Library Progress International *Print version ISSN 0970 1052*
Vol.44 No.3, Jul-Dec 2024: P.7388-7397 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

Aspect Based Sentiment Analysis of Distance Learning Using Natural Language Processing Techniques

Jitendra Soni

Institute of Engineering & Technology, Devi Ahilya Vishwavidyalaya Indore, India
jsoni@ietdavv.edu.in

How to cite this article: Jitendra Soni (2024) Aspect Based Sentiment Analysis of Distance Learning Using Natural Language Processing Techniques. *Library Progress International,* 44(3), 7388-7397.

ABSTRACT

Understanding public opinion on key issues is essential given the exponential growth of online discourse. This study investigates aspect-based sentiment analysis, concentrating on tweets pertaining to online learning. Measuring attitudes is becoming more crucial as the global education landscape shifts rapidly in order to provide more flexible and effective online learning environments.

The research is rigorous, beginning with a comprehensive statistical analysis of a substantial set of 202,645 entries from Twitter. Through exploratory data analysis, trends in the tweet distribution across different geolocations are uncovered, highlighting temporal patterns and global interaction dynamics. Two feature engineering techniques—type pre proxy and vectorization—allow for a thorough examination of textual material.

The main component of the project is aspect-based sentiment analysis, which determines sentiments for a number of factors like instructor participation, platform usability, and course content. Using the TextBlob library, this study provides a deep understanding of how users express their viewpoints on many aspects of remote learning by revealing attitudes at the category and aspect levels.

Visualizations like word clouds, geographic plots, and hourly sentiment distributions provide rich insights into the dialog. The project's innovative technique includes sentiment scoring for certain aspect categories, such as positive and negative sentiment scores for "course content" and "platform usability." These scores, which are based on the frequency and sentiment polarity of phrases, offer a dynamic picture of the variables that most influence moods.

Dialogue can be deeply understood by visualizations such as word clouds, geographic plots, and hourly sentiment distributions. Sentiment scoring for certain aspect categories—such as positive and negative sentiment scores for "course content" and "platform usability"—is one of the project's novel techniques. Based on the frequency and sentiment polarity of phrases, these scores provide a dynamic picture of the factors that most affect moods.

In summary, our study advances the rapidly evolving field of sentiment analysis and provides a comprehensive framework for examining attitudes in the intricate realm of online conversation, especially with regard to remote learning.

Keywords: online education, social media, sentiment analysis based on aspects, sentiment scoring on Twitter, and educational technology.

Introduction

The introduction of online learning has caused a paradigm shift in the rapidly changing field of education, changing not only how knowledge is transferred but also how students interact with course materials. Numerous viewpoints and opinions have been generated by this significant change, and they have been passionately expressed on social media sites. Twitter is particularly lively for these kinds of discussions because it is a real-time microblogging platform. This work analyzes and assesses the emotions contained in tweets on distant learning using Aspect-Based Sentiment Analysis (ABSA), a sophisticated lens to show the subtle parts of these expressions.

By adding the idea of aspects or traits, Aspect-Based Sentiment Analysis is a complex methodology that goes beyond conventional sentiment analysis. This method connects emotions to particular literary elements, enabling a more nuanced comprehension of sensations. ABSA is an essential tool for analyzing the subtle differences between opinions stated on Twitter in the context of remote learning. It provides a comprehensive analysis that breaks down emotions for a range of topics, including the usefulness of online platforms and the emotional effects of remote learning.

The paper explores the unique characteristics of ABSA, highlighting its capacity to offer improved entity-level comprehension, skillfully manage language nuances, and efficiently handle sentiment ambiguity. Furthermore, ABSA's real-time monitoring capabilities complements the organic flow of social media debates effectively.

Social media is a place where additional viewpoints regarding online learning are shared, which has benefits and drawbacks for study. It's difficult to convey the subtleties of emotions in tweets because of their informal, concise style and wide range of languages and idioms. But these difficulties also present opportunities for cutting-edge and novel sentiment analysis techniques. By tackling these problems, the research hopes to clarify emotions and create approaches that can comprehend how the discourse on remote learning has evolved in the digital era.

