TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.02, I.02(2023)
https://www.tejasjournals.com/

# Consumer Sentiment Analysis Using Deep Learning

Neha Singh[a] and Yogendra Pratap Singh[b]
[a,b] Dept. of CSE, Goel Institute of technology & Management, Lucknow, India
[a]nehasin9140@gmail.com, [b]yogendra.p.singh@goel.edu.in

KEYWORD

*Sentiment analysis, Twitter data, AI algorithms, Decision trees, Emotion analysis, Web applications, Image processing, Multimodal analysis*

ABSTRACT

*This research focuses on sentiment analysis of Twitter data using various AI algorithms. The study explores the potential of decision trees for sentiment classification and examines the importance of emotion analysis in understanding public opinions. Challenges in sentiment analysis are discussed, including handling vague statements, expanding sentiment categories, and integrating image processing techniques. Future directions include developing web applications for user-friendly sentiment analysis and incorporating multimodal analysis to capture emotions from multiple sources. The findings underscore the significance of sentiment analysis in diverse fields like marketing, customer feedback analysis, and social research*

## 1. Introduction

In recent years, the convergence of virtual entertainment platforms, online survey systems, and user feedback portals has led to an explosion of user-generated content. This vast volume of data contains valuable information about people's opinions, sentiments, and evaluations towards various products, services, events, and topics. Extracting and understanding this unstructured data has become crucial for organizations, governments, and researchers to gain insights into public sentiment, consumer satisfaction, brand perception, and political trends. Sentiment analysis, also known as opinion mining, has emerged as a powerful tool to automatically extract, assess, and comprehend sentiments from textual data.

Sentiment analysis involves computationally identifying and categorizing opinions expressed in a piece of text as positive, negative, or neutral. Artificial intelligence systems have revolutionized sentiment analysis, enabling the development of powerful models capable of deriving meaningful insights from data and making accurate predictions even on previously unseen text. These predictions find applications in various domains, including market research, customer feedback analysis, political sentiment tracking, brand management, and much more.

The primary objective of this paper is to develop efficient machine learning models for sentiment analysis. To achieve this goal, the study aims to compile and preprocess a diverse range of textual data from various sources like social media, online reviews, news articles, and discussion forums, creating comprehensive datasets for training and testing the sentiment analysis models. The research involves the investigation and implementation of supervised, unsupervised, and deep learning algorithms to explore their strengths and weaknesses, identifying the most effective approaches. The developed models will be rigorously evaluated using standard metrics, comparing them against baseline methods and state-of-the-art techniques to assess their accuracy and efficiency. Additionally, the study seeks to delve into model interpretability, gaining insights into the influential features and linguistic patterns that contribute significantly to sentiment prediction, ultimately enhancing the understanding of the decision-making process.

This study's significance lies in its potential to advance opinion analysis, artificial intelligence, and natural language processing. Accurate sentiment evaluation models can benefit organizations by understanding customer preferences, enhancing products and services, and enabling data-driven decisions. Policymakers and political specialists can also utilize the research for assessing public sentiment.

**Corresponding Author:** Neha Singh, Goel Institute of technology and management, Lucknow
**Email:** nehasin9140@gmail.com

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.02, I.02(2023)
**https://www.tejasjournals.com/**

## 2. Related Work

Prabhsimran Singh, Karanjeet Singh Kahlon, Ravinder Singh Sawhney, and Singh. "Opinion investigation into the Indian government's demonetization of the 500 and 1000 rupee banknotes." 2017's ICT Express[1].

This study focuses on sentiment analysis using Twitter data related to the Indian government's demonetization policy. They classify tweets into categories like happy, sad, extremely sad, neutral, and no effect, employing keywords and hashtags like #demonetization for data collection [5].

Divakar Yadav, Geetika, and Gautam. "AI draws near and semantic investigation for opinion analysis of twitter information." Seventh international conference on contemporary computing (IC3), IEEE, 2014.

