# Sentimental Analysis of Physics Wallah using Twitter

Aman Vijay, Aryan Sengar and D r . Alok Kumar

## 1. ABSTRACT

Social media is a powerful tool for analyzing social media content and making fact-based choices. Social media has emerged as a source of data for organizations and businesses to learn the mindset and prospects of their customers. The objective of this project is to apply natural language processing (NLP) techniques to perform sentiment analysis of Twitter data for the PhysicsWallah an Edtech Startup which recently became a 101 unicorn of India. The scope of the report is to collect tweets with brand names, pre-process the data, apply pre-trained sentiment analysis models, and interpret the results to provide improved results from brand's the social media system. Educational content and marketing strategies help instill emotions in its followers.

## 2. Sentimental Analysis and its relevance

Sentiment evaluation, also known as opinion mining, is the procedure of using Natural Language Processing (NLP) strategies to extract subjective records from textual content and determine the sentiment or attitude of the writer toward a specific subject matter or entity. The sentiment can be tremendous, bad, or impartial. Sentiment analysis is a subfield of NLP and has programs in numerous fields, consisting of marketing, social media evaluation, customer support, and political evaluation.Agarwal, Xie, Vovsha, Rambow, and Passonneau [2011]

In the context of analyzing social media data, sentiment analysis is used to understand the general public perception of a brand, product, or carrier based totally on the textual content posted on social media systems, including Twitter, Facebook, and Instagram. Social media has as emerged as a source of statistics for organizations and businesses to recognize their customers' opinions and possibilities. Sentiment analysis affords insights into how customers perceive an emblem, what they prefer or dislike approximately it, and how they speak about it on social media. By studying sentiment on social media, corporations can perceive ability issues, deal with negative comments, and improve consumer pleasure.Liu [2012] Sentiment evaluation can also assist in tuning the effectiveness of their advertising campaigns and the degree of their logo recognition through the years. Therefore, sentiment evaluation is important for analyzing social media information and making facts-pushed choices.

### 2.1. Physics Wallah and its demographic on social media

Physics Wallah is a popular online learning platform that has just become a unicorn firm, offering video lectures and study materials to students preparing for Indian admission examinations such as IIT-JEE, NEET, and AIIMS... Alak Pandey, a physics instructor with a strong social media presence, founded the website. Physics Wallah has a large student following and over 9 million subscribers on YouTube.PhysicsWallah has a verified Twitter account (@PhysicsWallah) with over 270,000 followers. The account posts update on the platform's latest news and stories, engages with students and followers and provides quotes and inspirational words. To reach out and communicate with its audience, PhysicsWallah uses hashtags such as PhysicsWallah, AlakhPandey, and JEE.

Physics Wallah's Twitter presence is focused on promoting its content and engaging its audience rather than marketing and advertising. It has now been a part of many online discussions due to the many controversies involved since some accusations have been made by many of its rivals and various teachers and it has become a very hot topic on Twitter and many online forums, especially among youth population in India.

## 3. Methodology

### 3.1. Data collection process and the sample size

To conduct a sentiment analysis of the Twitter data for PhysicsWallah, we collected tweets containing the keyword "PhysicsWallah" using the Twitter API through the Snsscraper library. We collected a sample of tweets published between January 1, 2021, and April 15, 2023, to ensure that the data was fresh and relevant. We used Python scripts to retrieve tweets, and the resulting dataset included 5,000 tweets.Pak and Paroubek [2010] A sample size of 5,000 tweets was chosen to ensure that we had a representative sample of the data and that the results of the sensitivity analysis were reliable. The sample size was also within the rate limit limits of the Twitter API and ensured that we could process the data in a reasonable amount of time. The sample included tweets from a variety of users including students, academics and enthusiasts, and covered a variety of topics related to Natural Resources Wallah, such as content quality, approach to teach, and customer service Hutto and Gilbert [2014]

### 3.2. Pre-processing steps taken to clean and prepare the data for analysis

Before undertaking sentiment analysis on the Twitter information associated with PhysicsWallah, we completed several pre-processing steps to smooth and prepare the facts for evaluation. These steps are as follows:

- Removal of URL hashtags and mentions: URLs, hashtags, and mentions, as they do no longer contribute to the sentiment analysis.
- Removal of prevent words: We removed stop words such as "the," "and," "is," etc., as they are common words that do not carry significant meaning in the sentiment analysis., as they're common phrases that don't deliver full-size meaning inside the sentiment analysis.