1 Data Analysis

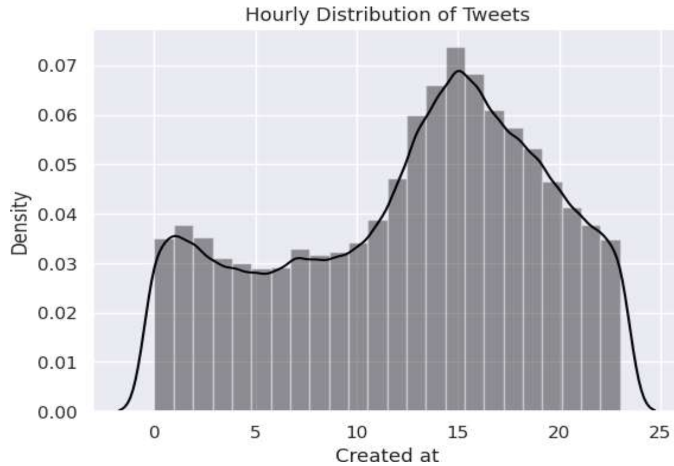
To extract valuable insights from an exploration of Twitter sentiments regarding remote learning, a thorough data analysis is necessary. In order to uncover significant findings, trends, and patterns, the data was carefully examined. The total number of entries was 202,645; columns were included for content, location, username, favorite content, number of retweets, and generated at.

Importing the necessary libraries—which included re, matplotlib, pandas, numpy, and seaborn—was among the first things to perform. The dataset had dimensions of (202,645, 8) when it was loaded, and its columns contained data that was essential to our sentiment analysis. The second investigation's main focus was on the content column, which collected textual comments from Twitter users regarding distant learning.

	Unnamed: 0.1	Unnamed: 0	Content	Location	Username	Retweet-Count	Favorites	Created at
0	0	0	innovate an innovative approach #quoteoftheday...	NaN	PaulBillygraha1	0	0	2020-08-02 04:56:27
1	1	1	The pandemic is raising concerns about how tee...	Worldwide	IAM__Network	0	0	2020-08-02 04:49:43
2	2	2	STI: Staying Education-ready in the New Normal...	Worldwide	IAM__Network	0	0	2020-08-02 04:32:36
3	3	3	Digital Learning Through Digital RCRT\n.\n.\nR...	NaN	digitalrcrt	0	0	2020-08-02 04:30:12
4	4	4	Upswing Classroom: Out and Out Virtual School,...	India	etr_in	1	0	2020-08-02 04:00:21

Histograms and geographic plots are two instances of the visual aids that were used to provide the dataset a more understandable visual representation. These visual tools showed spots with considerable engagement and shed insights on how tweets were distributed across different geographies. The community's perception of the attractiveness and resonance of particular tweets was reflected by the frequent distribution of likes and retweets. Moreover, the examination of the temporal component of tweet generation clarifies the differences in discourse surrounding distance learning throughout time.

The ability to see patterns in the dataset was mostly made possible by exploratory data analysis, or EDA. We now have a better grasp of the mechanisms underlying the emotions associated with remote learning on Twitter thanks to EDA, which can be used to identify activity spikes and trends in user participation.



A map of the geographic distribution of tweets was made to aid with this investigation. This graph not only gives our investigation a dynamic new dimension, but it also provides a persuasive synopsis of the international conversation on remote learning. Prominent nations are discussed extensively, offering a concrete illustration of how people around the globe are reacting to the topic on Twitter [7].

2 Featured Engineering

In order to derive significant representations from textual input and conduct advanced sentiment analysis on tweets linked to remote learning, feature engineering becomes a crucial stage.

Pre-processing the textual data to improve feature quality was the first step. The goal was to reduce the material to its most basic elements by eliminating stopwords, or common words that frequently do not contribute much meaning to a sentence. Furthermore, as hashtags and punctuation are commonly employed in Twitter discourse, they were eliminated to guarantee that the sentiment analysis concentrated on the important language components of the tweets. This phase improved the dataset and laid the framework for a sentiment analysis that would be more precise [10].

