2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT)
CHRIST (Deemed to be University), Pune Lavasa Campus, India. Oct 14-15, 2022

2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT) | 978-1-6654-5361-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/TQCEBT54229.2022.10041491

Perception and Practices of EdTech Platform: A Sentiment Analysis

Dr Abhijeet Birari
Assistant Professor
Christ University, India
abhijeet.birari@christuniversity.in

Rahul Janardanan
Student, MBA
Christ University, India
rahul.janardanan@mba.christuniversity.in

Dr. Rohit Sharma
Assistant Professor
Christ University, India
rohit.sharma@christuniversity.in

Dr. Prajakta Yawalkar
Associate Professor
Christ University, India
prajakta.yawalkar@christuniversity.in

Dr. Lakshmi Subramani
Academic Manager
S P Jain School of Global Management
lakshmi.subramani@spjain.org

Abstract: **Virtual and digital learning being the new normal, pandemic outburst and unexpected disruption in the functioning of educational services have paved way for online learning services. Considering the fast-track growth of the education technology (EdTech) industry, in order to sustain, it is imperative for the industry to understand the underlying issues by capturing the end users' perception. The primary purpose of this research is to examine the perception of users towards EdTech platforms A sample of 600 reviews regarding three major EdTech platforms were scraped from MouthShut.com as textual data and analysed using lexicon-based method. The polarity of the sentiments pertaining to the reviews of different platforms was analysed using sentiment analysis. Furthermore, the topic modelling on the reviews was performed using natural language programming. The results revealed a positive sentiment of users towards the EdTech services and platforms. The most influential factors are faculty expertise, interface user-friendliness, syllabus, and pricing model. Our findings help EdTech service providers to understand which factors are driving this dramatic shift in student behaviour so they may develop better strategies to attract and retain consumers. Despite the rise in EdTech platform popularity, this is the first study to investigate perception of EdTech users comprehensively.**

Keywords: EdTech platform, Perception, Online Learning Services, Sentiment Analysis

I. INTRODUCTION

The unpredicted outbreak of the Covid 19 pandemic resulted in discrepancies in the normal functioning of the educational sector. The traditional model of education was put to a halt by 30 March 2020. Online mode of learning took shape as the predominant mode of education through different online platforms. This gave a boost to EdTech Industry in India and many other companies which have been into educational technology which benefited from this new trend. Pioneers of Indian EdTech like Byju's and other platforms increased their sales in folds. The Pandemic opened doors for many EdTech service platforms to provide their services in a much innovative and upgraded manner. Working professionals, students, and educational enthusiasts used the lockdown period to either upskill or utilize the online mode to obtain certifications and deeper learning. This growing importance of the online educational approach enables exponential growth in academic service providers' online courses and develops their marketing strategies.

The transformative power of digital learning has vital effects on consumer behaviors, and it is essential to understand the consumer perceptions and sentiments towards the service and its attributes. The study aims to understand the consumer's attitude towards different platforms of the EdTech industry prominent in the Indian market. The emergence of new fields like text mining and Natural language process (NLP) and machine learning techniques, it has become possible to mine the opinion of people on various issues.

For the study, majorly three service platforms are taken for comparison and analysis to understand consumers' sentiments across the three and to understand the consumer perception. Three platforms that cater to three different target segments have been taken to understand better and factor variations.

II. RELATED WORK

An extensive survey related to the methods and techniques of Sentiment analysis and Topic modelling was carried out to get a clear understanding. The following table represents the techniques, dataset, and findings.

III. PHILOSOPHICAL UNDERPINNINGS

During the lockdown situation, many people started upskilling through online certification on technical and programming courses. The requirements for all the audiences were similar among different academic levels. We took three different platforms which cater to varying levels of academics. Byju's delivered to high school and preuniversity segments, whereas Udacity and Simplilearn were upskilling courses for working professionals and students.

To analyse and achieve the study's objectives, we need to identify the factors involved in the service industry that companies consider to shape their competitive strategies and customer satisfaction.