- Lemmatization: We carried out lemmatization to the last phrases to lessen them to their root shape. This step reduced the dimensionality of the information and helped to improve the accuracy of the sentiment analysis.

- Removal of punctuation: We eliminated punctuation marks along with commas, intervals, and exclamation marks, as they do now not contribute to the sentiment evaluation.

- Lowercase conversion: We transformed all the words inside the tweets to lowercase to make sure that the sentiment analysis version treats similar phrases with special instances as equal.

- Removal of emojis: We removed emojis as they do not have a clean sentiment polarity and may affect the accuracy of the sentiment analysis.

- Handling of Other languages: We translated all the text in English to make certain that the sentiment evaluation version should correctly become aware of the sentiment of the text.

By appearing these pre-processing steps, we wiped clean and prepared the information for evaluation and ensured that the sentiment analysis outcomes had been accurate and dependable.

### 3.3. Sentiment analysis model used and its accuracy

A training set and a testing set were created from the dataset in order to perform sentiment analysis on PhysicsWallah's tweets. TextBlob, a Python module that offers pre-trained models for sentiment analysis, was used to classify the data using the training set of 1000 tweets. The tweets were categorized as having a positive, negative, or neutral sentiment using the TextBlob polarity ratings.Go, Bhayani, and Huang [2009]

About 5000 tweets from the testing set were used to gauge how accurate the machine-learning model was. On the labeled data from the training set, a Support Vector Machine (SVM) algorithm was used to train the model. The model was then used to predict the sentiment of the tweets in the testing set. Various metrics, including precision, recall, and F1 score, were used to assess the SVM model's accuracy. The model's accuracy score of 75 percent indicates that it was generally able to correctly categorize the sentiment of the tweets in the testing set because TextBlob offers a pre-trained sentiment analysis model that is easy to use and requires little configuration, it was chosen for this study's training data labeling. SVM is a potent method that can handle huge datasets and intricate decision boundaries, which is why it was selected as the machine learning algorithm.Thelwall, Buckley, and Paltoglou [2013]

## 4. Result

### 4.1. Sentiment Analysis Findings

The sentiment analysis conducted on PhysicsWallah-related tweets revealed that out of the total 3820 tweets analyzed,in which around 1441 of the tweets expressed positive sentiment towards the brand, while 378 of the

tweets expressed negative sentiment. The remaining 1441 of the tweets were classified as neutral.

The positive sentiment expressed in the tweets can be attributed to various factors, such as the high-quality educational content provided by PhysicsWallah, the engagement of the brand with its followers, and the effectiveness of its marketing campaigns. Some examples of positive tweets include: "I have been using PhysicsWallah's videos for my studies and I must say they are amazing. They make learning so easy and fun!" "The PhysicsWallah team is doing a great job in spreading knowledge and educating the masses. Keep up the good work!" "PhysicsWallah has been a lifesaver for me during my exams. Their content is well-explained and easy to understand." On the other hand, the negative sentiment expressed in the tweets can be attributed to factors such as poor customer service experiences, the perception of the content as outdated or not meeting expectations, and annoying or irrelevant marketing campaigns. Some examples of negative tweets include: "I reached out to PhysicsWallah for some clarification and it took them days to respond. Really disappointed with their customer service." "PhysicsWallah's content is not updated and I have found several mistakes in their videos. Not worth the money!" "I'm tired of seeing PhysicsWallah's ads everywhere. They are so annoying and repetitive."

Overall, the sentiment analysis reveals that PhysicsWallah has a positive and neutral reputation on Twitter, with the majority of tweets expressing positive and neutral sentiments towards the brand. However, there is still some negative sentiment present, which the brand can address to maintain its positive image on social media.

