STUDENT'S PERCEPTION TOWARDS ONLINE LEARNING DURING COVID-19 PANDEMIC- AN EMPIRICAL STUDY USING MULTIPLE LINEAR REGRESSION

- *DEGA BALAJI VARA PRASAD, AVALA JAYA ISWARY ABHIRAM*

Abstract

Covid-19 jeopardized academic activities and gave a challenge to all the educational institutes to think of new ways to teach. Online learning has vastly started growing in the institutes, posing teachers to adapt to the new technologies and tools. Indian institutes have adapted these tools but still, there is a lack of knowledge on how much this online teaching has impacted students to their usual learning method of face-to-face classes. In this study, we try to find how much the technical constraints of an average Indian student like device suitability, bandwidth availability, resource accessibility are impacting their learning. Using the Multiple Linear Regression model with software like python, R, and Tableau we try to predict the marks of a student based on the variables which contains all the constraints.

We also use some statistical concepts like correlation to find linear dependency among the variables and try to visualize it, and consensus factor for getting more insights on the Likert scale ordinal variables. In this study, we created different multiple linear regression models based on different variables from our primary data, collected through the Google forms questionnaire. Our study says that more than 40% of students are satisfied with their marks and have responded positively that online classes are better than face-to-face classes. Around 55% of students use only smartphones and 35% uses both smartphone and laptop for online classes, comparatively, students with only smartphones are less fortunate to get good marks.

As this e-learning wave is a recent development, teachers, as well as students, are in the process of adapting this new teaching and learning methodology. In this prevailing situation of virtual teaching and setting of new normal of teaching-learning methodology, it becomes all more important to get to know the opinion of learners and to explore learners' inclination towards this novel teaching methodology, such as their degree of adaptation and amendment if any they would like to suggest for the same or want to reject it altogether. Against such a backdrop, this study aims to examine the perception of e-learning during the COVID-19 lockdown period.

Introduction

The effect of information technology on human life is immense, and its role in education too cannot be subsided. In the current scenario of the COVID-19 pandemic, the contribution of information technology has gained momentum due to the closure of educational institutions that raises challenges for students' learning. During this quarantine time, information technology is serving as the solution for the ongoing learning process through innovative and learning management systems. It has provided an opportunity for educators to implement IT solutions for teaching, as well as evaluation for the completion of the course work of students.

The efforts of stakeholders namely teachers, students, and institutional administrators are on for the optimal use of the technology and efficient learning process. The ultimate goal is to minimize the learning gap that arises due to lockdown. Educational institutions and students across the world have accepted and appreciated the online platform of learning. The reasons for this acceptability are ease of use, learning flexibility, and a controllable environment. However, despite its multiple advantages, there are quite a few limitations of e-learning such as social isolation, face-to-face interaction between teacher and student, connectivity issues, etc.

E-learning has never been adopted and accepted as real learning or the formal mode of education before this ongoing pandemic that compelled us to resort to electronic learning solutions the world over. Now, at the hour of pandemic crisis, most educational institutions are exploring and approaching e-learning to make it easy for students to work out a new normal. Also, various e-teaching software is being explored by teachers or educators to bring maximum possible ease for their students.

Method

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent variable y. The population regression line for p explanatory variables $x_1, x_2, ..., x_p$ is defined to be $\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$. This line describes how the mean response μ_y changes with the explanatory variables. The observed values for y vary about their means μ_y and are assumed to have the same standard deviation σ. The fitted values $b_0, b_1, ..., b_p$ estimate the parameters $\beta_0, \beta_1, ..., \beta_p$ of the population regression line.

Since the observed values for y vary about their means μ_y, the multiple regression model includes a term for this variation. In words, the model is expressed as DATA = FIT + RESIDUAL, where the "FIT" term represents the expression $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$. The "RESIDUAL" term represents the deviations of the observed values y from their means μ_y, which are normally distributed with mean 0 and variance σ. The notation for the model deviations is ϵ.

Formula and Calculation of Multiple Linear Regression

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$$

where for i=n observations:

y_i =dependent variable

x_i =explanatory variables

β_0 =y-intercept (constant term)

β_p =slope coefficients for each explanatory variable

ϵ =the model's error term (also known as the residuals)

Question formulation, questionnaire preparation, and data collection

The major factors that affect a student's learning objective are technical constraints like the network bandwidth, device availability, accessing educational resources through smart devices. Since the physical availability of books was temporarily abolished during the covid period, and many people have not chosen to purchase a hard copy, accessing soft copies through smart devices was the only way. Through the network bandwidth attribute, we can

derive how a student is engaged in online classes by comparing the total data usage and data used online to know the interest of participation of students in online classes per day.