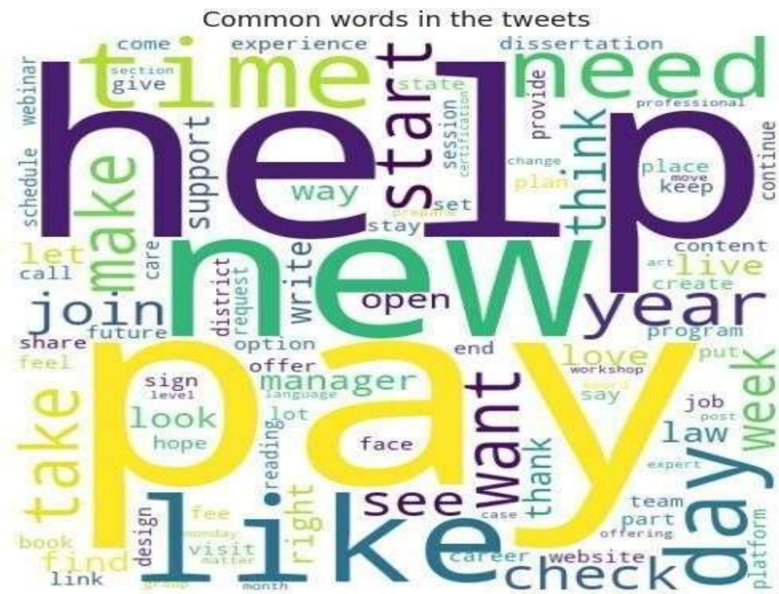
One important component of natural language processing (NLP) is tokenization, which is the process of dividing a text document into discrete pieces, or tokens. To obtain a structured representation of the terms included in each tweet, the Natural Language Toolkit (NLTK) package was used to tokenize the tweets. NLTK offers a wide range of tools for NLP-related tasks, making it a good option for decomposing textual content into its component pieces. Tokenization is a crucial phase that makes it easier to do further research and permits a more thorough examination of the emotions connected to certain words or phrases [10].

The pre-processing and tokenization processes were represented in the dataset by a new column called "processed," which was inserted with the appropriate label. Following the complete removal and tokenization of stopwords, hashtags, and punctuation, the refined text content was stored in this column. The original tweets that have been standardized and edited are included in the "processed" column. This clever feature facilitates a more targeted investigation of attitudes regarding particular facets of distant learning and makes it easier to proceed to the later phases of sentiment analysis [10].

3 Vectorization and Word Cloud

Vectorization and word cloud creation are crucial steps in turning textual data into useful insights for aspect-based sentiment analysis on tweets related to remote learning.

An analytically-ready numerical structure is created by vectorizing the processed textual data using the Text Frequency-Inverse Document Frequency (TF-IDF) approach. Weights are assigned to words by TfidfVectorizer based on their overall dataset infrequency as well as their frequency of occurrence in a given document [3]. Words that are prevalent inside a tweet and unique across the corpus can be prioritized using this strategy. TfidfVectorizer is used to accurately determine the relevance of words in the final feature matrix, both inside individual tweets and throughout the entire dataset. This conversion lays the groundwork for further research, in which the measurement of words is crucial o identify trends and points of view. [1].



