

## “Sentiments Of Social Media Users, Towards Web Education – A Data Mining Techniques.”

**Dr. Baig Muntajeb Ali**

(Associate Professor), Maulana Azad National Urdu University, Hyderabad,  
CTE – Darbhanga.

---

### **ABSTRACT**

In the 21<sup>st</sup> century, the Internet has caused a knowledge explosion, and every day quintillion bytes of data are created. Thus analyzing data and interpreting it has become an arduous task for educators. However, data analysis has become straightforward due to the advent of data mining techniques and machine learning. Recently the Corona Virus pandemic harmed the education system across the World, and the task of imparting education has been shifted from traditional face-to-face education to web education. Websites and social media in teaching and learning have created massive data. It has become a vital source of information for studying social media users' thoughts, emotions, feelings, opinions, and perceptions. Since many social media are available on Internet, this study uses the microblogging Twitter platform for sentiment analysis due its short messages and popularity. In this research, text mining is done using the Open Source Visual Programming Software Orange (3.28.0; Orange Data Mining, n.d.). The text mining technique involves four processes; in the first process, the text is extracted from posts of the Twitter platform; in the second process, the obtained data is pre-processed, and sentiments of Twitter are analyzed with the help of Orange (3.28.0; Orange Data Mining, n.d.) software. In the third step, the classification of sentiments is done using the V.A.D.E.R. (Valence Aware Dictionary and Sentiment Reasoner) modules. Finally, the obtained scores from text mining are used to analyze data. The research findings suggest that Twitter users' sentiments have moderate positive compound sentiment (.389) about web education and the use of technology in education.

Further, the compound polarity scores of 38 % of tweets about web education suggest the tweets were moderately positive and greater than zero. This study's moderate positive compound sentiments are due to the shift from traditional face-to-face (f2f) education to Web-based education during the pandemic. The educators, schools, colleges, and universities emphasized web-based teaching, learning, and evaluation during pandemics.

**Keywords:** Opinion Analysis, Sentiment Analysis, Pre-processing, Data Mining Techniques, Text Mining Techniques, Machine Learning Approach, and Social Networking Sites (S.N.S.).

### **INTRODUCTION**

People use the Internet and social media frequently in this digital age and express their emotions, feelings, attitude, and perceptions regarding events in posts and messages or tweets. Since there are myriads of social media users available on the Internet, in the present study, the Twitter platform is used for text mining and analysis of the tweets to study the sentiments of the users of the Twitter platform. On the Twitter platform, the messages are posted in tweets. They act as a vital source of information and gives valuable insights to the educators to improve the quality of teaching, content, and methodology. The tweets posted reflect the sentiments of Twitter users towards events, programs, and the quality of web education. In the tweets, the feelings shared by the users can be positive or negative; thus, it shows whether the user has positivity or negativity towards the web education system.

Further, users' perception analysis helps educators get feedback from the emotional reactions expressed by the users of the Twitter platform in improving the quality of web education. The analysis of data in the form of text is also helpful in knowing learners' performance. It empowers educators to understand the learners' problems related to teaching, learning, and evaluation in a most user-friendly way—the learners' performance utilizes text mining and machine learning to analyze the Twitter platform users' sentiments.

### **PURPOSE OF THE STUDY**

Artificial Intelligence (A.I. ), Artificial Neural Networks (A.N.N. ), Machine Learning (M.L. ), Big Data, Data Analysis, and Sentiment Analysis (S.A.) have all seen significant advancements in recent years. Therefore, recently these technologies are used in ecommerce, to study social-media users' behavior, perception, and sentiment analysis. The advent of web-based technology and social media has generated massive data. The analysis of generated data is highly beneficial to e-commerce enterprises in understanding social media feelings and opinions to make the quality of products and services better. In educational research, the use of A.I., Machine Learning, and Data Mining for studying sentiments are at nascent stage, and little research is done on it. Thus, the study aims to find users' sentiments of Twitter platforms about using web-based resources in education through text mining.