This work presents the Feeling Examination for Clients Audit Arrangement, using AI algorithms like Naive Bayes, Maximum Entropy, and SVM for sentiment analysis. Python and Natural Language Toolkit (NLTK) are used for preparation and processing [9].

Justin Zhan, Xing, and Tooth. "Opinion investigation utilizing item. audit information." 2015's Journal of Huge Information 2.1.

The study addresses core issues in opinion analysis and extreme sentiment classification using online survey data from Amazon.com. SVM, Random Forest, and Naive Bayes are used, along with the Scikit-Learn tool.

Akshay Amolik, et al. "Twitter feeling examination of film audits using AI methods." Global Design and Innovation Journal, version 7.6, 2016.

The research focuses on Twitter sentiment analysis for upcoming Bollywood and Hollywood films, using Guileless Bayes and SVM classifiers. SVM outperforms Naive Bayes in accuracy, but Naive Bayes is better in terms of precision and recall.

M. Bouazizi and T. Ohtsuki. "Multi-class feeling examination using the SENTA gadget." The sentiment score is calculated based on the viewpoint with the highest score, providing a more specific analysis of tweets with multiple opinions.

Geetika Gautam and Divakar Yadav. "Client audit grouping using twitter dataset." The study utilizes AI-based algorithms like Naive Bayes, SVM, and Maximum Entropy for customer review classification, achieving better results with Naive Bayes and SVM combined with semantic understanding and WordNet.

## 3. Methodology

However Gathering the information isn't extremely basic assignment. We really do need to think about such countless focuses while gathering the information. So in our proposal we will gather the dataset for preparing, testing and for opinion examination. This section concentrate on comprise how information will gather, how information will handled , put away and essentially how to order those information. Prior to continuing on toward this thing we should do conversation about proposed design.

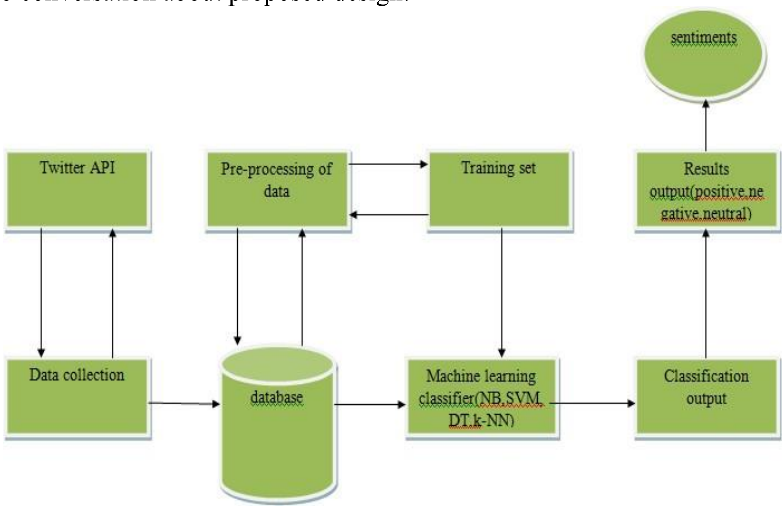



Figure 1: Flow Chart of Proposed Method

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.02, I.02(2023)
**https://www.tejasjournals.com/**

## 3.1. Proposed Methodology

As said before, our goal is to do an opinion analysis of material from Twitter. We will create a classifier by using several AI classifiers. When everything is ready, we will take several steps to conduct a feeling evaluation, as shown in the following outline:

## 3.2. Data Extraction

The researchers use the Python Tweepy library, which interfaces with Twitter's API, to download tweets. They utilize the Twitter Streaming API for real-time data updates and responses.

## 3.3. Data Collection

A Twitter account is required to access the Twitter API. After registering, the researchers obtain client keys, access token keys, client secret keys, and access secret keys, which are essential for retrieving data from Twitter.