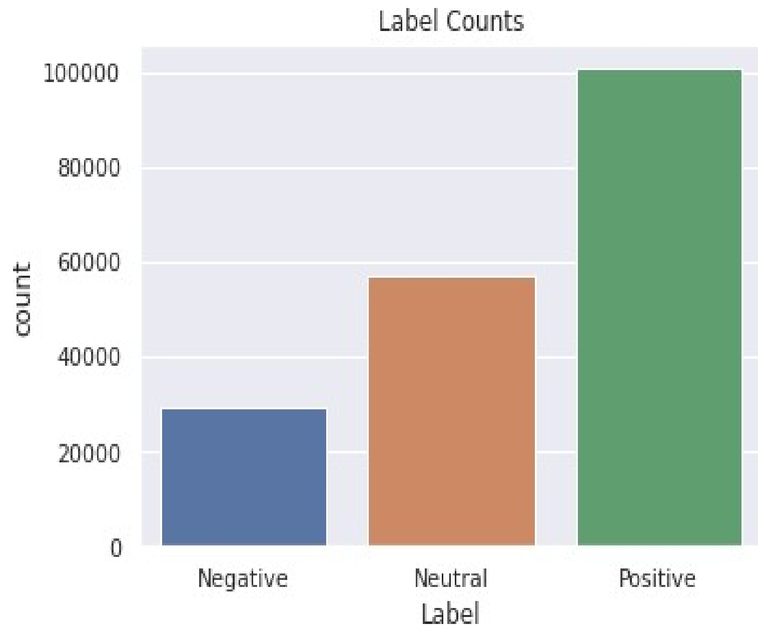
The WordCloud package may be used to represent the dataset as a word cloud, which makes it visually appealing to see the most frequently occurring terms. This technique displays processed text data as a cloud-like structure, where the size of each word represents its frequency [9]. The key ideas and points of view that are now influencing discussions on remote learning are succinctly and clearly summarized in this graphic. By giving a visual depiction of the phrases that are commonly used, the Word Cloud facilitates the identification of popular topics and trends and makes it simpler to recognize significant themes.

4 Aspect Based Sentiment Analysis

An essential stage in our analysis of tweets concerning remote learning is aspect-based sentiment analysis, which makes use of the TextBlob package to extract detailed information about the emotional tones expressed in the text.

TextBlob is a robust Python package that makes sentiment analysis and natural language processing tasks easier. TextBlob provides a quick and easy way to score text based on subjectivity and polarity using pre-trained models to measure sentiment. The degree to which a sentence conveys an opinion rather than being an objective statement is known as its subjectivity. From 0 (objective) to 1 (subjective), it goes. On the other hand, polarity, which ranges from -1 (negative) to 1 (positive), indicates how positive or negative a sentence is [10].

TextBlob was used to determine the subjectivity and polarity of every tweet in the collection. To ascertain the general emotion of a tweet, the library examines each word independently, taking into account its semantic orientation and contextual usage [2]. The degree of positivity or negativity of a sentiment is reflected in its polarity, which is a numerical value. Subjectivity also reveals how much of the remark is grounded on reality versus opinion. The aggregated polarity ratings of all tweets make it easier to assess the sample's overall sentiment distribution. Three sentiment categories—positive, neutral, and negative—account for the results. Through the establishment of a polarity value threshold, tweets are categorized in a manner that provides a thorough summary of the dominant



perspectives presented in the conversation around remote learning [4].

5 Words Cloud And Visualisation

Word clouds provide an eye-catching visual depiction of the terms that are most frequently used in tweets—both positive and negative. Word clouds were created independently for tweets categorized as positive and negative based on their polarity scores using the WordCloud library. The word cloud, where each word's magnitude represents its frequency of occurrence, gives users a quick visual understanding of the themes linked to both positive and negative emotions. By examining these word clouds, significant phrases that characterize both positive and negative expressions can be identified. This provides insight into the main factors influencing attitudes toward the subject of remote learning [8].

The distribution of emotions by country is shown via geographic visualization. The frequency of positive, neutral, and negative attitudes in each nation is graphically displayed using bar charts or maps. This research highlights geographic trends and variations in the emotional tones associated with Twitter disputes over distant learning, which contributes to a more nuanced understanding of how emotions change across different locales [10].

TextBlob is a robust Python package that makes sentiment analysis and natural language processing tasks easier. TextBlob provides a quick and easy way to score text based on subjectivity and polarity using pre-trained models to measure sentiment. The degree to which a sentence conveys an opinion rather than being an objective statement is known as its subjectivity. From 0 (objective) to 1 (subjective), it goes. On the other hand, polarity, which ranges from -1 (negative) to 1 (positive), indicates how positive or negative a sentence is [10].

