

# Prediction and Analysis of Data Mining Models for Students Underlying Issues during Novel Coronavirus (COVID-19)

Krishna

Assistant Professor Department of Computer Science,  
Acharya Bangalore B-School, Bengaluru-560091, India

Praveen Kumar V

Assistant Professor Department of Computer Science,  
Acharya Bangalore B-School, Bengaluru-560091, India

**Abstract**— Novel coronavirus widespread has impacted on lots of areas in an around the world. Among that Educational Sector comes under the picture where students are the elements which has been effected with higher rate .In this study the students problems has considered for research to find out the desire solution to reduce the depression of the students during covid-19.model has developed in this study to predict and analysis of students bottom line problems during COVID-19 finding the solution using students dataset of their various difficulties during COVID-19 of Bangalore. The random forest(RF), decision tree(DT), support vector machine(SVM), logistic regression(LR), K-nearest neighbor(KNN), and naive Bayes(NB) data set id tested by different Data Mining(DM) algorithms using weka data science tool to develop the models. The result model predicted the Major Problem faced by students during COVID-19 pandemic. Student's data set has been examined with decision tree is more accurate to identify the students problem. The accuracy of this algorithm has been noted 95.85% which can be considered the best developed model among the different models which is being developed by other algorithms including K-nearest neighbor, support vector machine, logistic regression, random forest, and. naive Bayes

**Keywords**—COVID-19; Data Model; Data mining; Data analysis; Clustering, Data Process, Data Science

## I. INTRODUCTION

COVID-19 has impacted on globally, due to this situation colleges and universities have closed in many countries this is short-term but many people are affected by this in all over the world: there is a big challenge for students and teachers in the education sector teaching and learning process. These nationwide closures are impacting over 60% of the world's student population. In this study different data models are developed with student's data of major problem faced by the students during COVID-19 survey. These models can be used in the Educational organizations and government to identify the solution and solidity of the problem. The models are evolved with the data which has been collected from survey questioner method and dataset instances of the student's problem were considered. Different data mining algorithms were examined on weka tool to design the models.

## II. LITERATURE SURVEY

2013, the author Sharma and Mamta, are implemented the DT model which compared with Naïve Bayes algorithm which has Comparison of different models to predict the result of students in the form of efficiency [12]. In 2009, Siraj,

Fadzilah, and Mansour Ali Abdoulha are used different algorithm with weka tool to better understanding, analysis and clustering of Graduates data [13]. 2009, Hoo Yann Seong are used Neural Network and DT with that they have predicted model by comparing with classifying and clustering of students' performance in their organization[14]. 2008, Espejo, and César Hervás used different algorithm to develop the data mining model and concluded with his research that DT is the best model for educational sectors in terms decision making[15]. 2007, Janecek, and Peter Haddawy DT and NB have considered consistently 3-12% more accurate when comparing with other models in predicting the performance of academic [16]

## III. METHODOLOGY

The dataset was obtained from questioner survey which has been collected from different college students in Bangalore. Which contain 734 instances with 7 attributes which are: are you Anxious about your Education during COVID-19 then what is the Scale of your Anxiety? What is your lack of concentration Scale on online sessions? (Under which Scale you find lack of concentration issue) Are you under Uncertainty or Dilemma about University and UGC Decisions? What is the Scale on Lack of Clarity and Communication about topic delivered on online sessions? Impact Scale on Lack of Physical Presence in understanding Class Are you facing Network issue during online class? Scale of Difficult to understand the online class without facial expression and body language Test mode: evaluate on training data these dataset are collected and analyzed with proper parameters for data analysis we have different types open source software and algorithms are available like KNIME, Sisense.,Rapid Miner, , Oracle Data Mining, Orange, SSDT (SQL Server Data Tools), Weka, Apache Mahout etc. these Software are useful to analyze the data more efficiently and predict the error and models using different algorithms. Among these tools weka has been selected to analysis and prediction of model based on dataset. When considering the function of data mining includes tracking patterns. One of the most basic techniques in data mining is learning to recognize patterns in data sets which contains the different faces like as shown in fig1

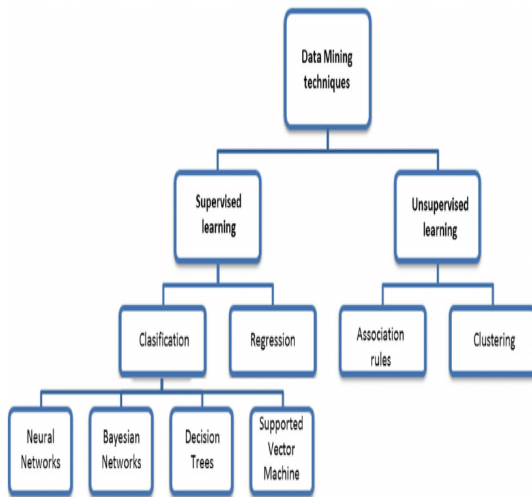


Fig 1 Method of Data Processing

In the fig a shows the process in data analysis and prediction methodology in the research using data mining to predict the model which is helpful to identify the major problems faced by the students during the covid-19 pandemic situation