### **LITERATURE REVIEW**

The literature review covers web education, text mining, pre-processing text, machine learning, sentiments analysis, Vader, lexicon approach, and natural language processing (N.L.P.). According to Olalekan (2020), Natural Language Processing (N.L.P.) is the most common technique used in the research to analyze social media users' opinions. The polarity of tweets is established during the sentiment analysis process; this sentiment analysis method is also called "Emotion Artificial Intelligence (A.I.)" or "Opinion Mining." According to Olalekan (2020), there are two approaches in the analysis of sentiments;

#### **Lexicon-based methods of sentiment analysis.**

Lexicon-based emotion analysis is a method of detecting sentiment orientations in a corpus of documents or messages in the form of posts or Twitter messages or a group of words from the semantic orientation of lexicons. Semantic orientation can be positive, negative, or neutral in

polarity. The lexical dictionary can be built automatically or by a time-consuming manual approach in this process (Gupta & Agrawal, 2020) .

### **Machine Learning-based methods.**

Humans learn by their experiences, but machines, such as computers, are taught to learn information from a set of data using data analysis techniques. Machine learning algorithms use computational methods to learn messages or information from a data set rather than relying on a previous equation model. As the sample size and the amount of data collected grows, the algorithms' ability to learn improves (What Is Machine Learning? | How It Works, Techniques & Applications - MATLAB & Simulink, n.d.).

According to Kharde (2016), the term opinion analysis means the process of mining the attitudes, emotions, opinions, beliefs, views, and behaviors of individuals from the tweets, messages, posts, documents, and database system with the help of N.L.P. The practice of categorizing textual views into different categories such as "positive," "negative," or "neutral" is called sentiment analysis or opinion analysis, and similar meanings include subjectivity analysis, opinion mining, and assessment extraction. In his project report Edward (2018), observes that Twitter social network data can be obtained by tracking keywords and extracting data. Moreover, the pre-processing data remove noisy data such as Uniform Resource Locator (URLs) @word(s) and #hashtags symbols, training model, and classifications. However, data mining has some constraints; the data is present in the unstructured form, and text in tweets may sometimes contain wrong English words like "gud" for good; thus, it becomes challenging to extract the required information from the tweets.

The observations by Mansoor et al., (2021), on the countries that were worst hit by the COVID-19 pandemic were done by using two methods of deep learning, namely, the Long Short Term Memory (L.S.T.M.) and the Artificial Neural Networks (A.N.N.). The study aimed to explore how Coronavirus affected areas of life such as work from home and online learning. Bhagat et al., (2021) investigated people's sentiments regarding online resources in learning during pandemic COVID-19 (Coronavirus Disease 2019). Bhagat et al., (2021), has extracted 1,54 research papers and articles from Google & DuckDuckGo., in their study, the scores are calculated for finding polarity and subjectivity of extracted research articles by applying the library Text Blob. The findings show that more than 90% of the research papers and reports have positive polarity, and 10% were slightly negative. The data demonstrate that blogs have a higher positive polarity than news articles and that blogs are also more opinionated than news articles.

The study by Rajput et al., (2020) was done on the tweets related to the Coronavirus pandemic using statistical analysis. According to the findings of this study of tweet's word frequency and attitudes, most tweets had a positive polarity, and only few 15% having a negative polarity. Barkur et al., (2020), studied the sentiments of Twitter platform users after the implementation of lockdown in India. In their study, the prominent sentiments were positive. Further, the results suggest that people trust the Government in implementing the lockdown; thus, Indians have positive sentiments of trust. In the study, Indians have trusted their Government in handling the lockdown successfully and will make available the necessary amnesties for their citizens.

A study by Cheeti et al., (2021) on the sentiment analysis of tweets during COVID-19 on education globally suggests that negative sentiment tweets are higher than positive ones. The low sentiments in the findings are due to the inconveniences faced in accessing information through the Internet, quality of content, technological issues, and economic constraints. The effectiveness and quality in distance learning using sentiment analysis of N.L.P. was studied by Nimasha Arambepola, (2020). The researcher extracted the Twitter data to study the opinion of Twitter users about distance education. The tweets were classified according to the polarity into three classes. The tweets' analysis was done using the techniques of sentiment analysis and statistical techniques. The result suggests that adequate Internet facilities and availability of resources enhance the distance learning approach.

