the prevailing opinion. Sentiment analysis is a method of studying and interpreting human emotions through the processing of natural language [6], [7], [8].

Sentiment analysis used to analyze YouTube comments will provide insight into an important metric called polarity [9]. The data's polarity ranges from +1 and -1; negative sentiment is completely presented by -1, 0 is neutral, and +1 is entirely positive. With this study, the public's reaction to this incident is analyzed. The comments related to the incident are extracted using the keyword 'air India urination case'. The major goal of this study is to understand the public's perceptions of the specific occurrence posted on YouTube. The polarity and factual accuracy of comments about the demonstrations are examined by gathering YouTube data. The intent is to perform a detailed study of the YouTube comment data using visualization frameworks. Moreover, to identify the study's challenges and issues and talk about how this research might be incorporated into potential future works.

The well-known supervised machine learning algorithms. Like, Support Vector Machine, Decision Tree, Naive Bayes, Logistic Regression, Random Forest, Gradient Boosting and Voting Classifiers are applied for the categorization and prediction of sentiment in comments. By mixing many models, ensemble learning enhances machine learning outcomes [10]. An

ensemble is a supervised learning algorithm because it can be trained and used to make predictions. Since text data cannot be processed by computers in its raw state, it must be cleaned before machine learning models are trained. Therefore, manual text conversion into a numerical format is needed [11]. In light of this, we investigate the outcomes of two NLP techniques- Term Frequency, Bag of Words, and Inverse Document Frequency Approach [12]. TF-IDF and BoW are some methods of NLP that assist in transforming comment phrases into numerical vectors.

TF-IDF highlights terms that are unique to particular comments or subjects in the YouTube comment collection. and work better for tasks requiring the identification of important phrases, such as sentiment analysis, topic modelling, etc [13]. BOW is an easy and effective way to represent text and better for text classification tasks where word order and semantics are less important [14]. It is a basic approach that might be useful in capturing the general emotion or topic of a statement. It does not require significant linguistic analysis, making it simple to execute.

1.1 BAG OF WORDS(BOW) AND TERM FREQUENCY(TF) - INVERSE DOCUMENT FREQUENCY(IDF)

- *Bag of Words (BoW)*: The Bag of Words model separates characteristics from a remark that may be applied to modelling, such as in our example of using machine learning algorithms to classify the sentiment of comments. It has two components: a group of well-known terms and a measure for assessing their presence. The sequence in which they occur is also neglected in BoW. The initial step is constructing a vocabulary from all the different terms in the comment's data frame. The final step is to list these terms and track how often they appear in every remark. Finally, you supply the model with the number matrix for training.
- *TF-IDF*: Since it assesses the significance of a word in a phrase, the TF-IDF method performs better than the BoW technique [15]. When evaluating word frequency, one typical problem is that frequently occurring terms start to take over the text yet may not have the necessary "informative material" for the model to distinguish between them appropriately. IDF determines the importance of a word. IDF is required since simply computing the TF is insufficient to understand the importance of words.

1.2 NAÏVE BAYES

The Bayes' Theorem, which requires predictor independence, is the foundation for the supervised learning method Naive Bayes [16], [17]. The classifier believes that one function in a class does not influence the presence of any other characteristic. That makes it possible to understand what the Bayes theorem asserts. In machine learning, choosing the optimal hypothesis (h) given the dataset is a specific need (d). Taking advantage of our prior understanding of the circumstance is one of the easiest approaches to selecting a hypothesis. One may compute the likelihood of a belief given prior knowledge using Bayes' Theorem [18]. The Naive Bayes method was implemented using the sci-kit-learn toolkit, and before that, we used a count vectorizer to transform the comments into a token count matrix.

1.3 DECISION TREE

In decision trees, the dataset properties are represented by core nodes. Decision rules are presented by branches, and leaf nodes, which provide the result, comprise a tree-structured classifier. The components of a decision tree include decision nodes and leaf or branch nodes. Decision nodes make decisions with numerous branches, while branch nodes represent the outcomes of such choices, indicate whether the emotion is positive, negative, or neutral, and have no extra components. Here, the collection of comments is first regarded as the root node or the place to gather information. The method employed by Decision Trees is entropy, which determines how the results are divided [19]. A decision tree's bounds are impacted by it. It should also be noted that entropy values range from 0 to 1 [20].

$$E = -p \cdot \log(p) - q \cdot \log(q) \quad (1)$$

The computation of entropy is described in Eq.(1). Here, p is the probability of a positive class, and q is the negative class.

1.4 LOGISTIC REGRESSION

It is a common machine learning method in the supervised learning technique. It makes predictions about a dependent data variable by investigating the correlation between one or more independent variables that are already present. The method of logistic regression has grown in significance in the field of machine learning. We have included it in our study because training logistic regression is extremely effective and easy to execute and analyze. Moreover, it classifies unfamiliar records fairly quickly.

1.5 SUPPORT VECTOR MACHINE

Considering a set of training examples, each of which is classified into one of two categories, a model created by an SVM training algorithm categorizes fresh instances into one from two groups, making it a conditional linear classifier without probability. The SVM best fits in situations when there are more dimensions than samples. SVM provides a hyperplane (line) that acts as a decision border that distinguishes between the two classes. Depending on which side of the hyperplane a data point falls, it might belong to a different category.

