Observation of Online vs Offline Learning Experience

Siddharth Padhiar
Data Science
MPSTME, NMIMS
Ahmedabad, India
padhiarsiddharth@gmail.com

Karan Mehta
Data Science
MPSTME, NMIMS
Mumbai, India
karanpmehta5161@gmail.com

Juhi Patel
Data Science
MPSTME, NMIMS
Mumbai, India
pateljuhi027@gmail.com

Dr. Siba Panda
Data Science
MPSTME, NMIMS
Mumbai, India
siba.panda@nmims.edu

***Abstract*—As the outbreak of COVID-19 increased in various countries. India is also majorly affected with the COVID-19 by that education system is affected, and it has transferred the traditional face-to-face teaching to online education platform. Considering student's perspective on both online and offline learning mode in India, we conducted a survey to collect the data. In that survey questionnaire, focus was on the factors and situation which can affect the education system. Using that data, we used Kruskal Wallis test to collect the evidence for which learning mode is better and Naive Bayes Algorithm, we were able to conclude the results.**

Keywords—Survey, Research Approaches, Learning Environments, Educational Reform, Student Teaching, Predictive Analysis, Descriptive Analysis, Data Analysis, Questionnaire

I. INTRODUCTION

Online learning is education that takes place over the Internet. However, online learning is just one type of "distance learning"- the umbrella term for any learning that takes place across distance and not in a traditional classroom. A traditional classroom approach is, learning space in which the teacher provides face-to-face instruction to students and communication between and among teacher and students is face to face [1]. The COVID-19 has pandemic changed learning in many unprecedented ways. In response to the outbreak of the epidemic, the online classroom has become a necessary way to maintain normal teaching order. Students had to not just move to online learning but also keep a social distance and it is quite challenging for some to adjust to the 'new normal' and missed the in-person interaction with their friends and professors. Universities need to know how students feel about distance education and for that small survey was conducted to know student's point of view.

Online and offline both the platforms have their own advantages and disadvantages. In online lectures students and faculties saves their travelling time, and they can spend more time with family members. In online lecture most of the faculties record their lectures so if there is any issue then students can access to those recording whenever they want to, so it is beneficial for students. But there can be lack of interaction with faculty in online lectures as compare to offline lecture. However, these online education platforms have problems such as network glitches and the inability to replay live sessions sometime. It is necessary to understand whether online learning platform is meeting the needs of teachers and students or not. Technical problems are one of the main stumbling blocks of online training, especially in India where technology is not as developed as other countries. Very often, there are compatibility issues with operating systems, browsers, or smartphones. All these factors can affect their studies. Sometimes, the online world, no matter how enriching it may be, can become too small for the student and they may need a physical space where they can resolve their queries and practice with real tools [2].

This research study aims to understand student's point of view related to this topic, whether they are comfortable with online learning or not. If they are not comfortable with online learning platform then what is the reason behind that and why they prefer offline i.e. traditional learning approach over online learning. The objective of this research is to know what learning platform students prefer and why. In this survey we asked students twenty questions related to both the platforms and some of their basic information. We also asked them to rate few of the factors which were important for this research. Here we were totally focused on student's opinion and their thoughts.

978-1-6654-9291-1/22/$31.00 ©2022 IEEE.

II. DATA COLLECTION

In this paper, data are obtained through a questionnaire survey. The questionnaire was circulated through various social media platforms and collected more views of students based on their online and offline study experience. Questionnaire was targeted to various age groups and different streams' students through which collected data is more versatile. Also, this questionnaire give summary about general characteristics and opinions of a group of students. Through this questionnaire successfully 319 students surveyed. So, this paper combines the two study experiences to comprehensively acquire online and offline study experience data of users [6].

Due to COVID-19, online teaching platforms requirements goes up. Simultaneously, there are some issues in online study platform in India and this questionnaire is collecting information about what are the factors and situations which are affecting the study experience. In the questionnaire, it is difficult to determine all the factors that affects online and offline study methods [3]. To obtain more targeted evaluation of student experience, this study adopted a questionnaire survey, whose targets were School, High School, Diploma, Undergraduate and Postgraduate students which contains different age groups. Also, it includes questions related to different education fields, personal information, interaction with faculties and classmates, learning experience in both online and offline platforms, technology they are using and issues while using it and accessibility of resources like library resources, labs, and extra curriculum activities etc [9]. By analysing relevant literature, we have designed the questionnaire with different parts, as demonstrated in Table I.