TABLE I. RELATED WORKS

Author	Dataset	Technique Applied	Key Findings
Albalawi et al. (2020) [1]	Publicly available datasets were analyzed in this study	Latent semantic analysis, latent Dirichlet allocation, non- negative matrix factorization, random projection, and principal component analysis	Latent Dirichlet allocation and non- negative matrix factorization methods delivered more meaningful extracted topics and obtained good results.
Koehler et al. (2015) [2]	Tweets	Sentiment analysis	SA remains untapped resource despite its scope and potential. Can be utilized for larger dataset from social media platforms.
Tong and Zhang (2016) [3]	Textual Twitter data	LDA, Text Mining	Less computation power restricted smaller sample selection. Can be used on data of Twitter, Wikipedia, social media.
Rao et al. (2020) [4]	Reviews, Social media posts	Topic mining using "Voyant tools"	Sentiment analysis is a supervised learning method for identifying positive, negative, and neutral opinions in text.
Bhagat et al. (2021) [5]	Newspaper articles, Blogs	Sentiment Analysis, Webscrapping	Results proved that more than 85% articles are positive and remaining are negative. Blogs proved to be more positive than newspaper articles
Asmussen and Moller (2019) [6]	Textual data from a large collection of papers for an exploratory literature review	Latent Dirichlet Allocation	Framework relevant for junior researchers due to usage of limited resources and faster reviews
Tashmin Khamis (2019) [7]	User Responses	Factor Analysis	The findings emphasize the importance of building trust and providing opportunities for individuals with the necessary skills to lead in a participatory and distributive manner while working with limited human and financial resources.
Melissa Laufer (2021) [8]	Questionnaire responses	SPSS	The research revealed that people are optimistic about digital education and the benefits it provides.
Yamini Chandra (2020) [9]	Survey Data	Chi-square analysis	Significant discrepancies were found between male and female students' fears of academic failure and their online and home environments.
Nidhi and Arti (2021) [10]	Twitter data. Interview responses	Sentiment analysis	People are neutral when sharing their feelings on social media.
Nimasha Arambepola (2020) [11]	Twitter Data set	Sentiment analysis	54% positive attitude from customers towards the distance learning, 16% shows neutral sentiments and 30% shows negative sentiments.
Biswarup Nandi (2017) [12]	Textual User Responses	Control Flow Graph (CFG)	It is very important to consider users sentiments and through sentiment analysis it is possible grasp the sentiments effectively
Frankvan Cappelle (2021) [13]	Computer Assisted Telephonic Interviews (CATI)	Multiple linear regression analysis	There is no one-size- fits-all solution for online learning during school closures.

- Pricing and subscription
- Faculty Expertise
- Syllabus Quality
- Content delivery

Therefore, it is imperative to understand customers' sentiments and understand whether they have positive or negative sentiments towards each service platform and its attributes.

To understand this, we shape up few questions to answer.

- How to understand the customer's emotions or sentiments about a specific platform through reviews and social media posts?
- How do we identify the topics in which they have a concern?
- Can we identify the main factors influencing customer behavior in selecting an online educational service?

After referring to the literature related to similar topics, the following techniques have the highest potential to solve the problems and bring up an analytical solution.

- Sentiment Analysis
- Topic modeling

IV. RESEARCH METHODOLOGY

Stage 1 - Data Collection

Customers reviews about different EdTech platforms in the Indian market was collected. The three prominent companies selected for the study are the following

- Byjus
- Udacity

- Simplilearn

For the study, the crucial data obtainable are the customer posts and reviews about the EdTech platforms, which gives us a clear idea of the customer's attitudes and sentiments towards the three platforms. The reviews will provide us with specific topics customers have concerns about. This approach will help us understand the attributes of the EdTech services that customers seek and value. Majorly, customer reviews from online websites like Mouthshut.com and official review portals of the service platforms were referred to acquire the data. Web Scrapping method was utilized to acquire the data from the websites. Twenty reviews of each three platforms were taken for the analysis, of which ten were of positive polarity, and the rest were of opposing polarity.

This study is aimed towards carrying out a comparative sentimental analysis towards user preference towards the electric vehicles namely Tata Nexon, Hyundai Kona, and Mahindra E20 in Indian context.

Reviews of the customers are been taken for the analysis of the data and to do a comparative analysis of electric vehicles. Such reviews are required to understand the sentiment, that the customers have towards electric vehicles. For the data collection, the web scrapping method is used to get reviews from various websites. Car websites, YouTube videos, expert views, social media, etc. are the websites from where the web scrapping method was being performed.