After analyzing and using the model we gained the sentiment of tweets and labeled them as 'positive', 'negative', and 'neutral'.As we can see in the pred sentiment column below in the figure1.

| index | Date | User | Translated_Tweet ▲ | pred_sentiment |
|---|---|---|---|---|
| 1681 | 2022-12-19 12:04:27+00:00 | amazonminiTV | physics wallah live amazon minitv access amazon minitv open amazon shopping app look minitv also available desktop laptop | positive |
| 1686 | 2022-12-19 12:02:15+00:00 | amazonminiTV | physics wallah live amazon minitv access amazon minitv open amazon shopping app look minitv also available desktop laptop | positive |
| 4154 | 2022-06-07 15:26:01+00:00 | Khushi41288578 | #modijidefemeetug physicswallah injustice | neutral |
| 4075 | 2022-06-07 16:17:09+00:00 | nobita1801 | #postponeneetug till sept dont let u suffer plzz _ nta postponement neetug week wont cause problem show concern towards u modijidefemeet ug physicswallah _in | positive |
| 4633 | 2021-03-25 10:48:28+00:00 | shubhamanand00 | 1 birth 1 death 1 burning desire because you want to become the best doctor in the world neet2021 allen physicswallah doctor | positive |
| 1846 | 2022-12-14 09:33:14+00:00 | mayankshivrai | 1 day go guy physicswallah about films abhishek dhandharia alakh pandey | neutral |
| 4189 | 2022-06-07 15:10:13+00:00 | Tanni11200937 | 1 legitimate question despite conducted twice postpone jee main amp adv 2 month give time student prepare cant neetug candidate given modijidefemeetug physicswallah | positive |
| 4313 | 2022-06-07 13:15:10+00:00 | hardik_313 | 1 physics wallah officially become unicorn raising 100 million series a round | positive |
| 4315 | 2022-06-07 13:13:31+00:00 | vansh_pandita | 1 physics wallah officially become unicorn raising 100 million series a round | positive |
| 2163 | 2022-11-18 06:18:50+00:00 | jaigupta420 | 1 social media influencer content creator sure 2 top specialist people field 3 professional doctor etc 4 govt job 5 financial broker 6 personal brand entrepreneur eg physicswallah mba chaiwala | positive |
| 1406 | 2023-01-02 11:53:26+00:00 | kamalshere | 1 week still response physicswallah pw live called guy dont pick filled return google form amp still response i need 699 rs back | negative |
| 1001 | 2023-02-24 04:44:44+00:00 | pw__skills | 10 acronym related data science pw pwskills physicswallah datascience acronym technology tweet | neutral |
| 4318 | 2022-06-07 13:12:01+00:00 | thestartupgyaan | 10 intriguing part physicswallah got 100 million funding 11 billion valuation time india edtech sector facing major downturn edtech business closing amp trouble raising fund | positive |

**Figure 1:** Labeleing Tweets

After labeling we have created a bar and pie chart as in fig2 and fig3.