TABLE I. QUESTIONNAIRE

Classification of Investigation	Content of Investigation
User's behaviour on teaching methods	Learning experience, device, technical issues, usage intention, content
User experience	Degree of satisfaction, interactivity, perceived value, accessibility, exam experience
Basic information	Gender, Age, education level and background, members in family, study rooms and area

III. DATA EXPLORATORY

Data exploratory is an approach to analyzing data in graphical representation. It is first step before applying any statistical techniques. Exploratory analysis gives idea about missing values, values range, and outliers in dataset. Also, it is useful to generate data insights and its structure. It checks assumptions associated with any model fitting or hypothesis test. Exploratory analysis is also useful to create a ranked list of relevant factors and eliminate unnecessary variables in data analysis. Based on the data we received by taken survey we performed exploratory analysis which gave us insights of the data.

In the questionnaire survey which we conducted, 67.7% of respondents were male while 32.3%% of respondents were female. Respondents which we mainly targeted were students, and 75.8% of the students who answered the survey were in between the age of 19 to 23. Most of the students were undergraduate and postgraduate students. From that 68.8% of the students were doing engineering and 14.1% of the students were from the science background. One of the questions in the survey was which device they use for online lecture and 72.7% of the students selected Laptop as a learning device and 27.3% selected Mobile. Key questions in the questionnaire were analyzed to understand the data characteristics.

On the question, which mode is preferred more based on the rating students have given for learning experience [1]. So, based on the platform which students have selected they have provided the rating from 1-5. In offline mode around 205 students gave their rating between 3-5 which we can consider as good rating, whereas in online very few i.e. around 47 students only gave their rating between 4-5 and if we compare both the platforms then offline mode has more weightage then online.

The results also shows that male and female students both feel that in offline mode they are more interactive as compared to online because here ratio of online mode is much less then offline.

One of the questions in survey where we asked students that do, they face any internet issue while attending online lectures. Out of 319 students around 237 students said that yes, they face internet issue and that students who selected yes also said that mode which they prefer is offline mode.

So here we can say that in India people face internet issue a lot, and because of that they prefer offline as a learning mode.

IV. METHODOLOGY

Due to Covid-19 the learning prospective of many educational bodies have changed. In online learning methodology there are certain tools which are widely used in industry such as Zoom, MS Teams, Google Meet, and Skype [4]. While in offline learning classroom is an only option. So, this project will help to find out preferable learning mode for students in India. Based on Individual's prospective and profile, these following research questions are

formulated. The key research questions this study aims to answer are as follows:

- Because of Covid-19 is there any difference in outcome of demographics on Mode preferred?
- Because of Covid-19 is there any difference in effect of degree of satisfaction on Mode preferred?
- Because of Covid-19 is there any difference in effect of technological aspects on Mode preferred?
- Because of Covid-19 is there any difference in effect of time spend for attending lectures, on Mode preferred?
- Because of Covid-19 is there any difference in effect of resource availability by educational body on Mode preferred?
- Because of Covid-19 do you feel distracted while studying online?

Data Collection questions were based on these related questions through surveying and statistical analysis was conducted in the following flow shown in Fig. 1. The data received by the questionnaires will help us to determine which learning mode is preferable for students [3].

A. Independent Variables

1. Demographics

The demographics of a person were obtained by incorporating variables such as gender, age, family details, living area, education details. The targeted group were students with different educational level (School, High School, Diploma, Undergraduate, Postgraduate), Fields (Commerce, Arts, Science, Diploma, Engineering, Medical, Management) and with different age group (Less than or equal to 13, 14 - 18 ,19 - 23, 24 - 27, 28 or more than that), Area(Metropolitan, Rural) and family details (Members, personal rooms) was collected and were important aspects which must be studied to get a proper impact understanding of it on learning mode preferred.

2. Degree of Satisfaction

The degree of satisfaction was a part of the survey in which students were asked about their experience they were getting on the mode preferred by various aspects like learning experience, online experience, offline experience, faculty interactions, teacher helpful, classmate's interaction in which all targeted group were asked to respond on a forced-choice Likert scale between 1-5. Asking about personal experience for preferred learning mode was also collected to understand the impact on learning mode and which mode is best for students.

3. Technical Issues

India is growing fast in terms of technology but it's not strong as compare strong to different countries like USA, China, Japan etc. Due to covid-19, it was a complete change and a challenging situation for universities a students' from shifting to online studies many students were unable to manage their own device and also unable to manage a proper internet connection this is because India is not much familiar with technology. So, this data was important to be collected from the targeted group and must be studied to get a proper impact understanding of it on learning mode preferred.