1.6 RANDOM FOREST

Another supervised machine learning approach is random forest. It consists of several decision trees. The algorithm consists of two phases: Constructing a random forest and predicting using the classifier produced in the first step [21]. The random forest algorithm starts by choosing random samples from a pre-existing dataset. A decision tree will then be built using this method for each instance. The forecast outcome from each decision tree will then be obtained. Voting will be done in this phase for each expected result. Finally, choose the prediction result that received the most votes as the final forecast result.

We employed random forest to compare its precision to the decision tree method and determine how much better it would be. We found that the accuracy is much higher with fewer estimators than with the decision tree.

1.7 GRADIENT BOOSTING CLASSIFIER

Ensemble classifiers combine several separate classifiers, which can be helpful since doing so can produce more accurate predictions[22]. The boosting strategy connects several weak learners (predictors with low accuracy) (and a model with strong accuracy) to create a powerful learner. This approach advances stage-wise to construct an additive model. Each prediction in gradient boosting aims to outperform the one before it by lowering the errors [23].

1.8 VOTING CLASSIFIER

The Voting Classifier (VC) is an ensemble learning technique combining several basic models to create the best result. It normally accumulates the outcome of each classifier submitted to this classifier and predicts the output class with the greatest number of votes. However, instead of forming separate, specialized models and evaluating the validity of each, we create a single model trained using these models and predict output based on their aggregate majority of votes for each output type [24]. The voting classifier supports two different voting methods.

- *Hard Voting*: Hard voting is also known as majority voting, where each classifier casts a single vote, and the class with the most votes wins. When classifying data, a hard-voting ensemble predicts the type with the most votes by adding them for clear class labels from other models.
- *Soft Voting*: When using soft voting, each classifier offers a probability value that a specific data point belongs to a particular target class. It adds the predicted probabilities for the various class labels to predict the class label with the highest sum probability. Our study introduced a voting classifier SDLR, which combines Support Vector Machine, Decision Tree, Logistic Regression, and Random Forest. We have applied SDLR with soft and hard voting and found some interesting results discussed in sections 6 and 7.

Using TF-IDF and BoW, the textual information was transformed into numerical weightage in vector format. Additionally, for prediction purposes, eight classifiers: Naive Bayes, Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, Gradient Boost, and Voting Classifier with hard and soft voting. The earlier studies on sentiment analysis of YouTube data are shown in the following section. Methodology and Building a Model are covered in Sections 3 and 4. Section 5 discusses Model Prediction, and Section 6 discusses the Results and Analysis of Experiments. Finally, the Discussion and Conclusion were mentioned in Sections 7 and 8, respectively.

2. RELATED WORK

The rise of Web 2.0 technology has made the Internet increasingly user-centric. YouTube is a data centre where people create massive volumes of data. The general public frequently uses it to share their opinions on various public issues. A social media platform, such as YouTube, creates data that may be utilized for multiple purposes, including subject and person analysis. Due to the rise of social networks, more users have been uploading data online, such as views on individuals, events, and goods.

Modern machine learning and NLP algorithms can analyze user conversations on social network platforms' emotions. It has developed as an essential tool for recognizing human behaviour, studying public relations, and resolving various problems. With the rise of AI, we now have these tools available to us [25].

There are many uses for sentiment analysis, and multiple academics have utilized it to extract explanations from YouTube data to investigate an occurrence or address a problem. Several important initiatives are progressing, including classifying YouTube videos into six emotion categories. Hateful acts have increased in social networks due to greater social media use. One of these behaviours that users must defend themselves against on YouTube, Facebook, Twitter, etc., is hate speech because it is among the most hazardous. Sentiment analysis can also predict hate speech from various social media content. The author in [26] describes a system for predicting hate speech from social media websites using machine learning and natural language processing.

There have been numerous academic attempts with two classes (positive or negative), three classes (two with neutral), or multiple classes (happy, sad, fear, surprise, and anger). Nevertheless, selecting the most accurate model can be difficult. As a result, efforts by Alhujaili et al. have been made to identify the polarity using sentiment analysis of YouTube comments [27]. A concept of a semiotic dictionary based on emoji is proposed in [28]. The approaches and strategies for sentiment analysis that can be applied to YouTube videos are examined in their study.

Additionally, it describes various techniques helpful in sentiment analysis and data mining studies. Based on this rating, the video's popularity, relevance, and quality are typically preserved. However, the number of views or likes can often cause unrelated or poor-quality films to rank higher in search results, which doesn't make sense. H Bhuiyan et al. proposed a sentiment analysis method based on Natural Language Processing (NLP) for user comments to reduce this problem [29]. This technique enables locating the most pertinent and well-liked YouTube videos for the search.

Usually, liking or disliking the videos' relevance and quality is the defining factor determining their ranking. The significance of a video can often only be determined after watching it, and as a result, irrelevant videos may be placed higher. The authors have used aspect-based sentiment analysis to examine the effects of various characteristics of the video's subject on video retrieval [30], [31]. In [32], They looked at related work on filtering YouTube spam comments and ran classification tests using several machine-learning methods. To determine whether the most popular video has the most positive opinion and to discuss the challenges involved in analyzing the sentiment of comments under YouTube videos, [33] set out to find the emotion of the top five videos (based on views) under the recipe for the Bengali delicacy "Rasbora". As the Internet grows throughout the globe, users post comments in many languages. Sentiment analysis in one language increases the risk of overlooking crucial information in texts written in different languages. Therefore, multilingual emotion analysis algorithms have been developed to assess data in several languages. A sentiment analysis model of Marathi news has been proposed to filter out positive information [34]. In [35] and author mainly studied comments on YouTube and tweets from Twitter to grasp attitudes in the Arabic language.

