Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Sentiment Analysis to Compare Tweets About Online Education During and After COVID-19

[1]Nishita Sharma, [2]Nandni Das Singhal, [3]Nidhi Panghal, [4] Nitish Kumar Sengar, [5]Meharban Ali

[1,2,3,4,5]Department of CSE, MIET, Meerut

Abstract: COVID-19 affected the world drastically in various terms. It has been transmitted throughout the globe with infectious air whose impacts were not assumed or expected. The education system faced many challenges due to it and suffered a lot but the rise of online education gave people, institutions, and schools another hope to continue their learning by being connected and conducting classes over the internet. In this paper, we seek to study and compare the responses of people from the community during and after COVID-19 about online education learning by analysing the tweets from the social media platform called Twitter. For classification, Naive Bayes algorithm has been used and an accuracy of 85.4% has been achieved. By this study, we concluded that during COVID-19 most of the people had negative opinions about online education, but the after COVID-19 datasets show a certain rise in the figures of neutral and positive tweets and a decrease in the number of negative tweets.

Keywords – sentiment analysis, COVID-19, education, online learning, tweets, Natural Language toolkit, Naive Bayes

1. Introduction:

Sentimental Analysis is considered one of those fields that makes sense out of data that is in the form of text and also includes opinions that can be negative, positive, or neutral. Examining the sentiments that are conveyed on social platforms enables analysts to examine thoughts and opinions that are often not voiced out formally. This technique is the reason for the increasing use of e-learning, online text, and microblogging [8] platforms which include Facebook, WhatsApp, Twitter, and even online blogs. Social media provides people with a platform where they express their opinions.

The landscape of education went through significant changes during COVID-19 which emphasized the necessity of distance learning and so it became important to know how students and faculty are reacting to this using sentimental analysis. In this process, tweets of Indian users are collected in two different scenarios i.e. during and after the pandemic. The objective of the study is to find issues between mentor teaching quality and student learning by this we can help the institutions and administration. The textual comments are preprocesses using Natural Language Processing (NLP) techniques such as feature extraction and selection for further analysis.

As we know that sentimental analysis is a method of using algorithms that convert various text-related samples into positive, neutral, and negative sentiments [9]. For this, we use Natural Language Toolkit which uses a text classifier feature, and also employs these codes via strong powerful predefined machine learning operations to obtain results from human language data. There are different approaches to performing sentiment analysis. The text data is pre-processed using the removal of unwanted symbols and tokenization. Our training model process gives priority to 80% focus on training or laying a robust foundation to handle the system's capability and the remaining 20% is dedicated to testing the model which ensures that the performance aligns with the correct objectives.

In this process, we opted for the use of Bar graphs for a visual representation of our datasets. The choice of using bar graphs facilitates the clear presentation of different sentiments with analyzed data that enhance the project's

overall interpretability and insights. In this project, we are working on a comparative analysis of online learning on two datasets during and after COVID-19 respectively.

2. Proposed Work Plan

2.1 General architecture/ Flowchart/ DFD of the overall system to be designed.

The approach is to study the impact and opinion of people on online education during and after COVID-19 through sentiment analysis of Twitter, including collection of data(tweets) on online education learning from Twitter and then usage of NLTK library in Python to preprocess the data. We are first removing the special characters and unwanted symbols/spaces from the data and then moving forward towards tokenization for specified assessment. Later this modified data is passed to the VADER for the polarity detection process and then the model is trained with the Naive Bayes Algorithm to predict the sentiment [12] of the data. Below is the data flow diagram.

2.2 Description of various modules of the system.

A. Data Collection

Data Collection is termed as of collection or gathering the unstructured data which is in raw form. In this study, for collecting the data we have focused on the tweets of Indian users on Twitter, which is an extensively used social media platform [1]. There are two datasets consisting of data from mid-April to August 2020 to analyse the sentiment during the COVID-19 pandemic and another from April to August 2021 for the analysis after COVID-19. As Twitter does not give free authorized access to use its data from API, to fetch the data of these periods, we used a got old tweets [11] that is available online for free, it uses urllib to gather datasets from Twitter [2]. We specified certain queries for our dataset: Tweets only in the English language were preferred, keywords used are online education, online courses, remote teaching, online learning, virtual training, and online classes, and they should be of Indian origin. Thus, we have 1501 tweets from each dataset that would help us better analyse.