## 3.4. Pre-processing

The collected tweets are pre-processed to prepare them for sentiment analysis. Emojis, abbreviations, duplicates, retweets, hashtags, and URLs are removed or replaced with appropriate terms. Slang and stopwords are eliminated to focus on meaningful content.

## 3.5. Sentiment Analysis

The pre-processed data is fed into a new form classifier, which categorizes the tweets into positive, negative, and neutral sentiments. The accuracy of the classifier's predictions is also indicated.
Python, as a strong and versatile programming language, is used throughout the research. The Tweepy library facilitates data extraction from Twitter. Pre-processing tasks, such as removing emojis and stopwords, are carried out using Python's libraries.

Table 1: Removed and Modified Content

| Content | Action |
|---|---|
| Punctuation(!,?,.,") | Removed |
| #word | Removed #word |
| ULR's and web links | Remove URLs or replaced with "URL" and then added in stop words |
| Number | Removed |
| Word not starting with alphabets | Removed |
| All Word | Stemmed all word (Converted into simple form) |
| Stop words | Removed |
| Emoticons | Replaced with respective meaning |
| White spaces | Removed |

The study focuses on sentiment analysis related to "gaganyaan," and Twitter data related to this topic is gathered using Tweepy. The researchers use AI-based algorithms to categorize tweets into sentiment groups. The findings of this research can have implications in understanding public sentiments and emotions related to the subject matter.

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.02, I.02(2023)
**https://www.tejasjournals.com/**

The research involves extracting Twitter data using the Tweepy library and pre-processing the tweets to make them suitable for sentiment analysis. AI-based algorithms are employed to categorize the tweets into positive, negative, and neutral sentiments. The study's findings can provide valuable insights into public opinions and emotions related to the topic of interest.

After information has been cleaned, the next step in the process is to classify the data using a classifier. So, we'll use a classifier for directed learning.

### 3.6. Classification of machine learning classifier

To categorise tweets into parallel classifications that are good and negative, we created a classifier that consists of several types of controlled learning classifiers. Scikit-learn, a classifier, was built with the help of the Python library. Python's scikit-learn package is used to do AI tasks. In addition to being an open source library, it was built on well-known libraries like numpy, scipy, and matplotlib. Scikit includes excellent documentation and support, as well as several adjustment boundaries. It therefore has a wide range of tools for categorization, perception, relapse, bunching, and so on. We can integrate Scikit-Learn into our system using a simple Python instruction, like "pip."
introduce scikit-learn'.
There are a few classifier goes under the scikit-learn. Some of them we will make sense of underneath:

- Innocent bayes(NB)
- Support vetor Machine(SVM)
- Choice Tree(DT)
- K-closest neighbor(k-NN)