Disseminating quality education feedbacks plays a vital role for educators in providing valuable insights about the process of learning and teaching. As observed by Kastrati et al., (2020), providing feedback manually is efficient and effective when fewer students are in online courses. However, this task becomes problematic when many students are present in the online course, blogs, and Massive Online Open Courses (MOOCs). Kastrati et al., (2020) built a framework to analyze students' opinions in the posts automatically. The dataset contains about one lakh reviews from the Coursera student feedback dataset of 5989 students. The experiment's findings suggest that the framework utilized performs better in identifying aspect categories and categorizing sentiment.

Furthermore, the results shows that the framework produces more accurate results than manual sentiment analysis techniques, which are more expensive. Cummins et al., (2010), used sentiment analysis to determine the variations in opinions in programming feedback between students and staff. The researcher studies the differences in perceptions of polarity, i.e., positive, negative, and neutral feedback, among students and examiners.

As observed by Stieglitz & Dang-Xuan, (2012), sentiments affect the re-tweet ability of the messages, i.e., tweets in politics. In their study Stieglitz & Dang-Xuan, (2012), observes that there is a significant effect of words quantity in tweets with affective dimensions, and polarity (such as positive, negative and neutral) emotions related with the political parties, and the re-tweetability (i.e., rate of tweets). The researcher Mubarak Bin Naina Hanif & Saptawati (2014), conducted a correlation analysis on the influence of users and sentiments towards Indonesian presidential elections. According to the findings, the negative correlation scores on presidential candidates are 5% higher than the positive correlation scores, indicating that low feelings have a more substantial impact on increasing influence value and vice versa.

The study by Aiyanyo et al. (2021), examines the emotions of Twitter users using machine learning and sentiment. The findings show that the negative sentiments are related to mental fatigue, loss of employment, responses by the Government, and concerns of students. Positive sentiments are related to the availability of course materials, online community support available, creativity, and access to the classroom. Yang et al., (2020), investigated the emotions related to sensitivity to social risks and the Weibo users' life satisfaction during the COVID-19 pandemic. The findings show that Weibo users had a more negative score of emotions like anxiety and depression concerning the social risks; thus, it suggests that the polarity of positive emotions of Weibo users was reduced during the COVID-19 pandemic. Furthermore, the users of Weibo were more worried about health and family issues and less about the aspects of leisure

and friends. Wang & Zhang (2020), has highlighted the effective use of sentiment analysis opinion mining in reflecting the students learning situations. The method adopted in the study consists of text data pre-processing, identification of topics, opinion analysis, pattern detection, and visualization. The suggested model mentioned in the study accurately detects students' sentiment patterns on a spectrum of issues, providing a valuable practical benchmark for enhancing the effectiveness of information services in the classroom.

Nabeela Altrabsheh, Mohamed Medhat Gaber (2013), noted an application of data mining and sentiment analysis in education to know the learners' opinions. According to Nabeela Altrabsheh, Mohamed Medhat Gaber, (2013), the Naïve Bayes and Support Vector Machine (S.V.M.) techniques are superior in analyzing educational data; further, these two techniques can combine for the analysis of feedback in real-time. A study by Ortigosa, (n.d.), on the analysis of sentiments used the hybrid approach of lexical-based and machine-learning techniques to classify messages. The obtained results shows that this approach has a very accurate (83.27%) analysis of Facebook posts' sentiments. In eLearning, users' sentiments are very useful for analyzing the opinion of the students, and their feedback helps improve the quality of eLearning. The pattern of the information can be used to support personalized learning by analyzing the users' emotions, behavior, and feelings for suggesting the activities on the eLearning platform. Furthermore, the feedback regarding eLearning can be obtained by analyzing the sentiments and opinions of the students.