4. Miscellaneous Variables

Miscellaneous Variables includes time Spend on lectures, resource availability, study compromised and distraction during lectures. Due to COVID-19 many universities were not able to conduct normal class similar to offline classes. In survey students were asked about this all variables and answer received were different individually based on targeted group educational level.

B. Dependent Variables

The key part of our dataset is the dependent variable as our analysis hinges on the significance of the independent variables affecting this variable. However due to COVID-19 many of the universities shifted the lectures to online. So, to find out which learning mode is better for targeted group a question was asked "Which learning mode do you prefer?" which includes options (Online, Offline) and to state which learning mode is best for targeted group analysis was done accordingly.

V. DESCRIPTIVE ANALYSIS

The data distribution table in this section enable us to draw relevant insights from the data and obtain a general understanding of the background of the respondents. Table given in this section describes the distribution of the categorical variables in terms of percentage of the number of observations in the respective categories to the total number of observations.

From given table we can observe that there is a significant increase in the percentage of people preferring offline learning approach (i.e. traditional face-to-face learning), whereas there is decrease in online learning approach [8]. It is also observed that students feel less interactive to their fellow classmates and faculties in online learning platform. Also, with given table we can also see that almost 62.69% students feel distracted while studying online and because of that they have selected offline learning mode as their preferred mode. 65.83%

students said that resources are not available while studying online due to the pandemic and because of that mode which they prefer is offline learning [10].

TABLE II: DISTRIBUTION OF DATA FOR CATEGORICAL VARIABLES

Variables	Category	"Mode Preferred 1 (in %) Offline"	"Mode Preferred 2 (in %) Online"
Gender	Female	26.018	6.269
	Male	54.858	12.852
Age Group	14 -18	9.717	0.626
	19 - 23	61.128	14.733
	24 - 27	7.836	3.134
	28 or more	2.194	0.626
Education Level	Diploma	2.507	0.313
	High School	4.702	0.626
	Postgraduate	17.862	3.76
	School	2.194	0
	Undergraduate	53.605	14.42
Education Field	Arts	1.253	0.313
	Commerce	6.269	0.626
	Diploma	0.940	0.313
	Engineering	56.112	11.598
	Management	1.253	0.940
	Medical	1.880	0.313
	Others	3.134	0.940
	Science	10.031	4.075
Members in Family	2	1.253	0.3136
	3	15.673	4.075
	4	39.184	7.213
	5	13.166	4.388
	6 or more	11.598	3.132
Personal Room	No	38.244	4.702
	Yes	42.633	14.42
Device Preferred	Desktop	4.0752	1.253
	Laptop	58.934	13.79
	Smartphone	17.868	4.075
Internet Issue	No	19.435	7.523
	Yes	61.442	11.598
Studies compromise	No	13.166	12.852
	Yes	67.711	6.269
Area	Metropolitan	66.77	13.793
	Rural	14.106	5.329
Time Spent	1 - 3 hours	20.376	4.388
	3 - 5 hours	24.451	7.210
	5 - 7 hours	27.899	5.6426
	7 - 10 hours	8.150	1.888
Resource accessibility	No	65.830	14.420
	Yes	15.047	4.702
Classmate Interaction	Much	65.203	8.777
	Very Much	15.673	10.344
Distraction in Online study	No	18.181	10.97
	Yes	62.695	8.150
Learning Experience	Very bad	3.7617	0.626
	Bad	8.1504	1.880
	Good	33.542	3.134
	Very good	29.780	9.090
	Excellent	5.6426	4.388
Online Exam Experience	Very bad	8.777	1.567
	Bad	17.554	3.134
	Good	29.467	5.015
	Very good	17.554	6.269
	Excellent	7.523	3.134
Offline Exam Experience	Very bad	2.507	1.253
	Bad	3.134	1.880
	Good	9.404	7.210
	Very good	42.00	5.956
	Excellent	23.824	2.821
Faculty	Very bad	7.210	1.567
	Bad	15.987	1.880
Interaction	Good	32.6018	5.642
	Very good	21.6300	6.583
	Excellent	3.4482	3.448
Teachers Helpful	Very bad	3.1347	1.2539
	Bad	8.150	2.1943
	Good	26.33	4.702
	Very good	32.601	7.523
	Excellent	10.658	3.448

VI. RESULTS AND FINDINGS

This section focuses on the questionnaire validity test created to determine the relationship between the independent and dependent variables. During the epidemic period, total of 319 responses were received. After the data collection, R and Microsoft excel tools were used to perform validity test and reliability test. The corresponding results are mentioned in the following sub-section [7].