The primary use of sentiment analysis is sentiment classification, which begins with a subjective examination. First, the text identifies whether it contains objective or subjective opinions expressing personal sentiment. A sentence, paragraph, and document-level sentiment analysis are all possible. Or we can say that emotion categorization aims to identify the words or sentences that best convey viewpoints, emotions, and attitudes. For instance, “wonderful” has a good connotation, whereas “ugliness” has a bad sense. Study in this area has increasingly advanced from analytical research on straightforward sentimental terms to more document-level sentiment and complex sentence analysis; the number of subjective articles conveying personal opinions online is growing [36]. The second stage is to classify polarities. The goal is to recognize individual responses to certain events. Sentences and phrases are organized into positive, negative, and neutral classes to assess the sentiment orientation and the viewpoints represented in the text. Instead, words and phrases are grouped according to feelings such as joy, sorrow, warmth, amusement, and surprise. The numerous emotional categories P Ekman includes in the emotion annotation and representation language encompass the majority of emotions [37].

India is a comprehensive part of a democratic society; anyone can express their opinion on any societal incident here. Recognizing the emotions behind online conversations is critical for understanding an event or an incident because it allows us to consider a larger audience and participants who can be both direct and indirect. Neogi et al. used Twitter data on farmers’ protests to learn more about the global public’s feelings about particular events [38]. Moreover, in [39] the author uses conversations on Twitter about the conflict between Russia and Ukraine recently to determine people’s sentiments.

Individual opinions can also be changed due to external influences. In [40] the negation approach in emotion mining has already been researched [41], [42]. A sentence’s polarity can frequently be determined by the emotive words or phrases that it contains. However, the polarities may be reversed by words or phrases’ negation; the scope of each negative word or sentence that comes before it determines the contextual polarity. Negations can take many forms and change the meaning of entire sentences and individual words. The area where meaning is altered is called the “scope of negation”.

Additionally, negations can change the sense of statements subtly. Using emoticons to categorize polarity is another area of interest [43], [44]. The first emoticon, which effectively conveyed emotion in written text, was used in 1982. Now emoticons are commonly used on social networking sites, discussion boards and blogs to express opinions. Based on past study and categorization of emoticons, the emoticons were divided into several emotion categories, such as happy, sad, furious, flirty, and sleepy. Sentiment strength detection is the third phase. Sentiment strength detection is further examined when sentiment categorization is finished. For example, videos of good quality and that are worth viewing again are both favourably orientated but may have different emotional polarity levels. As a result, to extract more sentiment information from sentiment analysis, it is necessary to analyze further how strongly positive and negative sentiments are expressed.

Machine learning approaches for sentiment categorization may be separated into supervised and unsupervised. Recent

comparisons of these approaches indicated that the accuracy of the supervised learning method is high, but the time taken to train on annotated data is more. On the other hand, unsupervised learning depends on the utilized sentiment corpus or dictionary. Unsupervised learning performs in real-time, despite the experimental outcomes being less precise than supervised learning [45]. Words, phrases, and sentences that are part of articles are measured. Following processing, the data are correlated in a vector space where each term or phrase denotes a dimension. Through Supervised learning, the associated weight for each dimension may be found. The TF-IDF and Bag of Words (BoW) were used as the primary technique for creating feature vectors in this investigation.

3. METHODOLOGY

In this section, we go through the detailed strategy used to find the people’s opinions about an incident that happened last week of November 2022. On a flight operated by Air India, travelling from New York to Delhi, a drunk male passenger peed on a co-passenger, a senior woman in her seventies sitting in a business class aisle seat [5]. Fig.1 illustrates the systematic procedure we used to predict and evaluate the opinion of a specific YouTube user. The first step in the process is to get the data from YouTube, followed by several important steps, such as cleaning, translating and preparing sentences to make them machine intelligible [46]. The next step is determining and categorizing a user’s sentiment using two criteria. Additionally, visualization techniques may be utilized to examine how people feel about certain issues. Finally, we plot the performance data and forecast the comments using machine learning methods.

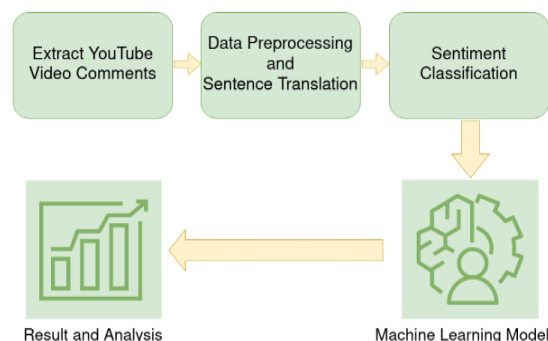


Fig.1. Step-by-step approach to sentiment analysis

3.1 DATASET

10,000 YouTube comments were collected in January 2023 from more than 50 videos uploaded related to the Air India urination incident. We have extracted raw data by using Google YouTube API in Python. It requires authentication using private access tokens and an API key. The incident happened and came into the news in the last week of November 2022, and we have chosen the starting date as after January 2023. Although we can select any date after this happens, we waited so long because we wanted to collect more and more comments. The only reason for choosing a particular incident in our study is to analyze textual content to determine public sentiment. We are not trying to convey our personal views on it. We used the keyword ‘air India flight urination case’ as a search query wherein all the comments

given to relevant YouTube videos were extracted. The comments from 9331 individual responses were gathered and kept in CSV format. Finally, the dataset has four attributes as Published_At, Comment_ID, Comment and Video_ID.