The ratio of data used for online classes to total data usage can justify how much he is interested in online classes, while by collecting 5 points Likert scale data on the perception of the online classes through which we measure the satisfaction and perception on online classes of the student, and with the similar scale we can also collect the degree of occurrence and interest of participation on variables like; participation in an online class, using social media, cheating in exams, and get bored in classes, etc...

Considering all the questions and formulating the basic hypothesis on what are the variables that could affect the good prediction of our response variable, we formulated our null hypothesis, and now we have to do data processing to make the data ready for model training and evaluate different models and compare the accuracy and residual values for good model selection.

Univariate exploratory data analysis

Most of the raw data collected in any survey whether collected through the normal survey, online survey, or even the secondary data, the data has at least some part missing or contains outliers in it. The variables were validated based on the data type of the variable, if it's numerical, it can be visually validated for outliers or missing values with a box plot that uses interquartile range to visually represent the count and how deviated it from the 25% and 75% quartiles. The line plots ware made to check for outliers, which can be easily recognizable by their peakedness. The best way to detect the outliers is to demonstrate the data visually. All other statistical methodologies are open to making mistakes, whereas visualizing the outliers gives a chance to take a decision with high precision.

After the univariate exploratory analysis, we would have chosen some numerical variables to be good data, now we can check for regression coefficients and select which numerical variables could suit our model better. We create a heat map as shown in the figure, which shows the linear relationship between each of our numerical variables to the response variable, and the higher the correlation, the better it can impact the response variable for good

accuracy and better prediction. Also, using the heatmap, we reduce the multicollinearity the exists in the data set.

Bivariate exploratory data analysis

While the univariate data analysis is used to find the outliers and to remove the anomalies in the data, using the bivariate exploratory data analysis, we compare the Likert scale or categorical ordinal data with the response variable to find if the ordinal scale categories have any positive preference or any patterns concerning to the response variable by taking sides on the distribution of response variable, makes our testing variable more preferable to be select for model training. The more the categories have differences or according to the consensus factor, the more the variable strongly states on one opinion, the more it's possible the variable to show impact on good and accurate prediction.

We also do bivariate analysis with predictor variables to find the correlation among them which may cause multicollinearity in our model, which leads to over fitting of our model. After the bivariate analysis, we selected a total of 16 variables and the remaining are preserved to find perception on online classes with Likert scale data. While the numeric data selected has higher good correlation, the categorical variables may also have high correlation, which can only be tested by creating a model.

Feature engineering

One hot encoding method, data preprocessing after variable selection to decrease effort in cleaning non-selected variables, and handling missing values. The 5 scale data is converted to 3 scale data to reduce the bivariate numeric variables that comes from one hot encoding method applied on the categorical variables, and, also to find the perception of students instead of calculating the consensus factor. We have selected our variables, now using our variables, we build our first model and select a better model using the accuracy indicating factors line AIC, Log-Likelihood, BIC and adjusted r square etc…

Basically, all machine learning algorithms use some input data to create outputs. This input data comprise features, which are usually in the form of structured columns. Algorithms require features with some specific characteristic to work properly. Here, the need for feature engineering arises. I think feature engineering efforts mainly have two goals, preparing the proper input dataset, compatible with the machine learning algorithm requirements and improving the performance of machine learning models. The features you use influence, more

than everything else, the result. No algorithm alone, to anybodies knowledge, can supplement the information gain given by correct feature engineering.

Missing values are one of the most common problems you can encounter when you try to prepare your data for machine learning. The reason for the missing values might be human errors, interruptions in the data flow, privacy concerns, and so on. Whatever is the reason, missing values affect the performance of the machine learning models.

Model building

We tried building 5 models through which we used stepwise variable method for variable selection. The first model consists of all the variables, just to observe the difference in accuracy and model indicator factors from all variable model to our final model. Model 2 is built and the variable causing multicollinearity and which has regression coefficient closer to zero are removed, since the collinearity causes reduction in the power of our model by increasing the importance of other variable that is highly correlated with, which causes over involvement of an unnecessary variable in our model. After removing all the unnecessary variables with less significance and by comparing the AIC value with a threshold of 2 for consideration, we selected all the variables that are Sloley needed to a good model and made our model 5. We found the model accuracy to be 81%(r squared value, 0.806).

Conclusion:

We have successfully completed the survey and built out a model for marks prediction using multiple linear regression. Also, with the additional categorical data that we collected, we draw inferences about student's perception of online classes. One of the crucial variable in the study is a number of hours spent on self study, more than 40% of the students spend more than 5 hours on self study excluding the time they spend on online classes, which was not expected in this stressful pandemic. Nearly 50% are having 5 hours of online classes and most of the students use 1 to 1.5 GB data just for online classes where they use 2 GB data on an average daily. More than **22% have said that their network speed is so bad that they couldn't even join the class sometimes**, even if they join, the video buffers and gets disconnected often. Even the remaining have the network issues, but there is no problem to continue studying through online classes.