#### IV. VISUVALIZATION OF OUTPUT

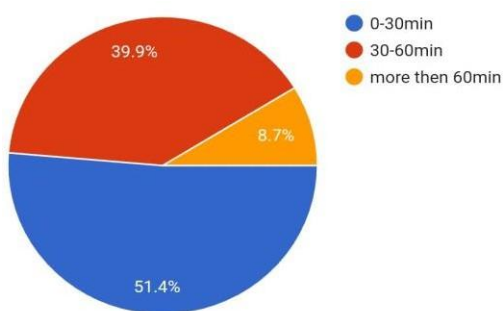


Fig 2a.Lack of Concentration

The fig 2a contains the graphical presentation of online sessions where the students comes under lack of concentration. Which contains the session time duration with three scales first scale is 0 to 30 minute, second scale 30-60 minute and last scale is more than 60 minutes among these scale 51.4% students are facing the lack of concentration. And considered as students problem in online session during COVID-19.

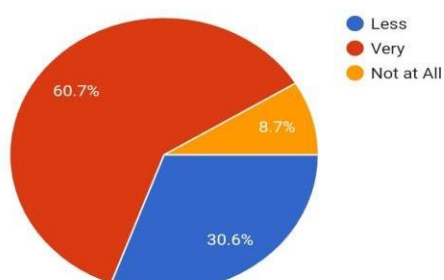


Fig2b Anxiety level

The Fig 2b represents the graph of anxiety level among the students about education in the pandemic situation.it has been noted 60.7% students are higher in scale and which is considered as the problem faces by students.

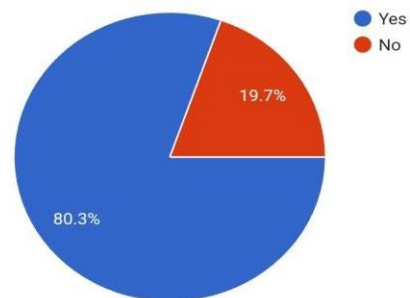


Fig2c. Uncertainty

From fig2c shows the maximum 80.3% of students are under uncertainty about university decision.

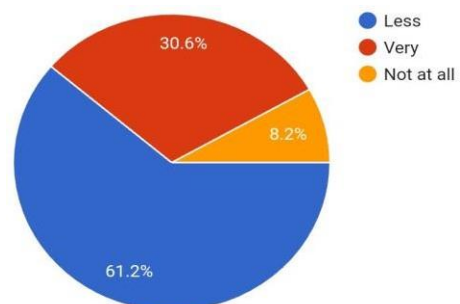


Fig2d Lack of clarity and Communication

The fig2d indicates the issues of lack of clarity and communication about topic delivery during online session 30.6% students are considered as height scale of issue but when considered with other scale it has not been considered as the major problem of students.

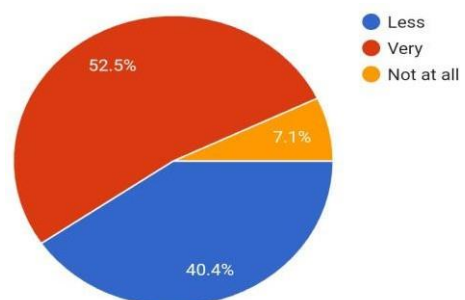


Fig2e Lack of physical presence and understanding

In the Fig2e indicates the impact scale on lack of physical presence in understanding the session and which is recorded as 52.5% students are in higher risk.

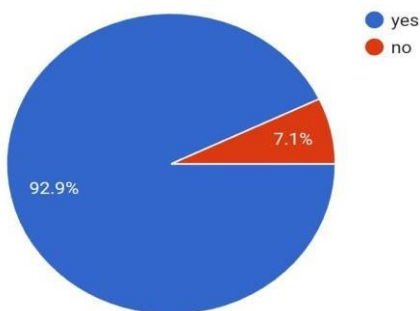


Fig 2f Network issues

In this graph Fig2b has been observed that 92.2% of students are facing the network issues during the online sessions and this scale is considered as another problem faced by the students during COVID-19

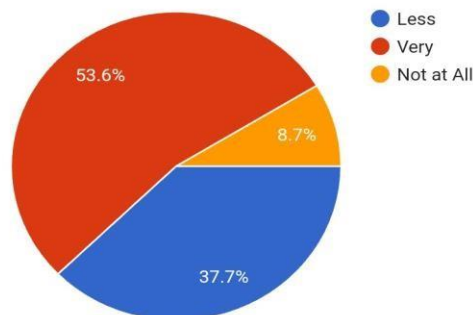


Fig 2g Facial Expression and Body language

This Fig2c is the Scale presentation difficulties to understand online session without facial expression and body language. 53.6% students are considered as major problem and 37.7 % of students as massive. by the analysis it is also be considered problem among one of them.