## 4. Result and Analysis

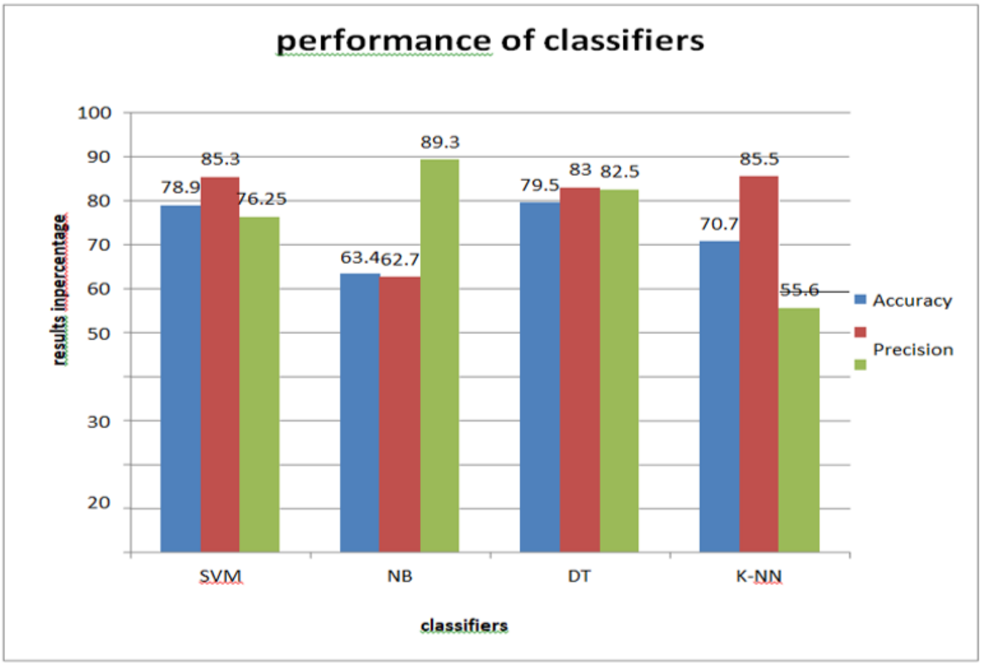
We collected a total of 1431 tweets from Twitter with the use of the Tweepy Programming interface, and we conducted sentiment analysis on those tweets using a controlled learning classifier, such as the support vector machine, k-closest neighbour, credulous bayes, and choice tree. These classifiers can help us locate common metrics like exactness, correctness, and review. The table below shows the results of these classifiers' operations.

Table 2: results of classifier on “Gaganyaan” dataset

| Supervised learning techniques | Accuracy | Precision | Recall |
|---|---|---|---|
| SVM | 78.9 | 85.3 | 76.25 |
| Naïve Bayes(NB) | 63.4 | 62.7 | 89.3 |
| Decision tree (DT) | 79.5 | 83.0 | 82.5 |
| k-NN | 70.7 | 85.5 | 55.6 |

We conducted a relative analysis to identify the most common tactics among all the ways. It is discovered that DT has the most exactness of all, at 79.5%, whereas innocent Bayes has the lowest exactness of all, at 63.4%. We may infer that DT is the best method for determining the precise result. Moving on to accuracy, K-NN obtained 85.5 accuracy, which is the highest of all. That suggests that K-NN produces results that are far more meaningful than the unimportant ones. Additionally, Credulous Bayes received 89.3% reviews, the highest rating of any classifier, indicating that our classifier successfully delivered the majority of the significant results.

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.02, I.02(2023)
https://www.tejasjournals.com/




## 4.1. Results of four weeks of twitter data

We will recapitulate the results of each and every week here. Our classifier divided the data into categories like political, neutral, and positive. Additionally, total the tweets in their entirety.
Positive tweets indicate out that people have shared their perspectives on a subject, whilst negative tweets point out that people have shared their perspectives on the opposite of the subject. While a nonpartisan tweet refers to a combined analysis.

**First week results of classified data**

| Week 1 | | | | |
|---|---|---|---|---|
| Date | Negative tweet | Neutral tweet | Positive tweet | Total tweet |
| 15/08/18 | 15 | 42 | 312 | 369 |
| 16/08/18 | 10 | 104 | 168 | 272 |
| 17/08/18 | 02 | 63 | 85 | 150 |
| 18/08/18 | 03 | 09 | 04 | 16 |
| 19/08/18 | 02 | 21 | 18 | 41 |
| 20/08/18 | 02 | 13 | 01 | 16 |
| 21/08/18 | 02 | 03 | 01 | 06 |
| Total | 36 | 255 | 589 | 870 |

The data from the first week shows that a total of 870 tweets were collected, of which around 255 were impartial, about 255 were neutral, and about 589 were positive. In which on the first date, our classifier separates 15 unfavourable, 42 neutral, and 312 favourable tweets. These things demonstrate that people were more curious on first dates. Additionally, they provide an answer for the topic. The good overall evaluation of the most recent week shows that people are in favour of our mission.