The web scrapping method is a method that is used to collect data for any analysis and to process large data sets from particular URLs/websites Web scraping is the process of obtaining structured web data in an automated manner.

It's also known as web data extraction. Web scraping is used for variety of reasons, including pricing monitoring, price intelligence, news monitoring, lead generation, and market research. Web data extraction is utilized by people and enterprises who wish to make better decisions by utilizing the huge amount of publicly available web data.

Stage 2- Data Preprocessing Corpus

The data gathered in stage 1 would be uploaded into the orange program for additional analytical processing. The steps taken when analyzing the data are as follows:

Transformation:

The given review data would be converted to lowercase and then processed further by eliminating any tags and URLs if any were discovered. It also eliminates accents if any are detected.

Tokenization:

We would separate by regular expressions and use or maintain just words if we used Tokenization. The text document is divided into words, sentences, and bigrams at this stage. Tokenization is a crucial step in the process of separating unstructured data into distinct components. The review data-sheet is treated to eliminate white spaces in this research study. The emoticons and other special symbols are preserved as a part of every document in the review corpus.

Text Normalization:

To analyze the data set and extract some contextual meaning/text from the page, the Wordnet Lemmatization approach would be employed. WordNet categorizes semantic relationships between words, allowing it to analyze and extract relevant contextual words from a text source.

Filtering:

We shall eliminate stop words from the English language using this method. This phase examines the word selection and eliminates any undesirable words, symbols, or other elements. To delete (and, or...in) from a text document, use the stop word. In this study, mathematical symbols, hashtags, and special characters were removed from the reviews data. The filtered list would then be improved with the TD- IDF technique.

Keyword Extraction:

Stage 3 begins after the text pre-processing phase in stage 2 has been finished. The number of times a given word token appears in the data set list is calculated using the Term Frequency technique, and its relative significance ratio, when compared to other terms, is calculated using the TD-IDF score.

This approach assigned a score to each word token based on its relative significance. The most commonly occurring words are displayed as a word cloud.

Stage 3- Topic Modeling

To find sub-themes in a text corpus, the Topic Modelling technique is utilized. The purchase of electric vehicles is influenced by a number of factors, including price, exterior design, residual value, features and technology, safety.

Pre-processed tokens will be turned into a dictionary with a word index and a count of the number of times they appear in the corpus. Topic modeling is essentially an experimental technique that is complemented by subject–area experts with extra knowledge. The ten subthemes created by the LDA algorithm will be manually labeled by the researchers in the current study.

For the purpose of future study, topic modeling with the Latent Dirichlet Allocation Process will be used. Let's go over the inputs and outputs now:

Input is the corpus that has the text document of the review collected.

The output would contain differentiation of the document with specific topic name extracted from the document and then weights are assigned to each topic as well as to words as per its importance or repetition in the text document and also as per the importance ratio when compared to other topics.

