

A DATA MINING APPROACH TO STUDY THE MENTAL HEALTH OF INDIAN STUDENTS DURING COVID-19 PANDEMIC

Dr.K.Kasturi, Associate Professor, Dept.of Computer Science, School of Computing Sciences, VISTAS, Chennai-117. Mail-Id: kasturi.scs@velsuniv.ac.in.

D.Devisanthy, Mphil Scholar, Dept.of Computer Science, School of Computing Sciences, VISTAS, Chennai-117. Mail-Id: devisanthy123@gmail.com.

Abstract

Objective: Since many students were being affected by the long break to the schools, lockdowns and change in their learning process, there is a need to analyse and find the psychological stress, anxiety and depression parameters of students during the COVID-19 Pandemic. **Methods:** In Data Mining using Quota Sampling technique, an online survey was conducted to the students of schools and colleges. The student's challenges, study patterns and stress levels have been analysed using their statistical scores for their responses to the questions. Using logistic Regression DM technique regression analysis was performed. **Methodology:** In order to explore the impacts of this pandemic on the lives of students, we conducted a survey of school and college students' ages from 15 to 24 from the various educational institutions in Tamil Nadu, India. The student data from the schools and colleges were collected through online surveys created by Google forms. The data were on the basis of analysing demographic details, stress prevalence, financial conditions and learning process of the students.

Keywords: COVID 19, Pandemic, Stress Prevalence, Students Depression, Logistic Regression, Quota Sampling.

Introduction

Data mining is an approach for analysing huge observational datasets that combines computer science and statistics knowledge. The goal of data mining is to identify unknown patterns in datasets and to present the information in novel ways. Data mining is a broad term that refers to a variety of analytical techniques. The outbreak of COVID-19 affected the lives of all sections of society as people were asked to self-quarantine in their homes to prevent the spread of the virus. The lockdown had serious implications on mental health, resulting in psychological problems including frustration, stress, and depression. In China, many studies assessed the health of adults who have been under the confinement period of one month during the corona pandemic. The paper shows the evidences to pay more attention to the people who were not infected by this virus, but they are depressed economically a lot by stopping their works during the lockdowns [1].

Related work

The study highlighted the impact of the COVID 19 potentially in the field of education ,that too in Universities and also deals with the mental health of their students , international students and academic staff members[2].Another study proposes the mental health of the youth after the occurrence of COVID 19 in the country of China. The study comprises of COVID status, Questionnaire regarding General health status .Also indicates the psychological distress and negative coping styles among the youth mental health status [3].

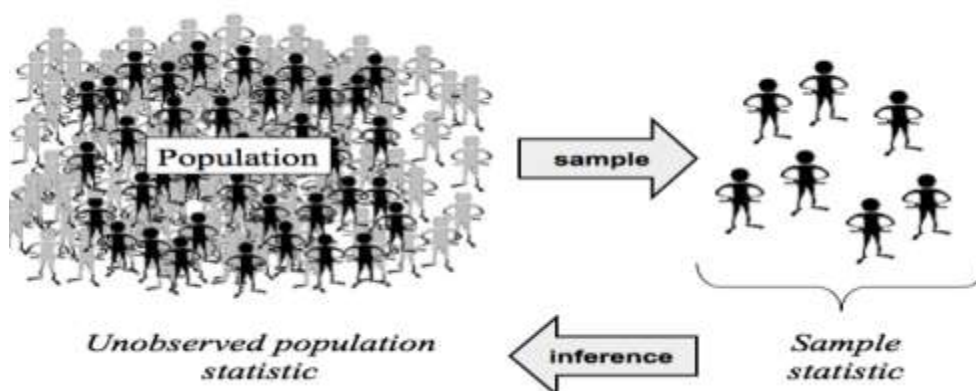
A study reveals about the mortality rate, psychological factors, students affected with psychological imbalances and challenges during the confinement period .Does statistical analysis and concludes to monitor the students continually to make them free from stress and the fear of pandemic. The

outcome reveals that more than 50 % of students showed moderate levels of stress and that too female students had more association with stress levels [4][5].

Sampling is the selection of subset of a large population in the survey of collecting data and that sample should represent the entire population, that is to make the unobserved population statistics also possible [6]. The flexibility in using Google forms such as feasible perspective, free of cost, covering wide areas makes online surveys significant in data collection with good quality for research purposes [7][8].