B. Data Pre-processing

Coming up to the next step which is Data pre-processing. It is the step in ML in which the raw data obtained from various sources is transformed into a usable format or structured manner to implement accurate analysis of it [4]. As our gathered data consists of various unwanted symbols, emojis, punctuation marks, extra space, etc, hence there is a need to remove these unwanted things. The first stage is the removal of the unwanted spaces and punctuation marks and then comes the usage of the NLTK library to implement data tokenization [3,2,5]. Tokenization is the separation of the sequence of data in the form of text into smaller parts known as tokens [12]. It is an important process because it allows computers to understand and analyse human language to understand the emotion for it.

C. Polarity Analysis

Polarity analysis is the classification of the sentiment of a test whether it is positive, neutral, or negative. This can be achieved by a variety of techniques or methods in Python. To check the polarity in the study we used VADER, which is included in the NLTK library. VADER is defined as sentiment analysis for the data in text form, that is for the intensity of emotion. It is based on the dictionary that depicts features of emotion intensity termed as sentiment score [6]. Example: Words like ‘like’, ‘amazing’, and ‘enjoy’, all convey a positive sentiment. It is intelligent enough to understand the words like ‘did not like’ as a negative statement.

D. Training of Model

To train the model well, we have a classification process. We have used the clean data for training the Naive Bayes classifier in the study to develop a structured model. The Naive Bayes is a probability-finding machine learning model based on the theorem [10] given by Bayes. It gives fast, highly scalable real-time predictions, also it does not require much training data. The data that is tokenized will pass via the classifier model and every tokenized tweets will be differentiated with the trained dataset. We have the range for the probability of a tweet to be neutral (0.4-0.6). Hence, every tweet will be classified in the form of positive, neutral, or negative. With the

merging of the Naive Bayes classifier and test-based training, we obtain the analysis of the sentiment of each modified dataset from the collected data.

E. Testing

Testing is defined as the method of verifying and ensuring whether the model is free from bugs, fulfills the desired user requirements, and handles all exceptional cases. To test the model's accuracy, we split the labeled data into 80% of the labeled datasets as training data and 20% as data to be tested. After the application of the data that is trained and tested with the Naive Bayes algorithm, we get the precision of 85.4% in our system/study.

F. Visualization

To represent the results in the form of a graphical representation of data and information is defined as visualization. It is done to facilitate understanding, interpretation, and communication. Visualizing the results of sentiment analysis in the way of a bar chart and showing the distribution of predicted sentiments compared to the true sentiments, we have used the 'matplotlib' library. Matplotlib is a popular 2D plotting library for the Python programming language. It provides a wide range of high-quality, customizable visualizations for analysis.

3. Experimental Results and Analysis:

3.1 Description of the data set used.

For this analysis we collected tweets in the English language related to online education from Indian users between April 2020 and August 2020 as well as April 2021 and August 2021.

The keywords we used for our research were "online education", "online courses," "remote teaching," "online learning," "virtual training," and "online classes". The outcome of the study presented that there was a gradual increase in tweets related to online education in April 2020 and August 2020 as compared to the same period in 2021. We found that most tweets expressed positive sentiments about online education, with the user convenience and flexibility. However, we also found a significant number of tweets with negative sentiment, mostly related to technical issues.

We take the tweets that were separated into [15] 80.00 percent training data and a testing test of 20.0 percent, and the extraction from the tweets and approach applied to the classifiers are acquired with the accuracy of the datasets.

3.2 Calculate the efficiency of the designed system according to the parameter used to evaluate the model.

In this chapter we analyse the Naive Bayes classifier performance and accuracy with a sentimental analysis framework using Vader datasets, and the parameters taken are precision, and recall to measure the accuracy of specified method.