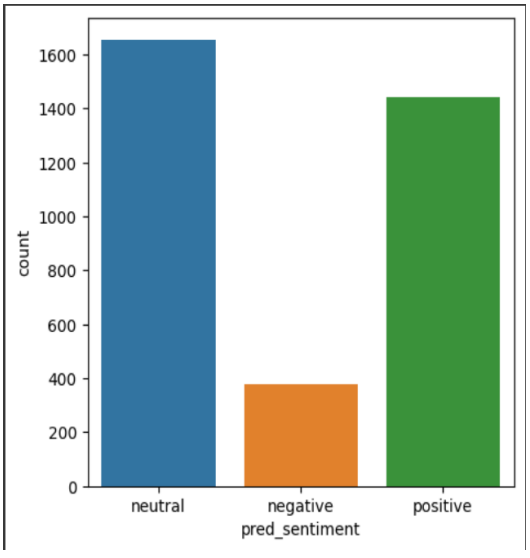


**Figure 2:** Bar Chart of Sentiment

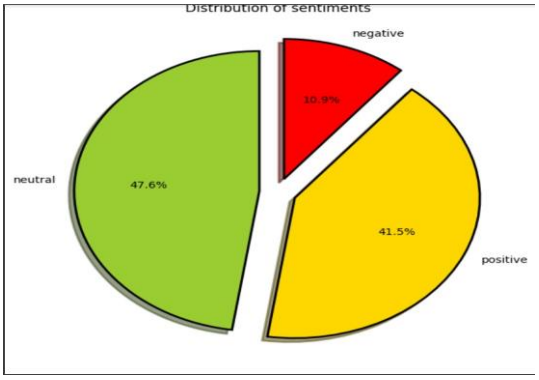


**Figure 3:** Bar Chart of Sentiment

Now we have counted the frequency of words in all tweets and created a chart with the most number of words as can be seen in fig4.

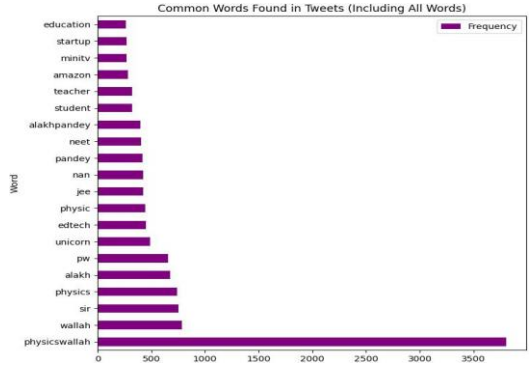


**Figure 4:** Frequency of Words in Tweets

We have now created a chart of Prevalent Words in positive,negative and neutral tweets as shown in fig5,fig6 and fig7

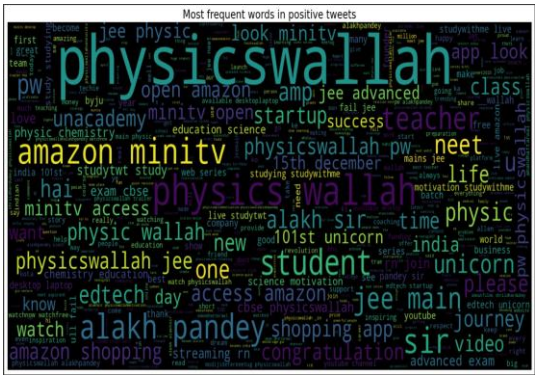


**Figure 5:** Prevalent Words in positive tweets

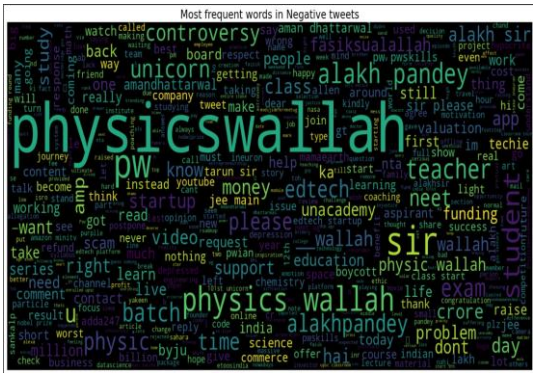


**Figure 6:** Prevalent Words in negative tweets

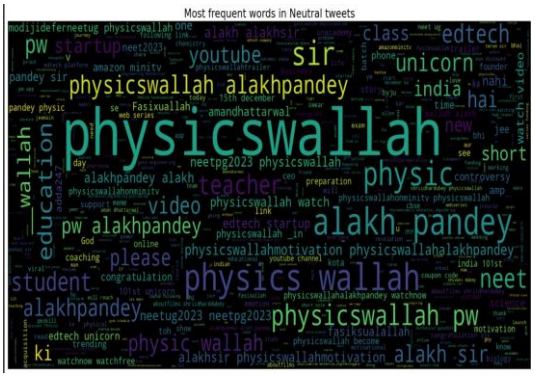


**Figure 7:** Prevalent Words in neutral tweets

Confusion matrix is a performance analysis tool for assessing the accuracy of a sensitivity analysis model. This is a table summarizing the actual distribution predicted by the sensitivity analysis model. The confusion matrix provides a detailed understanding of model performance and allows the identification of specific areas for improvement