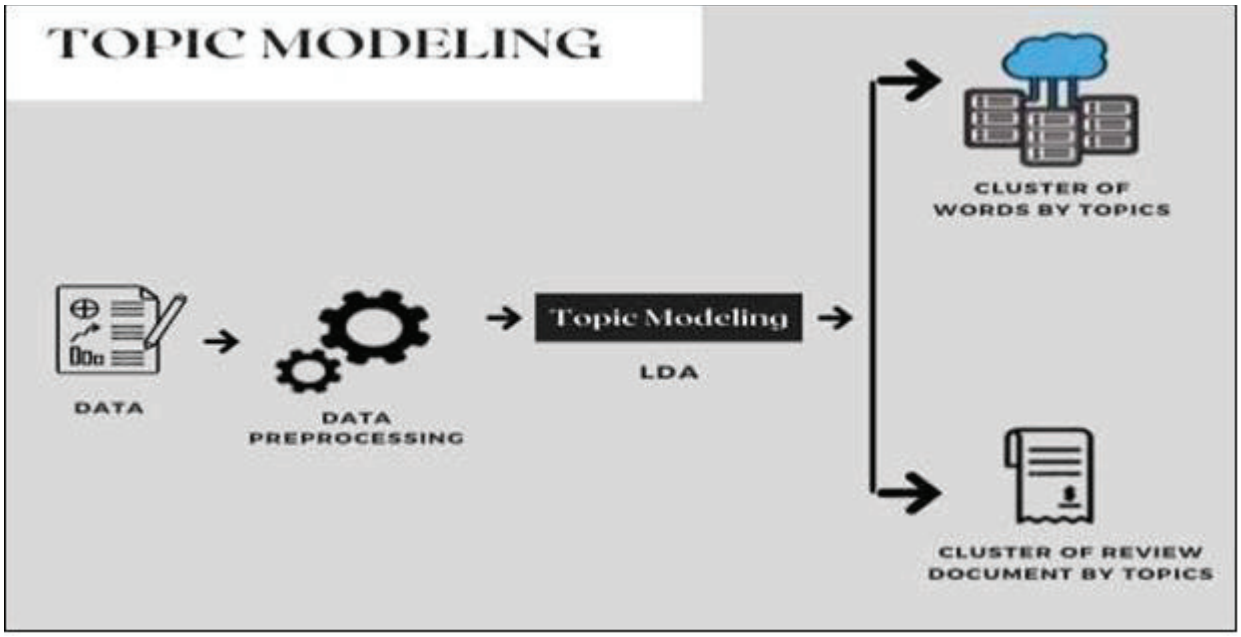

Fig 1: Topic Modeling flowchart

Topic modeling is a method of investigation. This technique is commonly used to deduce the underlying semantics of a corpus of text. When assessing the topic modeling technique's output, less attention is placed on accurately estimating population parameters; rather, the major purpose is to discover the best lens through which to see the data. The LDA technique is used to construct an abstract subject list of ten themes while capturing the sentiments towards electric vehicles.

Stage 4- Sentimental Analysis

The method of assessing whether a piece of writing is good, negative, or neutral is known as sentiment analysis. It's also known as opinion mining, and it involves determining a speaker's opinion or attitude.

Companies utilize it in the marketing area to establish strategies, understand customers' attitudes toward products or brands, howpeople respond to campaigns or new launches, and why consumers don't buy certain things.

Sentiment Analysis predicts sentiment for each document in a corpus. Sentiment analysis through Liu-Hu is used. The difference between the total of positive and negative words, normalized by the document's length and multiplied by a hundred, is the final score. The final score shows the document's emotion difference as a percentage.

After the preprocessing stage and the topic modeling stage, where the sub-topics are classified, the data review corpus reaches the final stage, where sentiment analysis is performed on the processed data, and then the reviews are assigned composite scores using the Liu-Hu model and thus are classified as positive or negative reviews.

A score of -1 is described as the most extreme negative, while a score of +1 is described as the most extreme positive.

Stage 5 - Thematic Analysis

Thematic analysis is a method for deciphering the meaning of data. The identified themes and keywords are investigated in this study in order to determine the meaning of the discovered themes and keywords. Thematic analysis is vital for categorising and dividing data into themes, as well as analysing and interpreting the primary subjects and themes from the data obtained.

V. EMPIRICAL RESULT AND ANALYSIS

A. Topic Modelling

LDA topic modeling has been used to identify the themes discussed in the reviews by the customers and the related keyword. The model obtained the major topics with high correlation to the study, and the most relevant keywords are manually obtained through extensive literature review. Here in the table (2), we can see 10 prominent themes which define the vital attributes of the EdTech service platforms and related keywords with higher weights are also labeled in the table.