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.02, I.02(2023)
**https://www.tejasjournals.com/**

**Second week results of classified data**

| Week 2 | | | | |
|---|---|---|---|---|
| Date | Negative tweet | Neutral tweet | Positive tweet | Total tweet |
| 22/08/18 | 02 | 01 | 01 | 04 |
| 23/08/18 | 03 | 04 | 01 | 08 |
| 24/08/18 | 05 | 11 | 04 | 20 |
| 25/08/18 | 02 | 03 | 05 | 10 |
| 26/08/18 | 05 | 08 | 10 | 23 |
| 27/08/18 | 10 | 22 | 21 | 53 |
| Total | 27 | 49 | 42 | 118 |

The data from the first week shows that a total of 870 tweets were collected, of which around 255 were impartial, about 255 were neutral, and about 589 were positive. In which on the first date, our classifier separates 15 unfavourable, 42 neutral, and 312 favourable tweets. These things demonstrate that people were more curious on first dates. Additionally, they provide an answer for the topic. The good overall evaluation of the most recent week shows that people are in favour of our mission.

**Third week results of classified data**

| Week 3 | | | | |
|---|---|---|---|---|
| Date | Negative tweet | Neutral tweet | Positive tweet | Total tweet |
| 29/08/18 | 09 | 11 | 20 | 40 |
| 30/08/18 | 05 | 12 | 16 | 33 |
| 31/08/18 | 05 | 13 | 03 | 21 |
| 01/09/18 | 03 | 03 | 12 | 18 |
| 02/09/18 | 04 | 03 | 02 | 09 |
| 03/09/18 | 08 | 23 | 20 | 51 |
| 04/09/18 | 03 | 04 | 02 | 09 |
| Total | 37 | 69 | 76 | 181 |

Third week records include more confident tweets. This week, many have expressed a more positive attitude on "gaganyaan." In any event, there has been an increase in the number of unfavourable tweets since the second week. This week, we saw more positive tweets than the previous week, but far more negative tweets than the prior week. We repeatedly see that the unfavourable tweets from the third week are essentially the same. However, the favourable tweets started to flip dramatically at this moment. On a neutral tweet, 69 people have contributed their varying viewpoints. Finally, towards the conclusion, we sent 181 tweets, which is more than the next week but significantly fewer than the first.

**Fourth week results of classified data**

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.02, I.02(2023)
**https://www.tejasjournals.com/**

| Week 4 | | | | |
|---|---|---|---|---|
| Date | Negative tweet | Neutral tweet | Positive tweet | Total tweet |
| 05/09/18 | 04 | 04 | 05 | 13 |
| 06/09/18 | 12 | 15 | 13 | 40 |
| 07/09/18 | 03 | 23 | 51 | 77 |
| 08/09/18 | 06 | 13 | 10 | 29 |
| 09/09/18 | 03 | 01 | 03 | 07 |
| 10/09/18 | 03 | 01 | 05 | 09 |
| 11/09/18 | 07 | 03 | 23 | 33 |
| 12/09/18 | 04 | 02 | 12 | 18 |
| 13/09/18 | 05 | 03 | 07 | 15 |
| 14/09/18 | 10 | 10 | 20 | 40 |
| Total | 57 | 75 | 149 | 281 |

Table 5.5 results report of classified data

This week's statistics are extremely obvious. A total of 281 tweets were received. 149 of those were uplifting tweets. Additionally, we received 57 and 75 tweets that were either unfavourable or neutral. Compared to the second and third weeks, there has been an increase in favourable tweets this week. And compared to the second and third weeks, more people have tweeted negatively. 75 individuals have voiced a variety of opinions