A. Reliability Test

In this study, Cronbach's Alpha was used to test the internal consistency of the questionnaire data, whose coefficient was between 0 and 1. In general, a coefficient greater than 0.6 indicates that the questionnaire can passes the internal consistency test. In contrast, a coefficient less than 0.6 indicates that some questions must be discarded. The reliability test results are presented in the Table. In this questionnaire, four Cronbach's Alpha coefficients were all greater than 0.6, indicating that the internal reliability of each first-level indicator of the questionnaire was high [12].

TABLE III: QUESTIONNAIRE DATA RELIABILITY RESULT

Index	Cronbach's Coefficient	Number of Questions
Degree of satisfaction	0.63	5
Personal factor of users	0.66	3
Quality of interaction	0.71	3
Education wise learning experience	0.61	2

B. Validity Test

The hypothesis is formulated for each of the independent variables with the dependent variable Mode Preferred.

H_0 = Variables Xi and Y are not dependent.
H_A = Variables Xi and Y are dependent.

The validity test can be divided into content validity and structure validity. The questions in this questionnaire scale used relevant literature for reference to ensure high content validity. As all the variables in questionnaire were categorical, Kruskal Wallis Test was performed to check the dependency

of independent variables with dependent variable. Some of the variables from the questionnaire passed the Bartlett Test and Kruskal Wallis test.

1. Bartlett Test

Pre-requisite to undergo Kruskal Wallis test is that dataset should pass the Bartlett's Test first. Significance value for passing the Bartlett's test is 0.05, so if any variable is having p value less than 0.5 significant value will get rejected. A p-value is the probability which signifies if the null hypothesis is true. Here there are total 19 independent variables, and out of these 19 variables 4 variables got rejected as they were having p values less than the significant value i.e., 0.5. The Bartlett's Test results are presented in below Table.

TABLE IV: BARTLETT'S TEST

Variables	P Value	Null Hypothesis
Internet Issue	0.1558	Accept
Gender	0.9117	Accept
Age	0.6163	Accept
Education	0.0004	Reject
Field	0.0202	Reject
Members	0.4929	Accept
Room	0.1765	Accept
Device Preferred	0.8852	Accept
Studies Compromised	0.0108	Reject
Area	0.0778	Accept
Time Spent	0.7693	Accept
Learning Experience	0.2659	Accept
Online Exam	0.5797	Accept
Offline Exam	0.1355	Accept
Resources Availability	0.2794	Accept
Faculty Interaction	0.1564	Accept
Classmate Interaction	0.0141	Reject
Teachers Helpful	0.1622	Accept
Distracted	0.0733	Accept

2. Kruskal Wallis Test (KW Test)

After completing Bartlett's test, 4 independent variables i.e. Field, Members, Studies Compromised and Classmate Interaction were not having any impact on the dependent variable because their p values were less than the significant value 0.05. So, with the remaining variables Kruskal Wallis Test was performed. 15 variables passed the Bartlett Test and on that 15 variable KW test was performed. Out of 15 independent variables, 7 were accepted and 8 were rejected as their p values were less than 0.05. The results of the Kruskal Wallis validity test are presented in below table [11].

TABLE V: KRUSHAL WALLIS TEST

Variables	P Value	Null Hypothesis
Gender	0.9263	Accepted
Age	0.0239	Rejected
Members	0.5278	Accepted
Rooms	0.0013	Rejected
Device	0.9076	Accepted
Internet issues	0.0155	Rejected
Area	0.06461	Accepted
Time Spent	0.7751	Accepted
Leaning Experience	0.0001	Rejected
Online Exam	0.0217	Rejected
Offline Exam	2.26E-06	Rejected
Resources Availability	0.2917	Accepted
Faculty Interaction	0.0013	Rejected
Teachers Helpful	0.6713	Accepted
Distracted	7.25E-08	Rejected

VII. NAIVE BAYE'S CLASSIFICATION

Naive Bayes' Classification can be used to Predict the overall performance of the model based on the prior knowledge and current evidence. After KW test, rejected variables which were impacting dependent variable were taken for model prediction. Total 8 independent variables (Age, Rooms, Internet issue, Learning Experience, Online Exam, Offline Exam, Faculty Interaction, Distracted) were rejected in KW test as p values were less than 0.05 and that 8 variables were taken to perform Naïve Bayes' Classification for predicting the accuracy of the model [5].