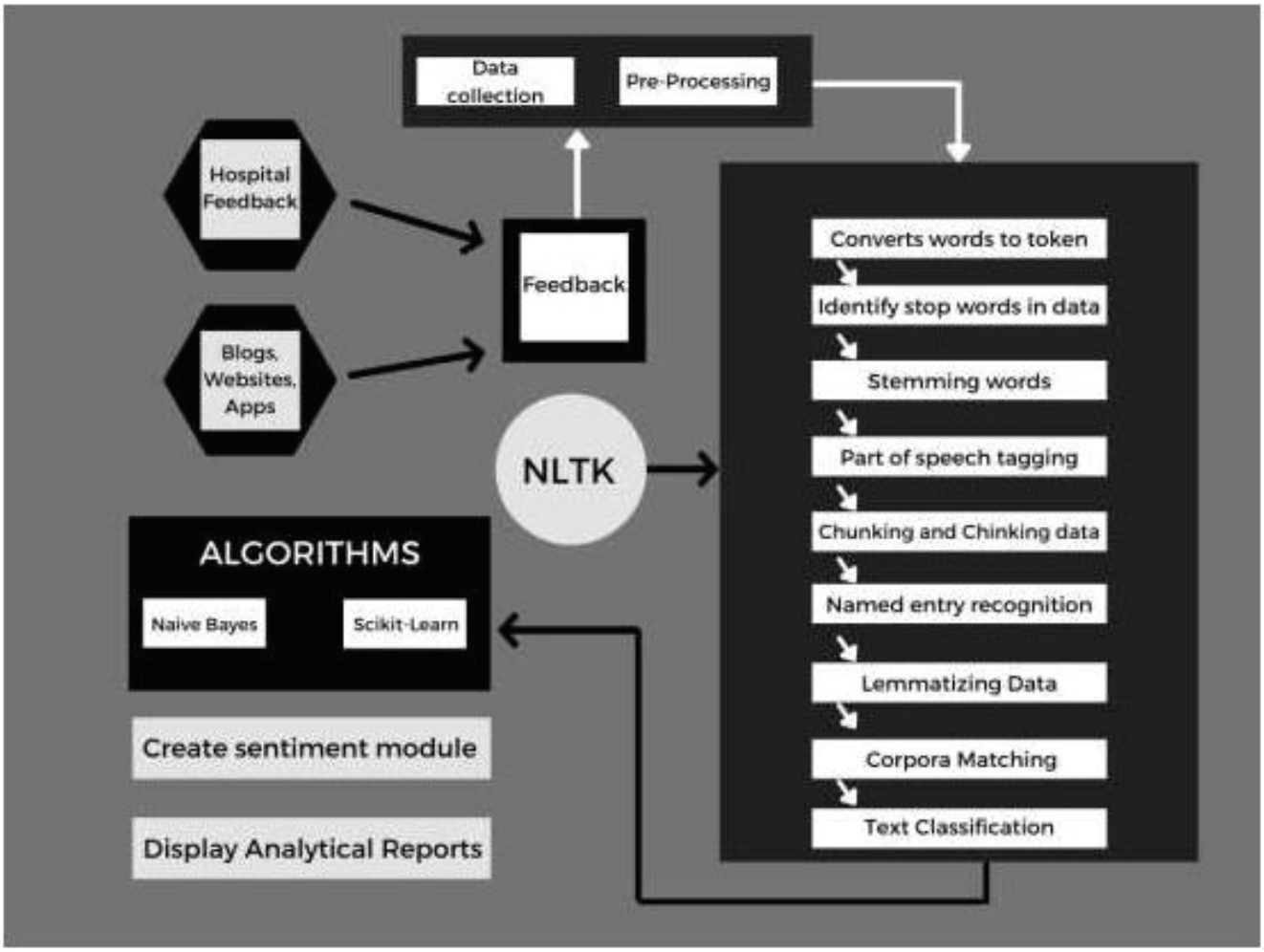

Fig 2: Sentiment Analysis using flowchart

TABLE II. TOPIC MODELING THEMES

Topic List	Theme	Keywords
Topic 1	Syllabus Quality	Worth, Content training, Value for money, Focused, Certification
Topic 2	Faculty Expertise	Good Trainer, Helpful, Clear Doubt, ExperiencedFaculty
Topic 3	Pricing	17k, High Investment, Loose money
Topic 4	Subscription	Enrolled, No refunds
Topic 5	Content Delivery	Understanding, Technical, Project, ThoroughlyExplained
Topic 6	Customer Service	Worst attitude, quick response, Follow up
Topic 7	Academic Levels	Hadoop, Java Developer, OOPS, Learn scratch,Beginner
Topic 8	Duration	4 hours, one year, 2 weeks
Topic 9	Interface Quality	Network Issues, Loose Data, Delay, No pause
Topic 10	Technology	Cloud Infrastructure, Downloaded Content

Multiple keywords relevant to the study were found using the word cloud formation. Many keywords are related to the customers' attributes in the EdTech service platform they opted for and its service satisfaction. Few keywords like "Experience, Subscription, Quality" explain the critical topics for discussion about the service.

Fig. 3: Word Cloud for Review Corpus

The fig (3) shows the merged word clouds of overall good reviews and bad reviews. We could see that the word cloud shows the major keywords that represent the consumer sentiments and emotions like "support, good, experience" and sort. The VADER model of sentiment analysis has been used for the analysis. The VADER model makes the important keywords bigger according to the relevance and the times it has been used by the customers.