3.2 DATA PRE-PROCESSING

One of the most critical components of data investigation is ensuring that the computer understands the collected data. Machines can only grasp binary digits 0's and 1's; they cannot understand words, photos, or movies. It takes numerous steps to give an input only made up of binary numbers. Therefore, processing the data is very necessary, and it necessitates a method known as data cleaning that entails converting raw data into a form that machines can interpret [47], [48]. Because we have a large text dataset comprising YouTube comments, to remove significant variations to avoid data inconsistency, we must clean it. The method for cleansing data is quite simple. Before data cleaning, we used a Python open-source pandas library to remove duplicate records. We discovered 505 identical items that reduced the dataset's structure or the number of comments to 8826 unique items.

As shown in Table.1, some comments contain additional languages than English. Since solely the sentiment categorization of English sentences is the main focus of our investigation. Hence, we need to convert them into English. Some records also contain transliteration (The alphabets of another represent words and phrases from one language; their sound is preserved) [49], [50]. Python open-source library called Googletrans converts sentences, including transliterated ones, into English. Special letters, punctuation, numbers, and emojis were also eliminated from the comments.

Tokenization is a technique for dividing huge texts into tokens [51]. An initial step in text data modelling is tokenization. Next, it helps to determine the meaning of the text by studying the word order. We have used a text normalization technique called lemmatization to convert any word to its basic root mode [52]. Finally, we added a new pandas column named "all_eng_text" to our current comment dataset data frame to hold the completely pre-processed comments.

3.3 LEXICON-BASED SENTIMENT ANALYSIS APPROACH

The sentiment scores of all words in the document are aggregated in a Lexicon-based strategy using a pre-prepared sentiment lexicon [53]. In this work, we have chosen to use a lexicon-based methodology to bypass the process of producing labelled data. The key benefit of using a lexicon-based method is that it is considerably simpler to grasp and can be rapidly updated by humans. The orientation of semantics may be identified and classified as neutral, positive, or negative using this method. Sentiment analysis, by definition, is a technique for retrieving polarity and subjectivity from text, while semantic orientation assesses the text's polarity and strength [54]. Adjectives and adverbs are utilized in this method to reveal the semantic flow of the text [55]. In the next step, the sentiment orientation value of YouTube comments is calculated using various adverb and adjective combinations. A lexicon-based sentiment analyzer with pre-defined rules for words or lexicons is called VADER and is used on the collected dataset [56]. In addition to indicating whether the word dictionary is positive, negative, or neutral, VADER also suggests whether a sentence is positive, negative, or neutral. The result from VADER is a Python dictionary with four keys and their corresponding values as 'neg' for negative, 'neu' for neutral, 'pos' for positive, and 'compound' for the compound score. The compound score is calculated by normalizing the other three (negative, neutral, and positive) between -1 and +1. Python NLTK VADER lexicon is used to analyze sentiment. The comments are represented numerically, and VADER assigns individual scores to each. Finally, each statement is labelled as Positive, Negative, or Neutral using the compound score.

3.4 Sentiment Categorisation

We have discussed that the comments are represented numerically, and VADER assigns individual scores to each. The values can be exactly 0 in some circumstances. For example, a score for sentiment will be awarded based on the polarity value, and the calculation is performed so that the opinion will be reported as negative if the score is less than 0 and positive if the polarity is higher than 0. In all other circumstances, the sentiment is considered neutral, and the score is set to 0, as shown in Table.2.

Table.1. Sample dataset of Air India incident

Published_At	Comment_ID	Comment	Video_ID
2023-01-18 T09:09:52Z	UgwD8Y-wb- xQ3KcEMjB4AaABAg	Nice video. In my opinion there should be restriction in number of alcohol drinks served in International air-lines. Also, the passengers who are drunk should not be allowed to board a flight. I have seen many drunk pas-sengers allowed to board in a domestic flight.	LUHp78uxuol
2023-01-09 T16:53:56Z	Ugwtsv5o0HdEw90OMbN4AaABAg	bhaie kyon plane main logon ko sharab pilate ho	l_Rqvov5QuU
2023-01-07 T09:55:53Z	UgygEQjCSXBq4dHo19l4AaABAg	अमेरिकन कंपनी ने हमे आईना दिखाया है. हमें शर्म फिर भी नहीं आती अभी भी दोषी को सजा दी जा सकती है. जो दी भी जानी चाहिए	erDwCvo_rFQ
2023-01-12 T10:06:41Z	UgwNyCsGzGZJhYBjDbp4AaABAg	To all the people who are supporting Mishra, you will have a special place in hell.	Ru2TdCum5QY

Table.2. The Positive, Neutral and Negative comments from vader

Comment	Polarity Scale	Sentiment
In many country due to this reasons indians are ban from many places	-0.5574	Negative
Let s start from self discipline learn how to behave in public places	0	Neutral
There are a lot of bad habits in us but I must say we lack patriotism and unity that is the main reason of such problems	-0.8176	Negative
Great job guys spreading awareness god bless you	0.8402	Positive
Great video and journalism Good insights into the incident Could u make a video of how freshers are treated in Indian uni and how that can be solved	0.8442	Positive
I travel US to India and heard a lot of bad reviews of air India But this hits limit I doubt I will ever travel to air India again	-0.7184	Negative

After categorizing the emotions, we looked at the total number of opinions and discovered many individuals have negative feelings towards the incident.