Based on Conditional Probability and Bayes' theorem:

$$P(c/x) = \frac{P(x/c)P(c)}{P(x)}$$

- $P(c|x)$ is the posterior probability of *class* (*target*) given *predictor* (*attribute*).
- $P(c)$ is the prior probability of *class*.
- $P(x|c)$ is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*.

A posterior probability means updated probability of an event occurring into consideration after receiving new information. It is used to compare the conditional or posterior probability for different classes.

After performing Naïve Bayes' Classification, with the accuracy of 86.6% model was able to predict which mode user will prefer according to the user data as shown in output.

TABLE VI: OUTPUT OF NAIVE BAYES'

Confusion Matrix		
	Offline	Online
Offline	75	2
Online	11	9

VIII. CONCLUSION

This study collected user experience data on both online and offline education platforms in India during the COVID-19 pandemic. Through data analysis of user reviews, we concluded that offline study platform is more preferable by students in India, while online study platforms encountered several problems, like Internet issues, unavailability of study rooms, problems in resource availability, lacking in faculty interaction and distraction factors affect the study. Through the collected data, it was found that field of study, family members, study compromised, and classmate interaction had no direct impact on their study, while study rooms availability, internet issues, examination experience, faculty interaction and distraction had impacted on their study. Descriptive analysis is more useful to generate more insights from data and give more understanding about categorical features in percentage. After collecting data questionnaire validity test is performed by Cronbach alpha through which questionnaire passes the internal consistency test. After that Validity Test is performed which is divided into content validity and structure validity. Before Kruskal Wallis Test, Bartlett's Test is performed for test the homogeneity of variances. In nonparametric test the Kruskal Wallis Test is used for categorical dependent variable. In addition, Naïve Bayes Theorem was used to predict user leaning mode, and accuracy reached approx. 78%. And through data analysis and statistical evidence it is proved that learning mode which students prefer is offline i.e. traditional face-to-face learning.

REFERENCES

[1] Wilbur, K. Evaluating the online platform of a blended-learning pharmacist continuing education degree program. Med. Educ. Online 2016, 21, 31832. [PubMed]

[2] Ryan, C.; Young, L.; McAllister, M. The impact of an online learning platform about nursing education on enrolled nurse preceptor teaching capabilities: A pre-post-test evaluation. Contemp. Nurse 2017, 53, 335–347. [PubMed]

[3] Has the COVID-19 pandemic affected the susceptibility to cyberbullying in India? [PubMed]

[4] Chan, T.; Sennik, S.; Zaki, A.; Trotter, B. Studying with the cloud: The use of online Web-based resources to augment a traditional study group format. Can. J. Emerg. Med. 2015, 17, 192–195. [PubMed]

[5] S. Karthika and N. Sairam A Naïve Bayesian Classifier for Educational Qualification [PubMed]

[6] Thor, D.; Xiao, N.; Zheng, M.; Ma, R.; Yu, X.X. An interactive online approach to small-group student presentations and discussions. Adv. Physiol. Educ. 2017, 41, 498–504. [PubMed]

[7] Botelho, J.; Machado, V.; Proenca, L.; Rua, J.; Delgado, A.; Joao, M.J. Cloud-based collaboration and productivity tools to enhance self-perception and self-evaluation in senior dental students: A pilot study. Eur. J. Dent. Educ. 2019, 23, e53–e58. [PubMed]

[8] Chapman, S.A.; Goodman, S.; Jawitz, J. A strategy for monitoring and evaluating massive open online courses. Eval. Programplan. 2016, 57, 55–63. [PubMed]

[9] Miri, B.; Gizell, G. Novice Researchers' Views About Online Ethics Education and the Instructional Design Components that May Foster Ethical Practice. Sci. Eng. Ethics 2020, 26, 1403–1421. [PubMed]

[10] Kamali, A.; Kianmehr, L. The Paradox of Online Education: Images, Perceptions, and Interests. US China Educ. Rev. 2015, 15, 591–601. [PubMed]

[11] Rehan Ahmad Khan Sherwani, Huma Shakeel, Wajiha Batool Awan, Maham Faheem & Muhammad Aslam: Analysis of COVID-19 data using neutrosophic Kruskal Wallis H test. [PubMed]

[12] Keith S. Taber: The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education [PubMed]