One of the main issues generated by using TextBlob annotated system is that if any tweet presented from the people's side is in the form of 'Yes' or 'No', instead of estimating it as '+' or '-' it tells that the collected tweet as 'neutral'. Because of this the neutral value graph also presents more percentages than the other two responses. To handle that issue we take VADER to analyse the sentiment which gives a more original outcome. To refine the result we come out of the issue it provides the maximum result scores after applying the Naive Bayes algorithm [13].

The formula for the accuracy check in Naive Bayes Theorem:
$P(t \mid x) = (P(x \mid t) * P(t))/P(x)$

where t is class, x is data, and $P(t \mid x)$ is the probability of class given the provided data [7,13,14].

The research of analysing sentiments of the Twitter dataset on a total of 1501 tweets in each timeline, regarding online education during and after the pandemic from April to August 2020 and April to August 2021 has given the analysis in the form of Table No. 1 mentioned below.

Table No.1: During COVID-19 stats

During COVID-19		
Sentiments of Emotions	No.of Tweets	Percentage
Positive	281	18.7%
Neutral	195	13%
Negative	1025	68.3%

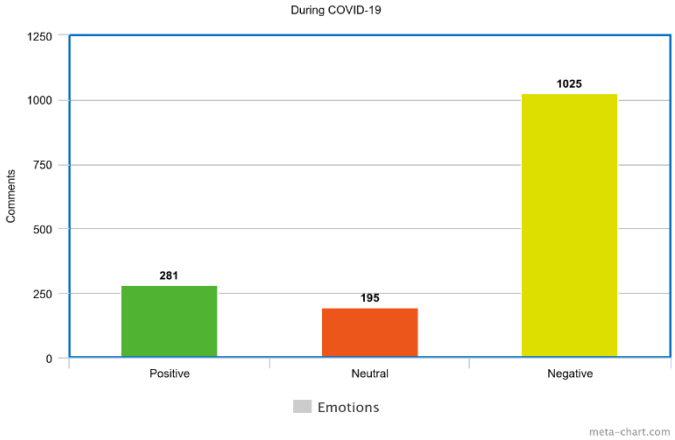

Figure No. 1: Representation of During COVID-19 stats

Figure No.1 is the graphical representation that suggests the data gathered includes a wide number of negative emotions and a slight difference between the % of positive and neutral tweets. As it is based on the dataset which is of during the months of COVID-19, people were facing challenges to cope with online learning.

Table No.2: After COVID-19 stats

After COVID-19		
Sentiments of Emotions	No.of Tweets	Percentage
Positive	299	19.9%
Neutral	1071	71.4%
Negative	131	8.7%

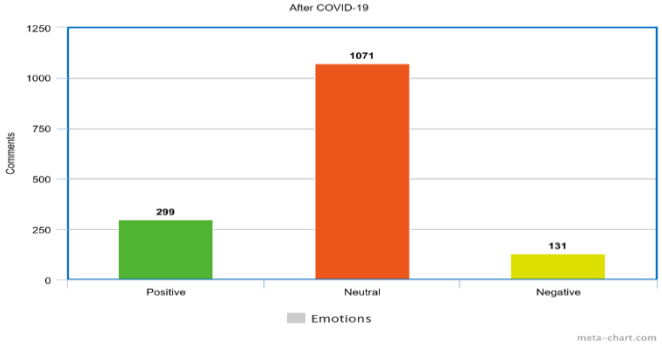

Figure No. 2: Representation of After COVID-19 stats

Now coming to the analysis of the datasets after COVID-19 which is from April 2021 to August 2021, Table No. 2 and Figure No. 2 clearly states that most people have neutral sentiments about online learning but also a slight rise in the number of positive tweets and a massive decrease in the parameter of negative tweets compared with the 2020 datasets.