Table.3. Algorithms accuracy and performance comparison using BoW

	Accuracy	Negative			Neutral			Positive		
		Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Naïve Bayes	64	0.58	0.87	0.70	0.84	0.45	0.58	0.62	0.53	0.57
Decision Tree	80	0.82	0.74	0.78	0.83	0.94	0.88	0.73	0.70	0.71
Logistic Regression	81	0.88	0.77	0.82	0.76	0.96	0.85	0.80	0.67	0.73
SVC	84	0.88	0.80	0.84	0.82	0.97	0.89	0.82	0.73	0.77
Random Forest	80	0.84	0.77	0.80	0.75	0.97	0.85	0.84	0.63	0.72
Gradient Boost	81	0.88	0.75	0.81	0.74	0.99	0.84	0.84	0.65	0.73
SDLR Soft Voting	82	0.86	0.78	0.82	0.81	0.97	0.88	0.80	0.70	0.75
SDLR Hard Voting	83	0.85	0.82	0.84	0.80	0.98	0.88	0.84	0.64	0.73

Table.4. Algorithms accuracy performance comparison using TF-IDF

	Accuracy	Negative			Neutral			Positive		
		Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Naïve Bayes	67	0.59	0.94	0.72	0.84	0.57	0.58	0.80	0.40	0.53
Decision Tree	79	0.81	0.75	0.78	0.82	0.92	0.87	0.73	0.70	0.71
Logistic Regression	78	0.81	0.79	0.80	0.73	0.89	0.80	0.81	0.61	0.69
SVC	82	0.85	0.81	0.83	0.80	0.92	0.85	0.80	0.71	0.75
Random Forest	79	0.82	0.77	0.79	0.75	0.95	0.84	0.82	0.61	0.70
Gradient Boost	80	0.88	0.76	0.81	0.74	0.98	0.84	0.82	0.63	0.71
SDLR Soft Voting	82	0.84	0.80	0.82	0.94	0.88	0.80	0.70	0.75	0.87
SDLR Hard Voting	81	0.81	0.84	0.82	0.80	0.94	0.86	0.87	0.60	0.71

4. MODEL PREDICTION

To fit the algorithms mentioned above into our dataset, we use the sci-kit-learn machine-learning package based on Python. This part also covers the acquired data's prediction, visualization, and analysis. A comparison of the accuracy and performance comparison of negative, neutral, and positive sentiment for each ML algorithm using BoW and TF-IDF are shown in Table.3 and Table.4, respectively.

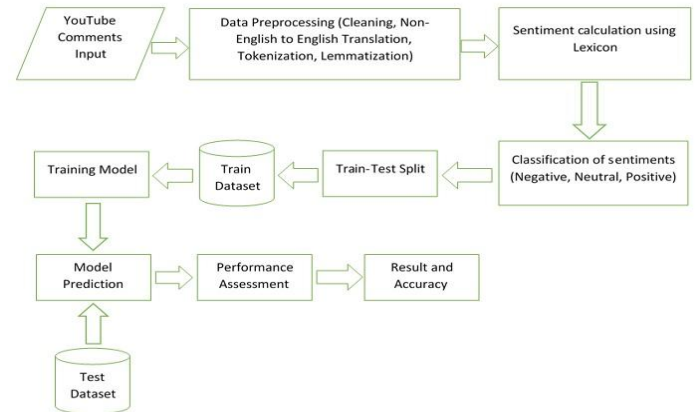


Fig.2. Sentiment Analysis Framework for YouTube Comment

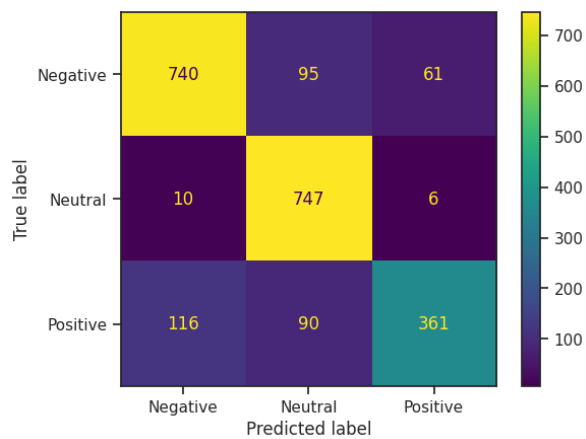


Fig.5.(a) Confusion matrix for SVM using the BoW vectorizer.
(b) Confusion matrix for SDLR hard voting using BoW vectorizer

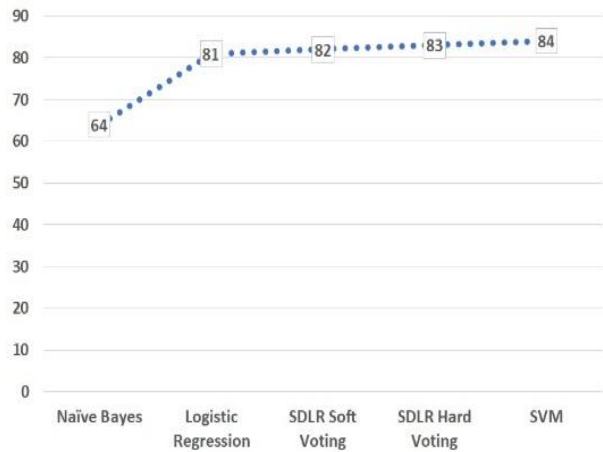


Fig.6. Algorithms accuracy chart for ML algorithm using BoW

6. DISCUSSION

This study focused on extracting and analyzing comments on the Air India flight urination incident. This incident occurred in the last week of November and was in the headlines of almost all digital and print media for a long time. Data on this incident continue to be abundantly generated on various social networking sites. As a result, standard techniques for processing such massive data are quite difficult; thus, we require significant computational capacity and methods to process it quickly. After analyzing the comments, it was found that most people want that alcohol should not to be served in flight or any public place.