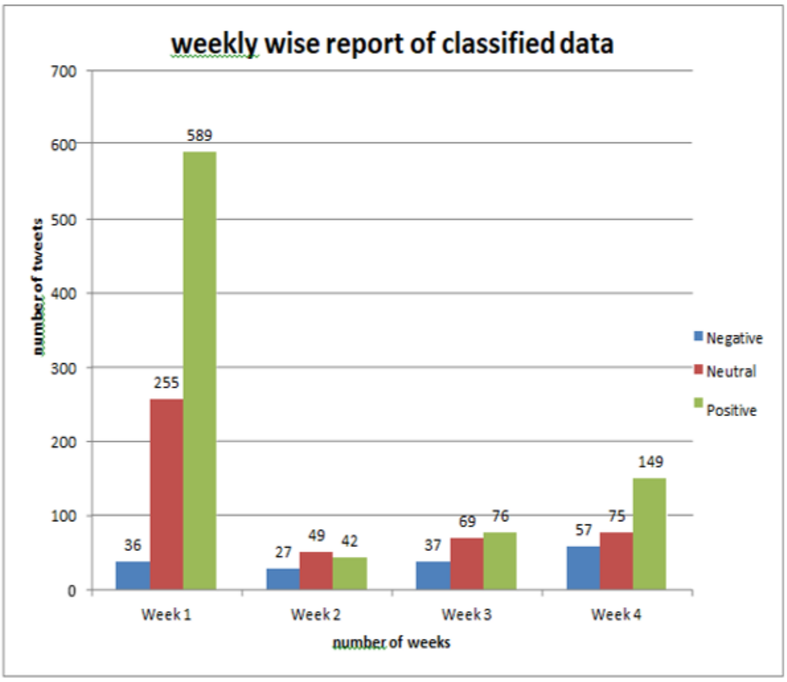



Figure 3: weekly wise report of classified data

This graph shows the relationship between the number of tweets and the number of weeks. Which demonstrates that people have tweeted a certain number of times per week. similar to the quantity of unbiased tweets, both favourable and negative.

# 5. Conclusion

In conclusion, sentiment analysis has proven to be a valuable tool for understanding public opinions and emotions expressed through various communication channels, especially on platforms like Twitter. The research has successfully employed AI algorithms like Decision Trees for sentiment analysis, which can be beneficial in gauging public attitude and sentiment towards specific projects or topics.
Future challenges in sentiment analysis include improving classifier architecture to handle phrases with various levels of relevance, expanding the categories used for sentiment classification, and incorporating image processing techniques to analyze visual content in tweets. Web application development can also enhance the accessibility of sentiment analysis, allowing users from different fields to analyze opinions more easily.
Managing vagueness in statements and extending sentiment classification to capture a wider range of emotions can lead to more comprehensive insights. Multimodal sentiment analysis, which combines various information modalities like text, images, sound, and videos, can further enrich the understanding of public sentiment.
Continuous sentiment monitoring can enable quick responses and adaptability to dynamic shifts in public sentiment, benefiting organizations in responding to customer feedback and concerns promptly. As technology continues to advance, sentiment analysis will become more refined and offer deeper insights into public opinion, making it an indispensable tool in various fields like marketing, customer feedback analysis, and social research.