According to Bravo-Marquez et al., (2013), Sentiments or opinions are lexical artifacts with numerous dimensions. When confronted with external stimuli in data or information, individuals think according to their particular perspectives and orientation. This unique perspective or orientation displays an individual's polarity of opinion. Individuals express their emotions through their opinions, serving as motivation and personal temperament. As a result, one might deduce that the user's emotions and the polarities, either positive or negative, are mutually influenced.

## **DATA SOURCE AND EXTRACTION**

In the Twitter platform, a large amount of data is generated in tweets, and the information from it is extracted with web scraping. According to (Beautiful Soup: Build a Web Scraper With Python – Real Python, n.d.), collecting information from the Internet in the form of data and information is called web scraping. Web scraping is a technique for extracting unstructured material from the web and converting it to structured data. According to Iacus, (2017), in web scraping, the web pages are downloaded and processed with the help of software; web scraping is also helpful in making similar queries in series.

### **Pre-Processing Text**

The process of cleaning and processing the unstructured text data is called pre-processing. As cited in Iacus, (2017), the following are the steps involved in pre-processing of data;

### **Tokenization**

In tokenization, the meaningful terms and phrases are selected, and then sentences and words are separated. The sentences are separated with the help of punctuation like period, question

mark, and exclamation mark; this process is called the Natural Processing Algorithms (Iacus, 2017).

### **Stop Words**

There are categories of words with significantly less semantic meaning in a text document, for example, prepositions, articles, and common nouns. The process of removing the fewer semantic words from the corpus is called the Stop Words (Iacus, 2017).

### **Stemming and lemmatization**

Stemming and lemmatization are text normalization and text mining techniques used in Natural Language Processing to pre-process text, words, and documents (Balodi, 2020). In the pre-processing of textual data, text normalization is crucial. Words in human language take on a variety of shapes depending on the syntactic framework, but their core content is mostly unaltered (Iacus, 2017).

### **CLASSIFICATION OF SENTIMENTS USING V.A.D.E.R. MODULE**

As cited in Mansoor et al., (2021), Sentiment analysis is a concept that needs supervised machine learning. Fine-grained sentiment analysis, aspect-based sentiment analysis, and emotion detection are all types of sentiment analysis. The classification of sentiment is binary; there are positive and negative categories. There are five groups in a fine-grained classification: strong positive, positive, neutral, negative, and strong negative.

Sentiment Analysis is also called opinion analysis, subjectivity analysis, and appraisal mining; in this technique, text analysis is done to find patterns, identify and retrieve information from the textual data in the social media or documents. Sentiment Analysis helps know online users' emotions, behaviors, and opinions by knowing the words' polarity (Nabeela Altrabsheh, Mohamed Medhat Gaber, 2013). As cited in Mansoor et al., (2021), the V.A.D.E.R. module is a rule-based model for analyzing sentiments from tweets, and it was developed by Gilbert. In this module, the polarity-oriented approaches are set up in such a way that they extract high-quality data from documents and messages. The polarity-oriented methods produce a categorical variable with positive, negative, and neutral values.

### **METHOD**

The text mining technique involves four processes;

#### **Data Source and Extraction**

In this study, Twitter is selected for text mining since it is the most popular social media and the messages are short compared to other social media. Text mining is a method of extracting unstructured information from the tweets of the Twitter platform. The Twitter API credentials are accessed through the Twitter developer account. The App is also created for accessing the tweets by registering to an application at <http://apps.twitter.com>. In this research, "5000" tweets were streamed using Web Education, Technology, extracted on 28-11-2021.

#### **Pre-processing Text**



compound polarity scores of 38% of tweets about web education indicate that the tweets were moderately positive and it is more than zero. There was a switch from face-to-face instruction to web-based education during the COVID-19 pandemic. Educators, schools, colleges, and universities focused on online teaching, learning, and evaluation during the pandemic period. The polarity scores are moderately high and not very high; this may be due to the digital divide; in rural areas. There is little access to the Internet, and students of low economic status could not afford to buy smartphones and digital devices for learning Bhagat et al., (2021). From the results, it can be inferred that the sentiments of Twitter users are moderately positive, which is not too high shows that the use of web-based technology in education is new to Twitter users. Moreover, the effective and efficient use of web-based resources in teaching, learning, and evaluation concern the access to quality of education for Twitter users during the COVID-19 pandemic.