They wish that some restrictions or punishment should be given to the culprit. In other words, we can say that people have a negative sentiment towards this incident. This study can also provide some directions to organization policymakers to take appropriate action. Some strict action was taken as Air India has decided to ban the accused from flying for four months, and the Directorate General of Civil Aviation (DGCA), the aviation regulator, fined Air India 30 lakh rupees are some example that supports our research outcome.

Several studies have been undertaken to categorize the user’s opinion because online information spreads quickly [58]. Our

research focuses on recognizing various sentiments by analyzing public comments and organizing the views according to polarity using a variety of machine-learning approaches. We have applied different ML techniques with TF-IDF and BoW factorization and found that BoW is more accurate than TF-IDF.

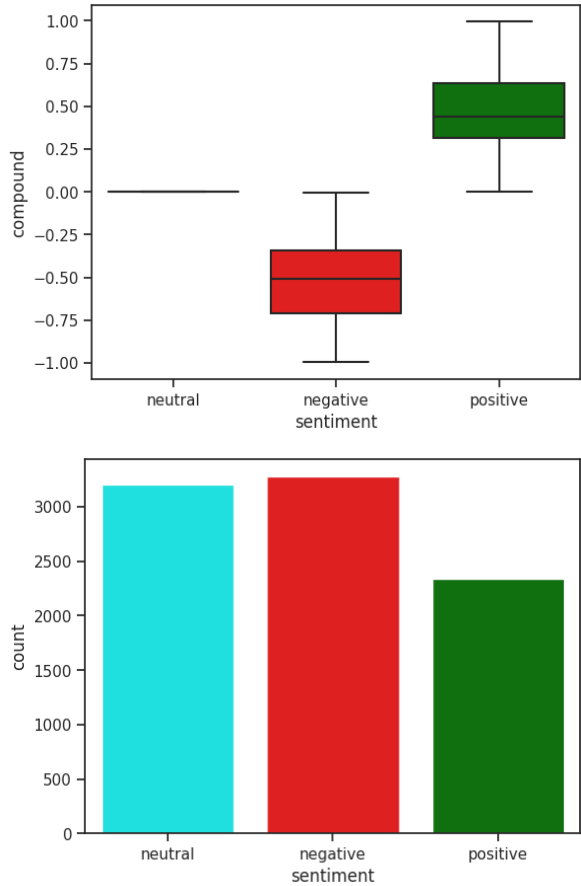


Fig.7. (a) Box plot. (b) Sentiment polarity range

7. CONCLUSION

Online platforms have enabled people to communicate their views, ideas, and opinions [59]. Social networks have increased in popularity for this purpose, spreading ideas and creating personal beliefs. Due to this, the flight incident saw a huge increase in YouTube video uploads where users shared their thoughts as comments. It has given rise to all types of people who are angry about the problem. In this paper, by creating a sentiment analysis algorithm, we looked at ways to comprehend people’s sentimentality and determine the direction of the incident outcome. The box plot shown in Fig.7(a) depicts a more negative sentiment polarity range than positive sentiment. The chart in Fig.7(b) indicates that most comments are negative, with neutral feelings coming in second and positive sentiments coming in last. In addition, for classification and prediction, eight typical machine learning models were applied and discovered that the support vector machine produced the best results.