The main obstacle in this online learning is the lack of computers. **Only 26% of students have a laptop/ desktop,** and most of them attend online classes through mobile, which they feel not

comfortable in attending lab classes or classes that have to be taught practically. While the count of students using mobiles for online classes are significantly higher than the students using laptops, around 50% of students prefer live online classes, 16.7% prefer pre-recorded classes with occasional doubt solving sessions and around 33.3% prefer self study with teachers assistance.

Social media has become part of every students' life, due to isolation, every one that are very socially interactive was in need of instant gratification. So people became more interested in posting pictures and updating their daily life through special features like reels in internet social platforms like Instagram. This has hugely and negatively impacted a student in his online learning process. Over 55% of students say they very often use social media while they attend online classes.

While the students are struggling with these problems, the exams have also been conducted through online mode. There is almost no technological way to reduce or prevent cheating, as students these days can always find loopholes to applications and programmes that are aimed at preventing cheating. The most straightforward and foolproof method of reducing or preventing cheating will be to include innovative questions, for which answers cannot be easily found anywhere. The fact that students copy in online exams has become a know fact to all, even the exam coordinators. **According to our survey, we found that 46% of students copies a great deal in online exams, 19% says they never cheated and the remaining were not willing to share their opinion. Some empirical studies done on relation between students and online classes, says, 73% students copies in online exams.**

Here is our final result. Students perception on online classes in covid pandemic. Nearly 60% of students disagree with the online classes and only 12.2% agree, whereas the remaining 28% stays neutral and 58.3% are not satisfied and, 41.7% say online classes have helped them in reaching their expected academic outcome.

References:

1. Khan, Mohammed A., Vivek, Mohammed K. Nabi, Maysoon Khojah, and Muhammad Tahir. 2021. "Students' Perception towards E-Learning during COVID-19 Pandemic in India: An Empirical Study" *Sustainability* 13, no. 1: 57. https://doi.org/10.3390/su13010057

2. Ding Y, Du X, Li Q, Zhang M, Zhang Q, Tan X, et al. (2020) Risk perception of coronavirus disease 2019 (COVID-19) and its related factors among college students in China during quarantine. PLoS ONE 15(8): e0237626. https://doi.org/10.1371/journal.pone.0237626

3. Yoo, DM., Kim, DH. The relationship between students' perception of the educational environment and their subjective happiness. *BMC Med Educ* 19, 409 (2019). https://doi.org/10.1186/s12909-019-1851-0

4. Nelwan, Melinda. (2020). Analysis of Student Perception on the Quality of Service Provided by a Private Higher Education Institution in East Indonesia. International Journal of Academic Research in Business and Social Sciences. 10. 10.6007/IJARBSS/v10-i4/7104.

5. Zayapragassarazan Z. COVID-19: Strategies for Online Engagement of Remote Learners [version 1; not peer reviewed]. *F1000Research* 2020, 9:246 (document) (https://doi.org/10.7490/f1000research.1117835.1)

6. Mamattah, R. S. (2016). Students' Perceptions of E-Learning (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-127612

7. Bączek, Michał MD; Zagańczyk-Bączek, Michalina MD; Szpringer, Monika MD, PhD; Jaroszyński, Andrzej MD, PhD; Wożakowska-Kapłon, Beata MD, PhD Students' perception of online learning during the COVID-19 pandemic, Medicine: February 19, 2021 - Volume 100 - Issue 7 - p e24821doi: https://doi.org/10.1097/MD.0000000000024821

8. Gismalla, MA., Mohamed, M., Ibrahim, O. *et al.* Medical students' perception towards E-learning during COVID-19 pandemic in a high burden developing country. *BMC Med Educ* 21, 377 (2021). https://doi.org/10.1186/s12909-021-02811-8

9. "Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques" Ho IMK, Cheong KY, Weldon A (2021) Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques. PLOS ONE 16(4): e0249423. https://doi.org/10.1371/journal.pone.0249423

10. https://towardsdatascience.com/feature-engineering-for-machine-learning-3a5e293a5114

11. https://towardsdatascience.com/statistics-in-python-collinearity-and-multicollinearity-4cc4dcd82b3f

12. https://towardsdatascience.com/how-to-upload-your-python-package-to-pypi-de1b363a1b3

Code:

https://github.com/BALAJI24092001/Student-Perception-Analysis/blob/main/StudentPercetionAnalysis.ipynb

Visualizations:

https://github.com/BALAJI24092001/Student-Perception-Analysis