### **FUTURE WORK**

The literature review suggests that text mining and sentiment analysis are primarily done in the English language. There is a research gap in text mining and sentiment analysis on Arabic, Persian, and Urdu scripts. The reason may be due to the insufficient availability of Lexicon dictionaries in these languages; thus, there is a need to develop sentiment analysis tools and research the future on text mining and sentiment analysis in the languages such as Arabic, Persian, and Urdu.

### **CONCLUSIONS**

The present study gives valuable insight to the educators in the form of text data from Twitter in making decisions regarding the application of websites, technology in enhancing learning and imparting education to the students. The analysis of the sentiments of tweets shows that the users of the Twitter platform have average positive sentiments scores and not very high sentiments scores about the use of websites and technology in education. The reason may be the sudden switch to web-based education from the traditional face-to-face education and digital divide among the people during the COVID-19 pandemic. As argued by Jung et al., (2020), online teaching approaches were used as emergency remote education during the COVID-19. This activity is distinct from planned and organized approaches such as correspondence education and open distance learning. Furthermore, social injustice, inequality, and the digital gap among disadvantaged learners have impacted the educational quality and access (Jung et al., 2020).



## REFERENCES

Adhikari, N. C. Das, Kurva, V. K., S, S., Kushwaha, J. K., Nayak, A. K., Nayak, S. K., & Shaj, V. (2018). Sentiment Classifier and Analysis for Epidemic Prediction. 31–48. <https://doi.org/10.5121/csit.2018.81004>

Aiyanyo, I. D., Samuel, H., & Lim, H. (2021). Effects of the covid-19 pandemic on classrooms: A case study on foreigners in South Korea using applied machine learning. *Sustainability (Switzerland)*, 13(9). <https://doi.org/10.3390/su13094986>

Balodi, T. (2020). What is Stemming and Lemmatization in N.L.P.? | Analytics Steps. Analytics Steps. <https://www.analyticssteps.com/blogs/what-stemming-and-lemmatization-nlp>

Barkur, G., Vibha, & Kamath, G. B. (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian Journal of Psychiatry*, 51, 102089. <https://doi.org/10.1016/J.AJP.2020.102089>

Beautiful Soup: Build a Web Scraper With Python – Real Python. (n.d.). Retrieved November 30, 2021, from <https://realpython.com/beautiful-soup-web-scraper-python/>

Bhagat, K. K., Mishra, S., Dixit, A., & Chang, C. Y. (2021). Public opinions about online learning during covid-19: A sentiment analysis approach. *Sustainability (Switzerland)*, 13(6), 1–12. <https://doi.org/10.3390/su13063346>

Bravo-Marquez, F., Mendoza, M., & Poblete, B. (2013). Combining strengths, emotions and polarities for boosting Twitter sentiment analysis. *Proceedings of the 2nd International Workshop on Issues of Sentiment Discovery and Opinion Mining, WISDOM 2013 - Held in Conjunction with S.I.G.K.D.D. 2013*. <https://doi.org/10.1145/2502069.2502071>

Cheeti, S. S., Li, Y., & Hadaegh, A. (2021). Twitter based Sentiment Analysis of Impact of Covid-19 on Education Globally. *International Journal of Artificial Intelligence & Applications*, 12(03), 15–24. <https://doi.org/10.5121/ijaia.2021.12302>

Cummins, S., Burd, L., & Hatch, A. (2010). Using Feedback Tags and Sentiment Analysis to Generate Sharable Learning Resources. *10th IEEE International Conference on Advanced Learning Technologies*, 653–657.

Edward, M. (2018). ONLINE SOCIAL MEDIA SENTIMENT ANALYSIS FOR NATIONAL SECURITY (Issue April).