REFERENCES

- [1] H. Crowel, H. Gribben and J. Loo, "Travel Content Takes off on YouTube", *Think with Google*, pp. 1-11, 2014.
- [2] K.S. Burns, "Social Media: A Reference Handbook", South Florida: Greenwood Press, 2017.
- [3] S. Aslam, "YouTube by the Numbers: Stats, Demographics and Fun Facts", Available at <https://www.omnicoreagency.com/youtube-statistics/>, Accessed in 2023.
- [4] M. Mohsin, "10 YouTube Stats Every Marketer Should Know in 2023", Available at <https://www.oberlo.com/blog/youtube-statistics>, Accessed in 2024.
- [5] "Most Commented News", Available at <https://timesofindia.indiatimes.com/mostcommented.cms?day=30>, Accessed in 2023.
- [6] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", *Proceedings of International Conference on Language Resources and Evaluation*, pp. 1-8, 2010.
- [7] A. Berger, S.A. Della Pietra and V.J. Della Pietra, "A Maximum Entropy Approach to Natural Language Processing", *Computational Linguistics*, Vol. 22, No. 1, pp. 39-71, 1996.
- [8] D.S.M. Nandhini and P. Gurunathan, "Joint Aspect-Opinion Extraction and Sentiment Orientation Detection in University Reviews", *International Journal of Information Technology*, Vol. 14, No. 6, pp. 3213-3225, 2022.
- [9] A. Srivastava, V. Singh and G.S. Drall, "Sentiment Analysis of Twitter Data: A Hybrid Approach", *International Journal of Healthcare Information Systems and Informatics*, Vol. 14, No. 2, pp. 1-16, 2019.
- [10] O. Sagi and L. Rokach, "Ensemble Learning: A Survey", *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, Vol. 8, No. 4, pp. 1-8, 2018.
- [11] M.M. Danyal, S.S. Khan, M. Khan, S. Ullah, M.B. Ghaffar and W. Khan, "Sentiment Analysis of Movie Reviews based on NB Approaches using TF-IDF and Count Vectorizer", *Social Network Analysis and Mining*, Vol. 14, No. 1, pp. 1-8, 2024.
- [12] Y. Zhang, R. Jin and Z.H. Zhou, "Understanding Bag-of-Words Model: A Statistical Framework", *International Journal of Machine Learning and Cybernetics*, Vol. 1, pp. 43-52, 2010.
- [13] V. Kalra, I. Kashyap and H. Kaur, "Improving Document Classification using Domain-Specific Vocabulary: Hybridization of Deep Learning Approach with TFIDF", *International Journal of Information Technology*, Vol. 14, No. 5, pp. 2451-2457, 2022.
- [14] G.K. Wadhwani, P.K. Varshney, A. Gupta and S. Kumar, "Sentiment Analysis and Comprehensive Evaluation of Supervised Machine Learning Models using Twitter Data on Russia-Ukraine War", *SN Computer Science*, Vol. 4, No. 4, pp. 1-11, 2023.
- [15] S. Akuma, T. Lubem and I.T. Adom, "Comparing Bag of Words and TF-IDF with Different Models for Hate Speech Detection from Live Tweets", *International Journal of Information Technology*, Vol. 14, No. 7, pp. 3629-3635, 2022.
- [16] F.J. Yang, "An Implementation of Naive Bayes Classifier", *Proceedings of International Conference on Computational Science and Computational Intelligence*, pp. 301-306, 2018.
- [17] I. Rish, "An Empirical Study of the Naive Bayes Classifier", *Proceedings of International Workshop on Empirical Methods in Artificial Intelligence*, pp. 41-46, 2001.
- [18] S. Zervoudakis, E. Marakakis, H. Kondylakis and S. Goumas, "OpinionMine: A Bayesian-based Framework for Opinion Mining using Twitter Data", *Machine Learning with Applications*, Vol. 3, pp. 1-7, 2021.
- [19] P.H. Swain and H. Hauska, "The Decision Tree Classifier: Design and Potential", *IEEE Transactions on Geoscience Electronics*, Vol. 15, No. 3, pp. 142-147, 1977.
- [20] Y.Y. Song and L.U. Ying, "Decision Tree Methods: Applications for Classification and Prediction", *Shanghai Arch Psychiatry*, Vol. 27, No. 2, pp. 1-6, 2015.
- [21] G. Biau and E. Scornet, "A Random Forest Guided Tour", *Test*, Vol. 25, No. 2, pp. 197-227, 2016.
- [22] R. Blagus and L. Lusa, "Gradient Boosting for High-Dimensional Prediction of Rare Events", *Computational Statistics and Data Analysis*, Vol. 113, pp. 19-37, 2017.
- [23] J.H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine", *The Annals of Statistics*, Vol. 29, No. 5, pp. 1189-1232, 2001.
- [24] A. Yousaf, "Emotion Recognition by Textual Tweets Classification using Voting Classifier (LR-SGD)", *IEEE Access*, Vol. 9, pp. 6286-6295, 2021.
- [25] N. Chintalapudi, G. Battineni, M. Di Canio, G. Sagaro and F. Amenta, "Text Mining with Sentiment Analysis on Seafarers' Medical Documents", *International Journal of Information Management Data Insights*, Vol. 1, pp. 1-9, 2021.
- [26] Z. Al-Makhadmeh and A. Tolba, "Automatic Hate Speech Detection using Killer Natural Language Processing Optimizing Ensemble Deep Learning Approach", *Computing*, Vol. 102, No. 2, pp. 501-522, 2020.
- [27] R.F. Alhujaili and W.M.S. Yafooz, "Sentiment Analysis for Youtube Videos with User Comments: Review", *Proceedings of International Conference on Artificial Intelligence and Smart Systems*, pp. 814-820, 2021.
- [28] D. Chauhan and K. Sutaria, "Multidimensional Sentiment Analysis on Twitter with Semiotics", *International Journal of Information Technology*, Vol. 11, No. 4, pp. 677-682, 2019.
- [29] H. Bhuiyan, J. Ara, R. Bardhan and M.R. Islam, "Retrieving YouTube Video by Sentiment Analysis on User Comment", *Proceedings of International Conference on Signal and Image Processing Applications*, pp. 474-478, 2017.
- [30] G.S. Chauhan and Y.K. Meena, "YouTube Video Ranking by Aspect-based Sentiment Analysis on User Feedback", *Soft Computing and Signal Processing*, Vol. 900, pp. 63-71, 2019.
- [31] K. Verma and B. Davis, "Implicit Aspect-based Opinion Mining and Analysis of Airline Industry based on User-Generated Reviews", *SN Comput Science*, Vol. 2, No. 4, pp. 1-5, 2021.
- [32] H. Oh, "A YouTube Spam Comments Detection Scheme using Cascaded Ensemble Machine Learning Model", *IEEE Access*, Vol. 9, pp. 144121-144128, 2021.