Aspect Category	Extended Aspects	Score Positive	Score Negative	% Positive	% Negative
Content	Content	250.456	70.054	78.14	21.85
	Image	62.897	42.432	59.71	40.28
	Quiz	65.123	54.876	54.27	45.73
	Discussion	40.721	80.51	33.58	66.41
Platform	Platform	150.65	20.34	88.10	11.89
	User friendly	30.87	50.331	38.01	61.98
	Stability	40.816	60.126	40.43	59.57
	Reliability	46.50	70.589	39.71	60.28
Connectivity	Connectivity	101.953	57.23	64.04	35.95
	Network	98.52	52.873	53.27	46.72
	Wifi	120.534	61.938	66.08	33.92
Faculty	Faculty	100.34	52.892	65.48	34.51
	Availability	80.653	23.847	70.11	29.88
	Responsiveness	73.273	87.467	45.58	54.41
	Teaching Methods	65.992	109.673	37.56	62.43
Cost	Cost	170.345	93.89	64.46	35.53
	Free	240.981	72.176	76.95	23.04
	Fees	90.178	34.767	72.17	27.82
	Money	82.45	63.979	56.307	43.693

6 Sentiments identified using context and aspects

The suggested method uses an attention encoder to extract useful information from text. More precisely, the encoder attention to the primary element under examination determines each word's relevance score. When encoding the word representations, it considers the connections between aspect and context words. These attention-based encodings are concatenated to generate an aspect retrieval vector once the attention score of each word has been ascertained.. The important details in the text are captured by this vector. Furthermore, Paragraph2vec extracts context clues from word sequences by learning their distributed vector representations. After that, an average is used to merge the aspect and context vectors. An LSTM network is fed this final unified representation, which integrates aspect and contextual semantics, in order to classify sentiment. The contextual meaning that goes along with it is provided by paragraph2vec, and the attention encoder allows the model to concentrate on the informative portions. The LSTM classifier's capacity for sentiment analysis has been enhanced by their combination.
Aspect and Context Generated using the fusion approach

Aspect	Context	Label	Score
Teaching Quality	engaging and knowledgeable instructor	POSITIVE	0.99
	clear explanations	POSITIVE	0.98
	boring lectures	NEGATIVE	0.95

	rd to follow	EGATIVE	94
Course Materials	well-structured course	POSITIVE	0.99
	useful materials	POSITIVE	0.97
	outdated materials	NEGATIVE	0.95
	lack of resources	NEGATIVE	0.92
Online Platform	user-friendly interface	POSITIVE	0.98
	easy to navigate	POSITIVE	0.97
	glitchy platform	NEGATIVE	0.94
	connectivity issues	NEGATIVE	0.93
Flexibility	self-paced learning	POSITIVE	0.96
	accommodating deadlines	POSITIVE	0.95
	inflexible schedule	NEGATIVE	0.91
	no accommodations	NEGATIVE	0.90
Social Interaction	lively discussions	POSITIVE	0.97
	collaborative environment	POSITIVE	0.96
	isolating experience	NEGATIVE	0.94
	lack of interaction	NEGATIVE	0.92
Workload	manageable workload	POSITIVE	0.98
	achievable expectations	POSITIVE	0.96
	overwhelming workload	NEGATIVE	0.93
	unrealistic expectations	NEGATIVE	0.91
Support	helpful support	POSITIVE	0.99
	responsive to questions	POSITIVE	0.98
	unhelpful support	NEGATIVE	0.94
	lack of guidance	NEGATIVE	0.92

8 Conclusion

This thorough examination of Twitter's attitude toward distance learning has produced insightful information on every aspect, enabling a more nuanced understanding of the variety of viewpoints and feelings around this global education revolution.