Proposed Work

Fig.No.1: Sampling Illustration



Data Collection

Student details were collected from the school and college students within the state of Tamil Nadu. Data were collected from the students studying 9, 10, 11 and 12 classes in schools and from students studying UG and PG courses in college and Universities. A total of 1243 valid data has been gathered from 642 girl students and 601 boy students.

Table No.1: Student Data Form

.No	Variables	Possible Values
1	Age	15Yrs – 18 Yrs : C1 19Yrs – 21 Yrs : C2 22Yrs – 24 Yrs : C3
2	Gender	Male, Female
3	Educational Standards	School, College, Universities
4	Residential Area	Pin code
5	Home Confinement	Yes, No
6	Feared by Pandemic	Never, Sometimes, Fairly Often, Very Often
7	Self-Management Stress	
8	Hurdles during Online classes	
9	Worried about Studies	
10	Depressive symptoms	
11	Economic crisis	

Quota sampling

This technique is applied in this study. Instead of researching a large population, by using few quotas, the researcher could get the whole picture of the population in less time. It allows subgroup selections to get desired outputs and ideas.

Steps in Quota Sampling:

1. Sample population can be divided into subgroups.
2. Find the Weightage of each subgroup.
3. To select the sample size.
4. To conduct survey according to quotas.

Logistic Regression

It is a statistical model that uses a different method for estimating the parameters, which gives better results. Feature selection also known as attribute or variable selection and it refers to the process of reducing the input variables for processing and for analyzing. Thereby finding the useful information or attributes from the existing data. This improves the quality of a built model by making it to process efficiently.

The regression machine tool effectively selects or discards the feature based on their usefulness and thus it finds the risk factors in the model and not just a prediction. It also improves the accuracy and decreases the training time. Using regression analysis the importance or significance of the variables can be derived.

Likert Plot

A Likert scale is a measure generally applied in research that employs survey questionnaires. It is the most extensively used approach to scaling responses in survey research, such that the term is often used interchangeably with rating scale, although there are other types of rating scales. The format of four-level Likert scale:

- Fairly Often
- Very Often
- Sometimes
- Never

Likert plots are often used to assess the participant's (student's) assertiveness to a particular question or a statement. The scale is called a summative scale because each item in the Likert scale can be analyzed by summing up the responses and scoring for a group of items can be created.

Experimental Results

R tool has been used for the implementation purpose. The student data using online survey consists of 11 simple questionnaire and the responses from the Google form have been recorded in .CSV files for the statistical analysis.

a. Selection of Significant Factors

Among 11 parameters listed in the Table1, Regression analysis has been performed using Logistic regression algorithm to identify the significant factors that are really affecting the students during the pandemic. According to Fig: 2 the parameters such as self-management stress, Financial stress and Hurdles faced during online classes are identified as the risk factors that are affecting the students' demographical, social and educational life during this pandemic.

Fig.No.2: Regression Model Analysis

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.666e+00  3.901e+00  1.196  0.23168
AgeC1        -7.998e-04  1.306e-02  -0.061  0.08116 .
Home Confinement  8.808e-01  4.973e-01  1.771  0.07654 .
Educational standard  1.193e-01  5.798e-02  2.059  0.03954 *
Worried about Studies -4.334e+00  1.730e+00  -2.505  0.01226 *
Depressive Symptoms  -3.829e+00  1.572e+00  -2.435  0.01488 *
Feared about Pandemic -4.909e+00  2.250e+00  -2.182  0.02911 *
GenderMale      -5.824e+00  5.349e+02  -0.011  0.99131
Self Mgmt Stress -2.294e+00  7.902e-01  -2.903  0.00369 **
Financial Stress  7.374e+00  2.697e+00  2.734  0.00625 **
Living Area      5.551e+00  5.349e+02  0.010  0.99172
Hurdles online class  4.126e+00  1.529e+00  2.698  0.00698 **
---

Signif. Codes:    0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 207.61 on 153 degrees of freedom
Residual deviance: 125.54 on 141 degrees of freedom
AIC: 151.54