Gupta, N., & Agrawal, R. (2020). Application and techniques of opinion mining. *Hybrid Computational Intelligence*, 1–23. <https://doi.org/10.1016/b978-0-12-818699-2.00001-9>

Iacus, S. M. (2017). Big Data and Social Science - A Practical Guide to Methods and Tools. In Journal of Statistical Software (Vol. 78, Issue Book Review 2). <https://doi.org/10.18637/jss.v078.b02>

Jung, I., Xiao, F., & Radio, S. (2020). A global outlook to the interruption of education due to COVID-19 Pandemic : Navigating in a time of uncertainty and crisis A global outlook to the interruption of education due to COVID-19 Pandemic : Navigating in a time of uncertainty and crisis. June.

Kastrati, Z., Imran, A. S., & Kurti, A. (2020). Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs. June. <https://doi.org/10.1109/ACCESS.2020.3000739>

Kharde, V. A. (2016). Sentiment Analysis of Twitter Data : A Survey of Techniques. 139(11), 5–15.

Mansoor, M., Society, E., & Gurumurthy, K. (2021). Global Sentiment Analysis Of COVID-19 Tweets Over Time \*. October 2020.

Mubarak Bin Naina Hanif, F., & Saptawati, G. A. P. (2014). Correlation analysis of user influence and sentiment on Twitter data. Proceedings of 2014 International Conference on Data and Software Engineering, I.C.O.D.S.E. 2014. <https://doi.org/10.1109/ICODSE.2014.7062491>

Nabeela Altrabsheh, Mohamed Medhat Gaber, M. C. (2013). SA-E: Sentiment Analysis for Education Nabeela. University of Portsmouth PORTS MOUTH Hants UNITED KINGDOM, June, 26–28.

Nimasha Arambepola. (2020). Analysing the Tweets about Distance Learning during COVID-19 Pandemic using Sentiment Analysis Nimasha. International Conference on Advances in Computing and Technology (ICACT–2020) Proceedings, 2756–9160. [https://www.researchgate.net/publication/347442838\\_Analysing\\_the\\_Tweets\\_about\\_Distance\\_Learning\\_during\\_COVID-19\\_Pandemic\\_using\\_Sentiment\\_Analysis](https://www.researchgate.net/publication/347442838_Analysing_the_Tweets_about_Distance_Learning_during_COVID-19_Pandemic_using_Sentiment_Analysis)

Olalekan, A. (2020). Twitter Sentiment Analysis with Orange + Vader + Power BI. Medium, Part 1. <https://medium.com/analytics-vidhya/twitter-sentiment-analysis-with-orange-vader-powerbi-part-1-184b693b9d70>

Orange Data Mining. (n.d.). Retrieved December 12, 2021, from <https://orangedatamining.com/download/#macos>

Ortigosa, A. (n.d.). Sentiment analysis in Facebook and its application to e-learning.

Rajput, N. K., Grover, B. A., & Rathi, V. K. (2020). Word frequency and sentiment analysis of twitter messages during Coronavirus pandemic. <https://arxiv.org/abs/2004.03925v1>

Stieglitz, S., & Dang-Xuan, L. (2012). Political communication and influence through microblogging - An empirical analysis of sentiment in Twitter messages and retweet behavior. Proceedings of the Annual Hawaii International Conference on System Sciences, 3500–3509. <https://doi.org/10.1109/HICSS.2012.476>

Wang, K., & Zhang, Y. (2020). Topic Sentiment Analysis in Online Learning Community from College Students. Journal of Data and Information Science, 5(2), 33–61. <https://doi.org/10.2478/JDIS-2020-0009>

What Is Machine Learning? | How It Works, Techniques & Applications - MATLAB & Simulink. (n.d.). Retrieved December 6, 2021, from <https://www.mathworks.com/discovery/machine-learning.html>

Yang, L. I., Li, Y., Wang, J. I. N., & Sherratt, R. S. (2020). Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. 8.