# 6. References

[1]. Gurkhe D., Pal N. and Rishit B. "Effective Sentiment Analysis of Social Media Datasets using Naïve Bayesian Classification." (2014).
[2]. Bouazizi, M., Ohtsuki, T.: Multi-Class Sentiment Analysis in Twitter: What if Classification is Not the Answer. IEEE Access. 6, 64486-64502 (2018).
[3]. Gautam, G., Yadav, D.: Sentiment analysis of twitter data using machine learning approaches and semantic analysis. 2014 Seventh International Conference on Contemporary Computing (IC3). (2014).
[4]. Amolik, Akshay, et al. “Twitter sentiment analysis of movie reviews using machine learning techniques.” International Journal of Engineering and Technology 7.6 (2016): 1- 7.
[5]. Mukherjee S., Malu A., Balamurali A.R, Bhattacharyya P.“TwiSent: A Multistage System for Analyzing Sentiment in Twitter”.
[6]. Davidov D., Tsur O., Rappoport A.” Enhanced Sentiment Learning Using Twitter Hashtags and Smileys”.
[7]. Neethu, M., Rajasree, R.: Sentiment analysis in twitter using machine learning techniques. 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). (2013).
[8]. Pulkit Garg, Himanshu Garg, Virender Ranga “Sentiment Analysis of the Uri Terror Attack Using Twitter” International Conference on Computing, Communication and Automation (ICCCA2017).
[9]. Prof. SudarshanSirsat, Dr.Sujata Rao, Dr.BhartiWukkadada” Sentiment Analysis on Twitter Data forproduct evaluation” IOSR Journal of Engineering (IOSRJEN) ISSN (e): 2250-3021, ISSN (p): 2278-8719PP 22-25.(2019)
[10]. Hetu Bhavsar, Richa Manglani” Sentiment Analysis of Twitter Data using Python” International Research Journal of Engineering and Technology (IRJET) Mar 2019e-ISSN: 2395-0056 p-ISSN: 2395-0072
[11]. Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirband” Machine Learning- Based Sentiment Analysis forTwitter Accounts” 2018 by the authors. Licensee MDPI, Basel, Switzerland.
[12]. Sahar A. El_Rahman, Feddah Alhumaidi Al Otaibi ,Wejdan Abdullah AlShehri “ Sentiment Analysis of Twitter Data”.
[13]. [India announces first manned space mission". Bangalore: BBC News. ^ Press Trust of India (25 April 2012). "Spaceflight stuck due to budget: CAG". Times of India. New Delhi. Retrieved 11 June2013.
[14]. Press Trust of India. "Human space flight mission off ISRO priority list". Retrieved 18 August 2013.
[15]. https://indianexpress.com/article/what-is/what-is-gaganyaan/
[16]. Priyadarshi, Siddhanta (23 February 2009). "Planning Commission Okays ISRO Manned Space Flight Program". Indian Express. p. 2.
[17]. Beary, Habib (27 January 2010). "India announces first manned space mission". Bangalore: BBC News.
[18]. E. Loper and S. Bird, “NLTK: the Natural Language Toolkit”, Proc. ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics ,vol. 1,pp. 63-70, 2002.
[19]. P. Pang and L. Lee, “Opinion Mining and Sentiment Analysis. Foundation and Trends in Information Retrieval”, vol. 2(1-2), pp.1-135, 2008.
A. McCallum and K. Nigam, “A comparison of event models for Naive Bayes text classification”, Proc. AAAI/ICML-98 Workshop on Learning for Text Categorization, pp. 41-48, 1998
[20]. G. Kontaxis, I. Polakis, S. Ioannidis, and E.P. Markatos. Detecting social network profile cloning. In Pervasive

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.02, I.02(2023)
**https://www.tejasjournals.com/**

Computing and Communications Work- shops (PERCOM Workshops), 2011 IEEE International Conference on, pages 295300. IEEE, 2011.

[21]. Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. ”Sentiment analysis algorithms and applications: A survey.” Ain Shams Engineering Journal (2014).

[22]. H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B.Y. Zhao. Detecting and characterizing social spam campaigns. In Proceedings of the 10th annual con- ference on Internet measurement, pages 3547. ACM, 2010.

[23]. Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu. The socialbot network: when bots socialize for fame and money. In Proceedings of the 27th Annual Computer Security Applications Conference, pages 93102. ACM, 2011.

[24]. Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schtze. Introduction to information retrieval. Vol. 1. Cambridge: Cambridge university press, 2008.

[25]. Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques.” Proceedings of the ACL-02 conference on Empirical methods in natural language processing- Volume 10. Association for Computational Linguistics, 2002.

[26]. Sharma, Anuj, and Shubhamoy Dey. ”A comparative study of feature selection and machine learning techniques for sentiment analysis.” Proceedings of the 2012 ACM Research in Applied Computation Symposium. ACM, 2012

.