Fig. 4: Good Review Word Cloud (Left)/Bad Review Word Cloud (Right)

The below word cloud is formed from the consolidated data of all the 600 reviews combining all three platforms. In the word cloud, we can see major topics and related keywords such as "Teacher" and "Quality".

Fig. 5: Consolidated Data Word Cloud

B. Sentiment Analysis.

Fig (5) Shows the Violine Chart visualization of the three platforms. The violin charts explain the polarity of sentiments across the platforms taken for analysis. For Udacity, we can see that the score ranges from -5 to 22. This says Byjus has good and bad reviews distributed among positive and negative polarity regions in almost equal proportion. Volume plot for the platform represents the usage of different relevant keywords in the data set and the number of times it has been used. The position of each word defines the polarity of sentiments which can be positive, negative, or even neutral, and the horizontal distribution explains the usage rate of each keyword. When we look at Simplilearn, we can see that it ranges from -20 to +30, it signifies Simplilearn has a good proportion of positive opinions from people, whereas Udemy's polarity ranges from -20 to +50, but the horizontal distribution is in areas between -20 and 20 which means people have more opinionated comments in this region of polarity.

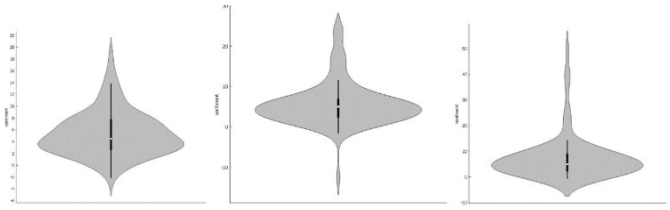

Fig. 6: Violine Plot Visualisation for Byjus, Simplilearn, Udacity

Fig 6 visualize the violin plot obtained from customer responses' overall negative, positive, and neutral sentiments—the compound values of the violin plot range from +1 to -1.2. We can see from the visualization that across all three platforms, customers have a uniform opinion in terms of negative and neutral sentiments. In contrast, the positive violin plot is broader in its volume, which defines the diversity of opinions within in the positive emotions of the customers.

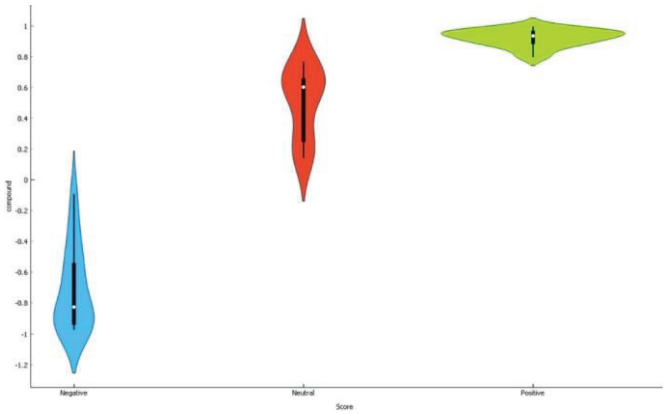

Fig. 7: Consolidated Violin Plot for Sentiment Score

C. Thematic Analysis

From the below table it was seen that Syllabus Quality, Pricing, Expertise of faculty etc. are some of the themes or topics which were mostly opinionated by the users. The keywords are also mentioned according to the frequency of usage. This explains that these themes or topics are to be taken into consideration while understanding the effectiveness of the EdTech industry during Pandemic and these are the concerned topics users mentioned in their reviews.

Syllabus Quality			**Pricing**		
Content Training	Focus		High	Investment	
Value for Money					
	Recommends		Loose		
Faculty Expertise			Subscription		
Good Training	Helpful		Enrolled	No Refund	
Experienced Faculty			Customer Service		
			Quick Response	Follow Up	
Content Delivery					
Project	Technical	Good Understanding	Academic Level		Duration
			Data Science	OOPS	One year 4 hours
Thorough Explanation			Scratch	Beginner	2 Weeks
Interface Quality		Technology			
Network issues	Loose Data	Cloud Infra Downloads			
Delay	No pause				