Number of Fisher Scoring iterations: 15

```

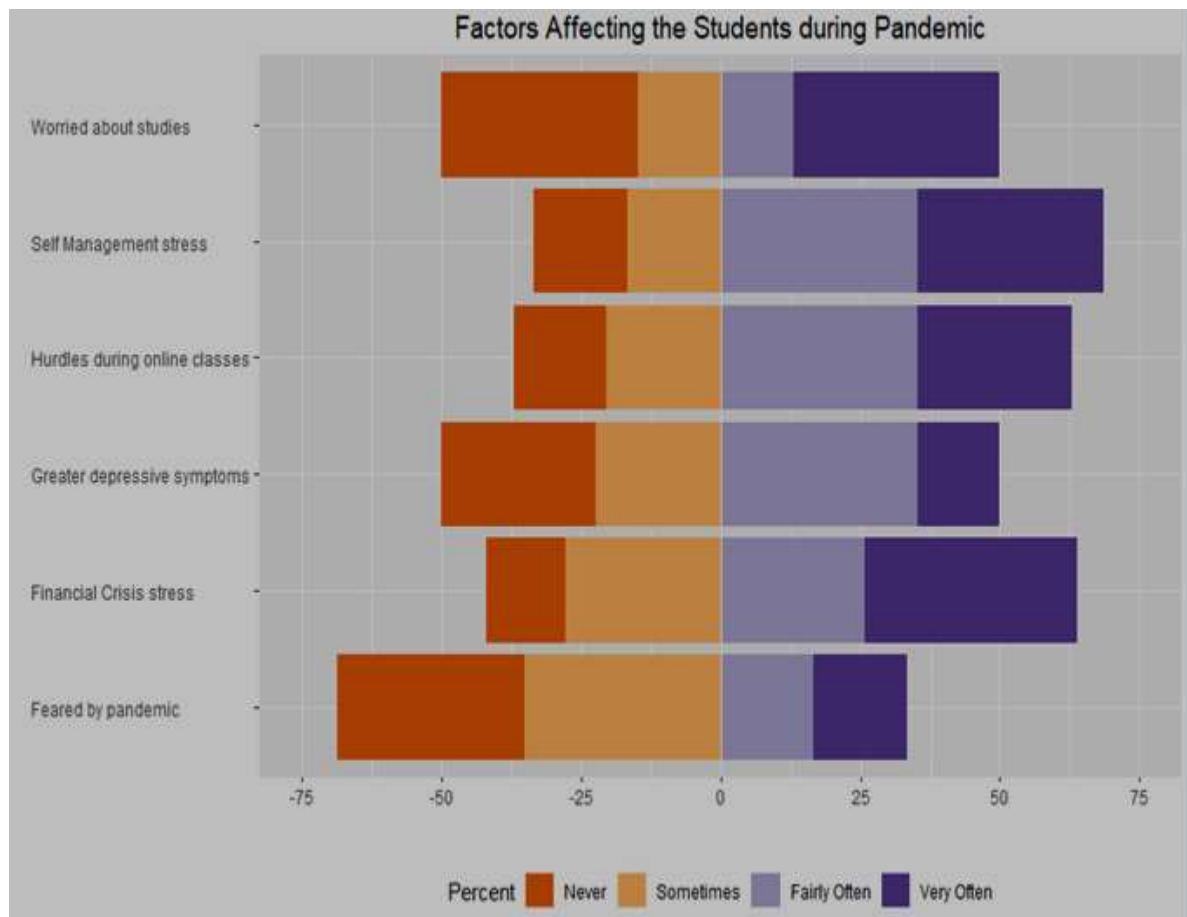
According to the Fig:2 the co-efficient 'Age' has more significance for the group C1 ,students of ages 15 to 18 than the other two age groups. The highest three levels of significance has been obtained for the coefficients which is indicated at the right end of the coefficient observation and it is mentioned as double asterisks (**) , single asterisk (*) , and a dot (.) based on their levels. The lower the p value, higher the significance of the parameter. The Fisher scoring iterations tells that the estimation of the model's fit is improved by different iterative estimations. At last the scoring took 15 iterations and stops with that when no improvement is done to the model additionally.

Table No. 2: Predictor p value for significant parameters

Parameters	Self-Management Stress	Financial Stress	Hurdles in online classes
p value	0.00369	0.00625	0.00698

b. Likert Plot Analyses

Fig.No.3: Likert Plot



In analysing the responses to the assertions made on a likert plot ,it shows that the bars from 0 to right for positive values shows the general agreement and whereas the bars from 0 to left for negative values shows the general disagreement.

According to the likert plot, in majority the bars are extending from 0 to right and it shows that the percentage of patients are agreeing with ‘fairly often’ and ‘very often’ for the factors affecting the students during this pandemic.

Conclusion

The study reveals the prevalence of stress factors that influence the student’s mental health during this pandemic. Other than the fear for pandemic the students find very difficult to manage themselves and suffer a lot in the financial crisis of their families in this pandemic situation. Due to the closing of educational institutions for long term and forced to learn virtually in online classes, they are stressed a lot to accept this pattern of change in the academic activities.

The work done shows with the evidence that among the students of age grouped from 15 to 24, most probably boy students have the more stress prevalence than the girl students .This study suggests that parents should monitor their children continually as they are stressed by the negative impacts caused due to the COVID-19 outbreak and the Government to take proper action on financial crisis to avoid dropouts from educational institutions.

References

- [1]. Zhang, Stephen X., et al. "Unprecedented disruption of lives and work: Health, distress and life satisfaction of working adults in China one month into the COVID-19 outbreak." *Psychiatry research* 288 (2020): 112958.