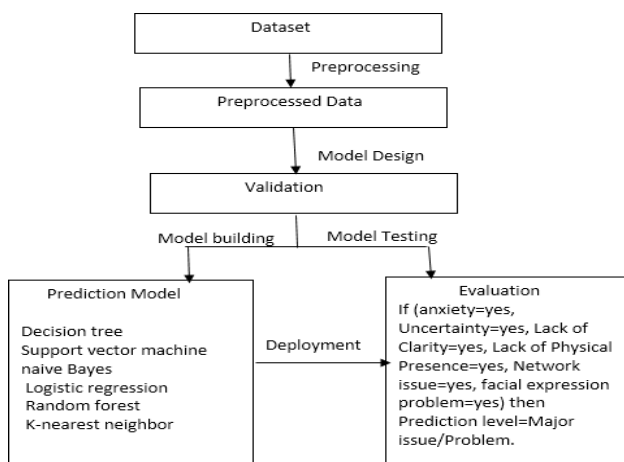


Fig2h: Data mining Method applied on data set

Data set has been applied on the data mining methodology to predict the model. Data set is applied the preprocess method to generate Preprocessed Data. This is further applied for different types of validation by the model design. After this face data set is passed to model building face by applying different data mining algorithm for examining the data set which deployment for evaluation process by applying the logic which contains if anxiety level of the student is yes, uncertainty problem facing is yes, Clarity and lack of physical presence is yes, network issue and facial expression issue is yes then the prediction level is considered as major issue or problem. The organizations of educational institutes can take decision and can find out the solution for the problems.

## V. FINDINGS AND DISCUSSIONS

The different Data Mining algorithm such as K-nearest neighbor logistic regression, random forest, decision tree, naive Bayes and support vector machine were examined on the data using WEKA tool. The model evolved with decision tree algorithm was examined with height accuracy of 94.85% and this model is considered as best model among others data models.

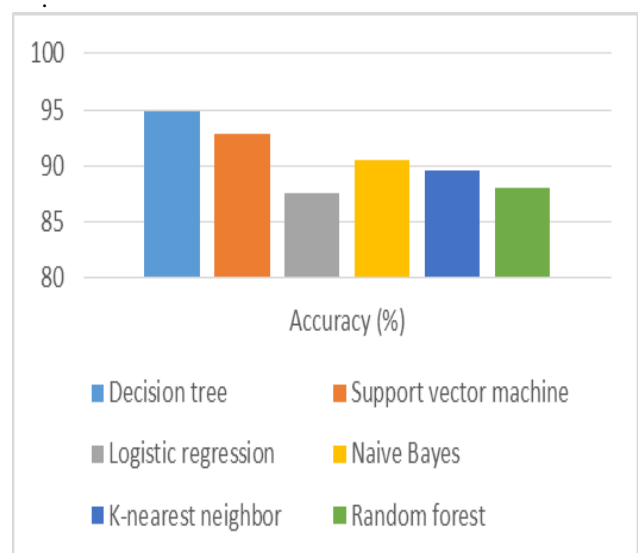


Fig. 3 Results Performance of the Different models

TABLE1.1 Performance evaluation

Sl No	Examined Data models	Accuracy in Percentage
1	Decision tree	94.85
2	Support vector machine	92.85
3	Logistic regression	87.52
4	Naive Bayes	90.49
5	K-nearest neighbor	89.60
6	Random forest	88.06

The model predicted a major problem faced by students, from the above accuracy performance table it has noted that DT has the highest accuracy Data Mining methods which is capable to predicting the major problem solution of the students with accuracy of 95.85% which has compared accuracy percentage which has been examined on student's dataset with different algorithm result. Which contains RF-88.06, LR-87.52, SVM 92.85, KNN-89.60, and NB-90.49.