VI. CONCLUSION AND FUTURE WORK

The research paper concludes that the method of sentiment analysis enables us to understand the consumer emotion, sentiments, and attitude towards the three platforms of EdTech taken for the study and analyze the factors customers most discussed. Web scrapping and orange software environment helped us to distinguish the positive, negative and neutral consumer sentiments. Customer reviews about Byjus through mouthshut.com show the most positive

responses when compared to Udacity and Simplilearn. Users are satisfied with the user interface and the content quality of Byjus. The platform was mainly delivered to students until the secondary level of education, but customers were delighted with the relevant syllabus they provided during the Pandemic. Compared to Udacity and Simplilearn, there was dissatisfaction concerning the faculty expertise and pricing. The smooth interface of Byju's makes learning much more effortless, whereas Udacity and Simplilearn have a more complicated UI. Delivery of content through the mobile app was a value proposition that users have positive polarity for. What makes Byju's unique and improves customer satisfaction is its personalized content and expert teachers all over India. In contrast, the communities of Udacity and Simplilearn are small, and minor customization is possible. Users show gratitude, happiness, appreciation through their comments in the case of Byju's. The pricing of Udacity and Simplilearn induces much dissatisfaction, and the inefficiency in aftersales service produced anger and sadness in users. Further studies can be done to analyze user sentiments using sentiment analysis and topic modeling as this study showed possible results for analysis.

REFERENCES

[1] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis," *Front. Artif. Intell.*, vol. 0, p. 42, Jul. 2020, doi: 10.3389/FRAI.2020.00042.

[2] Koehler, M.J., Greenhalgh, S.P., & Zellner, A. (2015). Potential Applications of Sentiment Analysis in Educational Research and Practice – Is SITE the Friendliest Conference?

[3] Z. Tong and H. Zhang, "A Text Mining Research Based on LDA Topic Modelling,"pp. 201–210, May 2016, doi: 10.5121/CSIT.2016.60616.

[4] M. Rao, D. M, and A. Palathil, "An Analysis of Customer Sentiments Towards Education Technology App: A Text Mining Approach," *SSRN Electron. J.*, Mar. 2020,doi: 10.2139/SSRN.3563577.

[5] K. K. Bhagat, S. Mishra, A. Dixit, and C.-Y. Chang, "Public Opinions about Online Learning during COVID-19: A Sentiment Analysis Approach," *Sustain. 2021, Vol. 13, Page 3346*, vol. 13, no. 6, p. 3346, Mar.2021, doi: 10.3390/SU13063346.

[6] C. B. Asmussen and C. Møller, "Smart literature review: a practical topic modelling approach to exploratory literature review," *J. Big Data 2019 61*, vol. 6,no. 1, pp. 1–18, Oct. 2019, doi: 10.1186/S40537-019-0255-7

[7] Tashmin Khamis.” The COVID-19 pandemic: a catalyst for creativity and collaboration for online learning and work-based higher education systems and processes” https://www.emerald.com/insight/content/doi/10.1108/JWAM-01-2021-0010/full/html

[8] Melissa Laufer and Anne Leiser “Digital higher education: a divider or bridge builder? Leadership perspectives on edtech in a COVID-19 reality” International Journal of Educational Technology in Higher Education, Volume 18, Article Number: 52 (2021)

[9] Yamini Chandra “Online education during COVID-19: perception of academic stress and emotional intelligence coping strategies among college students” ISSN: 2046-3162 Article publication date: 6 October 2020 Issue publication date:10 March 2021

[10] Dr Nidhi and Arti “Work-Life Balance During Covid-19(Pandemic) - A Social Media Analysis” PSYCHOLOGY AND EDUCATION (2021) 58(2): 11730-11737ISSN: 0033-3077

[11] Nimasha Arambepola “Analysing the Tweets about Distance Learning during COVID-19 Pandemic using Sentiment Analysis” Conference: International Conference on Advances in Computing and Technology (ICACT–2020) At: University of Kelaniya, Sri Lanka

[12] Nandi, Biswarup et al. “Text based sentiment analysis.” 2017 International Conference on Inventive Computing and Informatics (ICICI) (2017): 9-13, doi: 10.1109/ICICI.2017.8365326

[13] Frankvan Cappelle “An analysis of the reach and effectiveness of distance learning in India during school closures due to COVID-19” International Journal of Educational Development Volume 85, September 2021, 102439