202,645 tweets made up the initial dataset, which we examined in great detail. Fundamental insights were obtained by removing superfluous columns, looking at the tweet distribution across geolocations, and visualizing like and retweet rates. Sentiment analysis was made possible by the spatial representation of tweets and their temporal patterns.

In order to improve the textual data, feature engineering was essential. To make the language simpler, stopwords, hashtags, and punctuation were eliminated. Tokenization with the NLTK library produced a structured representation. Cross-analysis transitions were made easier and the sentiment analysis feature quality was improved with the addition of the "processed" column.

The TextBlob tool made sentiment analysis easier by determining the subjectivity and polarity of each tweet. A mixed bag of sentiment was seen in the overall sentiment distribution: Positive (100,794), Neutral (57,060), and Negative (29,198). A comprehensive image of the emotional tones is provided via visualizations that display the counts and scores of both positive and negative tweets.

Word clouds that highlight important terms in both positive and negative tweets can be used to provide a visual representation of the dominant themes. The sentiment distribution unique to each nation was visualized geographically, and the hours of highest sentiment expression were identified by analyzing hourly data on both positive and negative tweets.

The main conclusions show a wide range of opinions, with positive statements predominating in the discussion. Sentiment analysis based on features reveals complex feelings that correspond to various viewpoints about various aspects of remote learning. Geographic variations and temporal dynamics demonstrate how pervasive and dynamic Twitter attitudes are.
In conclusion, this study provides a methodological framework for further research as well as insight into the sentiment landscape around remote learning on Twitter. Sentiment analysis, feature engineering, vectorization, and visualizations are combined to create a comprehensive method for comprehending complicated emotions in the digital age of education.

References

[1] G. A. Ruz, P. A. Henríquez, and A. Mascareño, ''Sentiment analysisof Twitter data during critical events through Bayesian networks classifiers,'' Future Gener. Comput. Syst., vol. 106, pp. 92–104, May 2020.

[2] M. Atzeni, A. Dridi, and D. Reforgiato Recupero, ''Using frame-based resources for sentiment analysis within the financialdomain,'' Prog. Artif. Intell., vol. 7, no. 4, pp. 273–294, Dec. 2018.

[3] A. Abdi, S. M. Shamsuddin, S. Hasan, and J. Piran, ''Deep learning- based sentiment classification of evaluative text based onmulti-feature fusion,'' Inf. Process. Manage., vol. 56, no. 4, pp. 1245–1259, Jul. 2019.

[4] P. Kaewyong, A. Sukprasert, N. Salim, and F. A. Phang, ''The possibility of students' comments automatic interpret using lexicon basedsentiment analysis to teacher evaluation,'' in Proc. Int. Conf. Artif. Intell. Comput. Sci. (AICS), 2015, pp. 179–189.

[5] D. Dessí, M. Dragoni, G. Fenu, M. Marras, and D. R. Recupero, ''Deep learning adaptation with word embeddings for sentiment analysis on online course reviews,'' in Deep Learning-Based Approachesfor Sentiment Analysis. Singapore: Springer, 2020, pp. 57–83.

[6] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, ''What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach,'' Comput. Edu., vol. 145, Feb.2020, Art. no. 103724.

[7] Y. Wang, M. Huang, X. Zhu, and L. Zhao, ''Attention-based LSTMfor aspect-level sentiment classification,'' in Proc. Conf. Empirical Methods Natural Lang. Process., 2016, pp. 606–615. [8] B. Huang and K. Carley, ''Parameterized convolutional neural networks for aspect level sentiment classification,'' in Proc. Conf. Empirical Methods Natural Lang. Process., 2018, pp. 1091–1096.

[9] Z. Li, Y. Wei, Y. Zhang, X. Zhang, and X. Li, ‘‘Exploiting coarse-to-fine task transfer for aspect-level sentiment classification,’’ inProc. AAAI Conf. Artif. Intell., vol. 33, 2019, pp. 4253–4260.

[10] Zenun Kastrati , Ali Shariq Imran , Arianit Kurti “Weakly Supervised Frame- work for Aspect-Based Sentiment Analysis on Students’