- [2]. Sahu, Pradeep. "Closure of universities due to coronavirus disease 2019 (COVID-19): impact on education and mental health of students and academic staff." *Cureus* 12.4 (2020).
- [3]. Liang, Leilei, et al. "The effect of COVID-19 on youth mental health." *Psychiatric quarterly* 91.3 (2020): 841-852.
- [4]. Chhetri, Bijoy, et al. "Estimating the prevalence of stress among Indian students during the COVID-19 pandemic: A cross-sectional study from India." *Journal of Taibah University Medical Sciences* 16.2 (2021): 260-267.
- [5]. AlAteeq, Deemah A., Sumayah Aljhani, and Dalal AlEesa. "Perceived stress among students in virtual classrooms during the COVID-19 outbreak in KSA." *Journal of Taibah University Medical Sciences* 15.5 (2020): 398-403.
- [6]. Fricker, Ronald D. "Sampling methods for online surveys." *The SAGE handbook of online research methods* 12.3 (2016): 184-202.
- [7]. Dutton, Lenny. "Google Forms: build free surveys." *The School Librarian* 63.2 (2015): 83.
- [8]. Mohanty, Chitta R., Rakesh Vadakkethil Radhakrishnan, and Mantu Jain. "Re-exploring the online surveys: Research tools in COVID times." *Journal of Anaesthesiology Clinical Pharmacology* 36.3 (2020): 414-414.
- [9]. Xiao, Chunfeng. "A novel approach of consultation on 2019 novel coronavirus (COVID-19)-related psychological and mental problems: structured letter therapy." *Psychiatry investigation* 17.2 (2020): 175.
- [10]. Wu, Mengting, et al. "Mental health status of students' parents during COVID-19 pandemic and its influence factors." *General Psychiatry* 33.4 (2020).
- [11]. Taylor, John M. "Psychometric analysis of the ten-item perceived stress scale." *Psychological assessment* 27.1 (2015): 90.
- [12]. Zhang, Stephen X., et al. "Unprecedented disruption of lives and work: Health, distress and life satisfaction of working adults in China one month into the COVID-19 outbreak." *Psychiatry research* 288 (2020): 112958.
- [13]. De Boni, Raquel Brandini, et al. "Depression, Anxiety, and Lifestyle Among Essential Workers: A Web Survey From Brazil and Spain During the COVID-19 Pandemic." *Journal of medical Internet research* 22.10 (2020): e22835.
- [14]. Zhang, Ying-Li, et al. "Validity and reliability of Patient Health Questionnaire-9 and Patient Health Questionnaire-2 to screen for depression among college students in China." *Asia-Pacific Psychiatry* 5.4 (2013): 268-275.
- [15]. Doshi, Dolar, et al. "Assessing coronavirus fear in Indian population using the fear of COVID-19 scale." *International Journal of Mental Health and Addiction* (2020): 1-9.