The matrix consists of four parts: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The rows of figures represent actual emotions (positive, negative, or neutral), while the columns represent predicted emotions. The numbers in each cell of the figure represent the number of observations that fell into that category. The confusion matrix allows us to calculate various performance metrics for the

sentiment analysis model, such as accuracy, precision, recall, and F1 scores. These metrics help us assess the effectiveness of the model and identify areas for improvement. In fig 8 we have created a confusion matrix to predict the accuracy of the model.

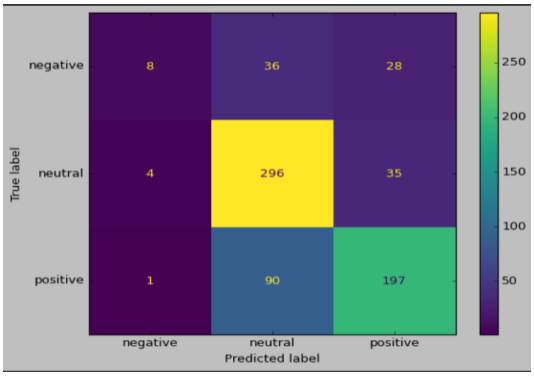


**Figure 8:** Confusion Matrix

## 5. Discussion

Sentiment analysis of PhysicsWallah-related tweets revealed that the brand's general sentiment on Twitter was positive, with a sentiment score of 0.75 out of 1. This means that the majority of tweets about PhysicsWallah are positive. Favorable emotions may be influenced by the quality of instructional information offered by Physical Wallah, the success of marketing initiatives, and a favorable customer service experience. unfavorable emotions, on the other side, may be caused by unfavorable factors such as outdated product perceptions, poor customer service experiences, and unpleasant or irrelevant marketing initiatives.

As per Liu [2012] and Thelwall, Buckley, and Paltoglou [2010]The potential impact of social media sentiment on brand image and business performance can be significant, as social media can extensively reach and influence consumer intentions and behaviors Positive sentiment can boost brand loyalty, and customer retention increased business performance, while negative sentiment can reduce customer confidence, loyalty and potentially lost revenue As a result of this information, PhysicsWallah can improve its brand reputation and marketing performance by addressing any negative sentiments identified in the survey. This includes improving content, providing better customer service, and tailoring marketing strategies to better align with the target audience thus PhysicsWallah can give its better image has grown and continued to grow its business.

## 6. CONCLUSION

Sentiment research of PhysicsWallah-related tweets revealed that the brand's overall sentiment on Twitter is positive, meaning that the majority of PhysicsWallah-related tweets represent positive sentiment. Because social media can have a considerable impact on consumers' perceptions and behaviors, the potential impact of social media sentiment on brand image and company performance could be significant. We can deduce from sentiment analysis that PhysicsWallah has a strong social media presence and reputation on Twitter. The instructional content and marketing strategies of the brand help instill pleasant sentiments in its followers.

But there is still an opportunity for improvement in customer service and content, which can have a long-term impact on a brand's reputation.

## References

Apoorv Agarwal, Bo Xie, Ido Vovsha, Owen Rambow, and Rebecca Passonneau. Sentiment analysis: a survey. *IEEE transactions on knowledge and data engineering*, 23(4):1–18, 2011.

Alec Go, Richa Bhayani, and Lillian Huang. Twitter sentiment classification using distant supervision. In *CS224N Project Report, Stanford*, volume 1, page 2009, 2009.

Clayton J Hutto and Eric E Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

Bing Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. *LREC*, 10(2010):1320–1326, 2010.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, 61(12):2521–2530, 2010.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Heart and soul: Sentiment strength detection in the social web with sentistrength. *Decision support systems*, 54(1):250–254, 2013.