- [33] Y.L. Chen, C.L. Chang and C.S. Yeh, "Emotion Classification of YouTube Videos", *Decision Support Systems*, Vol. 101, pp. 1-7, 2017.
- [34] M.S. Divate, "Sentiment Analysis of Marathi News using LSTM", *International Journal of Information Technology*, Vol. 13, No. 5, pp. 2069-2074, 2021.
- [35] M. Abdul-Mageed and M. Diab, "SANA: A Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis", *Proceedings of International Conference on Language Resources and Evaluation*, pp. 1162-1169, 2014.
- [36] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis", *Mining Text Data*, pp. 415-463, 2012.
- [37] P. Ekman, "An Argument for Basic Emotions", *Cognition and Emotion*, Vol. 6, No. 4, pp. 169-200, 1992.
- [38] A.S. Neogi, K.A. Garg, R.K. Mishra and Y.K. Dwivedi, "Sentiment Analysis and Classification of Indian Farmers' Protest using Twitter Data", *International Journal of Information Management Data Insights*, Vol. 1, No. 2, pp. 1944-1949, 2021.
- [39] G.K. Wadhwani, P.K. Varshney, A. Gupta and S. Kumar, "Sentiment Analysis and Comprehensive Evaluation of Supervised Machine Learning Models using Twitter Data on Russia-Ukraine War", *SN Computer Science*, Vol. 4, No. 4, pp. 1-8, 2023.
- [40] S.E. Uthirapathy and D. Sandanam, "Predicting Opinion Evolution based on Information Diffusion in Social Networks using a Hybrid Fuzzy based Approach", *International Journal of Information Technology*, pp. 1-6, 2022.
- [41] A. Hogenboom, P. Van Iterson, B. Heerschop, F. Frasincar and U. Kaymak, "Determining Negation Scope and Strength in Sentiment Analysis", *Proceedings of International Conference on Systems, Man and Cybernetics*, pp. 2589-2594, 2011.
- [42] L. Jia, C. Yu and W. Meng, "The Effect of Negation on Sentiment Analysis and Retrieval Effectiveness", *Proceedings of International Conference on Information and Knowledge Management*, pp. 1827-1830, 2009.
- [43] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. De Jong and U. Kaymak, "Exploiting Emoticons in Polarity Classification of Text", *Journal of Web Engineering*, Vol. 14, No. 2, pp. 22-40, 2015.
- [44] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai and A. Kappas, "Sentiment Strength Detection in Short Informal Text" *Journal of the American Society for Information Science and Technology*, Vol. 61, pp. 2544-2558, 2010.
- [45] P. Chaovalit and L. Zhou, "Movie Review Mining: A Comparison between Supervised and Unsupervised Classification Approaches", *Proceedings of International Conference on System Sciences*, pp. 1-5, 2005.
- [46] H. Malik and Z. Tian, "A Framework for Collecting YouTube Meta-Data", *Procedia Computer Science*, Vol. 113, pp. 194-201, 2017.
- [47] E. Haddi, X. Liu and Y. Shi, "The Role of Text Pre-Processing in Sentiment Analysis", *Procedia Computer Science*, Vol. 17, pp. 26-32, 2013.
- [48] V. Mohan, "Preprocessing Techniques for Text Mining - An Overview", *International Journal of Computer Science and Communication Networks*, Vol. 5, pp. 1-7, 2016.
- [49] S. Karimi, F. Scholer and A. Turpin, "Machine Transliteration Survey", *ACM Computing Surveys*, Vol. 43, No. 3, pp. 1-46, 2011.
- [50] C. Van Lieshout and W. Cardoso, "Google Translate as a Tool for Self-Directed Language Learning", *Language Learning and Technology*, Vol. 26, No. 1, pp. 1-19, 2022.
- [51] J.J. Webster and C. Kit, "Tokenization as the Initial Phase in NLP", *Proceedings of International Conference on Computational Linguistics*, Vol. 4, pp. 1106-1110, 1992.
- [52] V. Balakrishnan and L.Y. Ethel, "Stemming and Lemmatization: A Comparison of Retrieval Performances", *Lecture Notes on Software Engineering*, Vol. 2, pp. 262-267, 2014.
- [53] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexicon-based Methods for Sentiment Analysis", *Computational Linguistics*, Vol. 37, No. 2, pp. 267-307, 2011.
- [54] A. Szabolcsi, "Positive Polarity - Negative Polarity", *Nat Lang Linguist Theory*, Vol. 22, No. 2, pp. 409-452, 2004.
- [55] S. Jain, K. R. Seeja and R. Jindal, "A Fuzzy Ontology Framework in Information Retrieval using Semantic Query Expansion", *International Journal of Information Management Data Insights*, Vol. 1, No. 1, pp. 1-6, 2021.
- [56] C.J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text", *Proceedings of International Conference on Web and Social Media*, Vol. 8, No. 1, pp. 216-225, 2014.
- [57] F. Heimerl, S. Lohmann, S. Lange and T. Ertl, "Word Cloud Explorer: Text Analytics based on Word Clouds", *Proceedings of International Conference on System Sciences*, pp. 1833-1842, 2014.
- [58] N. Joseph, A. Kar and V. Ilavarasan, "How Do Network Attributes Impact Information Virality in Social Networks?", *Information Discovery and Delivery*, Vol. 49, No. 2, pp. 162-173, 2021.
- [59] W. Sutherland and M.H. Jarrahi, "The Sharing Economy and Digital Platforms: A Review and Research Agenda", *International Journal of Consumer Studies*, Vol. 43, pp. 328-341, 2018.