TABLE 1.2 Data analysis Parameters

Correctly Classified Instances	9.48719 %
Incorrectly Classified Instances	5.1281 %
Root relative squared error	53.2095 %
Total Number of Instances	734
Relative absolute error	16.4415 %
Root mean squared error	0.2108
Mean absolute error	0.0526
Kappa statistic	0.8125

Table 1.2 contains analysis parameters which generated during classification of decision tree data mining algorithm which contains kappa statistic 0.8125 which is considered as good kappa value which is applied on 734 instances while prediction of data model

## VI. CONCLUSION

In this paper, DM algorithms were evolved and prognosis the student's problems during COVID-19 from the survey data set. LR, RF, DT, K-NN, NB, and SVM algorithms were tested on the student's dataset using weka software. The model evolved with DT was examined to be the most methodical with the significant percentage accuracy of 95.85%, with Comparatively RF-88.06, LR-87.52, SVM 92.85, KNN-89.60, and NB-90.49. The models can be used in Education sector to find the solution of student's problems during COVID-19.

## VII. REFERENCE

- [1] K. Aftarczuk, Evaluation of selected data mining algorithms implemented in Medical Decision Support Systems (2019).
- [2] S. Palaniappan, R. Awang, Intelligent Heart Disease Prediction System Using Data Mining Techniques (2019)
- [3] T.H. McCormick, C. Rudin, D. Madigan, A Hierarchical Model For Association Rule Mining Of Sequential Events: An Approach To Automated Medical Symptom Prediction. (2019)
- [4] T.H. McCormick, C. Rudin, D. Madigan, A Hierarchical Model For Association Rule Mining Of Sequential Events: An Approach To Automated Medical Symptom Prediction. (2019)
- [5] P. Sudeshna ; S. Bhanumathi ; M.R. Anish Hamlin Identifying symptoms and treatment for heart disease from biomedical literature using text data mining Electronic IEEE Xplore: (2018) ISBN: 978-1-5090-4324-8 Print on Demand (PoD) ISBN: 978-1-5090-4325-
- [6] Na Deng ; Song Lin ; Caiquan Xiong ; Desheng Li; A Clustering Algorithm of Four Character Medicine Effect Phrases in TCM Patents IEEE Xplore (2018); DOI: 10.1109/ICEIEC.2018.8473529; ISBN: 978- 1-5386-5775-1
- [7] Shaila H Koppad ; Anupamma Kumar : Application of big data analytics in healthcare system to predict COPD: IEEE Xplore (2017); ISBN: 978- 1-5090-1277-0 C. Yang, W. N. Street, Der-Fa Lu, L. Lanning, A Data Mining Approach to MPGN Type II Renal Survival Analysis (2010)
- [8] Romero, C. and Ventura, S., "Educational data mining: A Survey from 1995 to 2005". Expert Systems with Applications (33) 135-146. 2007
- [9] Romero, C., Ventura, S. and Garcia, E., "Data mining in course management systems: Moodle case study and tutorial". Computers & Education, Vol. 51, No. 1, pp. 368-384. 2018
- [10] Sheikh, L. Tanveer B. and Hamdani, S., "Interesting Measures for Mining Association Rules". IEEE-INMIC Conference December. 2014.
- [11] Waiyama, K. "Improving Quality of Graduate students by Data Mining" Department of Computer engineering. Faculty of Engineering. Kasetsart University, Bangkok, Thailand. 2013.
- [12] Sharma, Mamta, and Monali Mavani. "Accuracy Comparison of Predictive Algorithms of Data Mining: Application in Education Sector." Advances in Computing, Communication and Control. Springer Berlin Heidelberg, 2011. 189-194.
- [13] Siraj, Fadzilah, and Mansour Ali Abdoulha. "Uncovering hidden information within university's student enrollment data using data mining." Modelling & Simulation, 2009. AMS'09. Third Asia International Conference on. IEEE, 2009.
- [14] Wook, Muslihah, Yuhani Hani Yahaya, Norshahriah Wahab, Mohd Rizal Mohd Isa, Nor Fatimah Awang, and Hoo Yann Seong. "Predicting NDUM Student's Academic Performance Using Data Mining Techniques." In Computer and Electrical Engineering, 2009. ICCEE'09. Second International Conference on, vol. 2, pp. 357-361. IEEE, 2009.
- [15] Romero, Cristóbal, Sebastián Ventura, Pedro G. Espejo, and César Hervás. "Data Mining Algorithms to Classify Students." In EDM, pp. 8- 17. 2008.
- [16] Dr. Niranjana Murthy M, Amulya M P Dr. Dayananda P and Pradeep H G, "Coronavirus – COVID-19 before and after solution through web application and App", International Journal of Advanced Science and Technology, Volume-29, Issue: 5, 2020
- [17] Nghe, Nguyen Thai, Paul Janecek, and Peter Haddawy. "A comparative analysis of techniques for predicting academic performance." Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual. IEEE, 2007.