4. Conclusion

The motive is to study and present a differentiated examination of people's responses during & after the pandemic. This started from a pandemic where all the educational institutions have been switched from head-to-head learning to a completely digital or online form using the Internet and gadgets. In this research, we focused on pointing out the sentiments and major topics of E-learning. In this people's tweets have been collected from a social media platform that we have taken Twitter data. It is used to judge the data on the attitudes of the students, sentiments, and false information towards COVID-19.

We have applied Naive Bayes which has provided valuable insights into the dynamic nature of sentiments towards online education. We have also used VADER which is specially designed for social media texts it is a rule-based sentiment analysis framework that can rapidly and effectively analyse the sentiments of text data by assigning the polarity to it i.e. (Positive, Neutral and Negative).

We are taking the datasets of 2 years i.e. April to August of 2020 year and another April to August of 2021 year. In the study, it was found that there were many negative statements about online learning during the timeline of April 2020 to August 2020(during COVID-19) as people were facing issues in managing their studies using the internet. After the pandemic i.e. timeline from April 2021 to August 2021 we found more positive responses as compared to negative responses in entire datasets.

During COVID-19 total recorded tweets were 1501 tweets, and 68.3% of tweets were taken down as negative tweets but after COVID-19 people understood the concept of online education and are in favour of this, so there are major changes in the negative tweets i.e. 8.7%. Therefore, this shows how people are accepting the reality of online education, but are still confused with it, as neutral tweets were noted down with 71.4%.

Life is all about learning and the pandemic has taught us that nothing is constant and humans are the ones who can adjust them according to the environment requirements. Hence, we all are in the process of learning about the new era, with new challenges, difficulties, and achievements.

References:

[1] Chintalapudi, N. Battineni, G. Amenta, F. Sentimental Analysis of COVID-19 Tweets Using Deep Learning Models. Infect. Dis. Rep. 2021.

[2] Swetha Sree Cheeti, Yanyan Li and Ahmad Haaegh "Twitter Based Sentiment Analysis of Impact of Covid-19 on Education Globally", IJAIA, May 2021.

[3] M. Umair, A. Hakim, A. Hussain and S. Naseem "Sentiment Analysis of Students' Feedback before and after COVID-19 Pandemic" International Journal on Emerging Technologies, June 2021.

[4] Abonia Sojasingarayar, "Data Preprocessing in Python", Medium website, Nov 2022.

[5] Arif Ridho Lubis, Santi Prayudani, Muharman Lubis, Okvi Nugroho, "Sentiment Analysis on Online Learning During the Covid-19 Pandemic Based on Opinions on Twitter using KNN Method", IEEE 2022.

[6] Aditya Beri, "Sentimental Analysis Using VADER", published in Towards Data Science, May 2020.

[7] AlShaikh, W. Elemedany. "Estimate the performance of applying machine learning algorithms to predict defects in software using weka", 4th Smart Cities Symposium (SCS 2021), 2021

[8] Impact of Covid-19 pandemic on education", Wikipedia, the free encyclopedia, 2021.

[9] "Sentiment analysis in aspect term extraction for mobile phone tweets using machine learning techniques" by Venkatesh Naramula and Kalaivania A., International Journal of Persvasive Computing and Communications, 2021

[10] Hardikkumar Dhaduk, "Performing Sentiment Analysis with Naïve Bayes Classifier", Analyticsvidhya, August 2022.GetOldTweets3 PyPi, Nov 27, 2019

[11] Mayur Wankhade, Annavarapu Chandra Sekhara Rao and Chaitanya Kularni "A survey on sentiment analysis methods, applications and challenges", Springer Link, February 2022.
[12] Hong Chen, Songhua Hu, Rui Hua and Xiuju Zhao "Improved Naïve Bayes classification algorithm for traffic risk management", EURASIP Journal on Advances in Signal Processing, 2021.
[13] Feng-Jen Yang "An Implementation of Naïve Bayes Classifier", IEEE, 2018.
[14] "Sentiment Analysis and Topic Modeling on Tweets about Online Education during COVID-19" by Patrick Bernard Washington, Sallem Ullah, Ernesto Lee, Septmenber 2021.