

Classification Approach for Evaluating Students Performance in Covid 19 Pandemic



Poonam Sawant, Sachin Gupta, Yogesh Sharma, Anamika Singh

Abstract: Student performance evaluation and analysis is a necessary task for improving students' quality now a days. The main aim of this research is to analyze students' performance during Covid-19 pandemic. Covid-19 pandemic impact has been extensive, affecting the education sector in India as well as world. In attempt to reduce the spread of Covid-19 government decided to temporarily close educational institutions. In response to schools and colleges closures, UNESCO recommended the use of distance learning programmes and online platforms to reach learners remotely and limit the disruption of education. This impacts not only on students' psychology, on their performance too. Although there are many systems that have been implemented predictive analytics till date, better advancements is needed. Machine learning classifiers and related technologies can be used efficiently in performance evaluation. At the end of this paper, we have proposed an Architecture of Student's Performance Evaluation System with classification techniques.

Keywords: Academic Performance, Covid-19, Machine Learning, C4.5, Naive Bayes.

I. INTRODUCTION

The impact of Covid-19 pandemic is witnessed in every sector around the world. Education sector is observed as badly affected sector in India. The closure of schools happened spontaneously in more than 100 countries worldwide. Decision of school closures left more than one billion learners out of the school. The lockdown has forced many educational institutions to abandon their offline classes and examinations and to choose the online learning. This impacts students learning ability as well as their performance [1] [2]. Student performance prediction and analysis is an essential task to plan efficient strategies to improve education system in India during such crises. Earlier many systems have been developed and implemented to predict student performance in traditional ways.

Manuscript received on March 30, 2021.

Revised Manuscript received on April 07, 2021.

Manuscript published on April 30, 2021.

* Correspondence Author

Dr. Poonam Sawant*, Asst. Professor, Sinhgad Institute of Management & Computer Application Pune, India
Email: poonam25m@gmail.com

Prof. Sachin Gupta, Asst. Professor, Sinhgad Institute of Management & Computer Application, Pune, Maharashtra, India
Email: sachingupta28aug@gmail.com

Prof. Yogesh Sharma, Asst. Professor, Sinhgad Institute of Management & Computer Application, Pune, Maharashtra, India
Email: yogeshs.scs@gmail.com

Dr. Anamika Singh, Director, Sinhgad Institute of Management & Computer Application, Pune, Maharashtra, India
Email: director_mba_simca@sinhgad.edu

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

But due to the need of faster and efficient systems advance development is necessary. Machine learning technologies such as decision tree, association rule, nearest neighbors, neural networks, genetic algorithms, exploratory factor analysis and stepwise regression can analyze relevant information results and produce different perspectives to understand more about the students' activities. They can be applied to the higher education process, which in turn helps to improve student's performance [5] [6]. From the literature study we found machine learning techniques implementation with python programming language. Various machine learning algorithms have been developed and implemented over a period of time with enhancement in performance and ability to predict performance. The well-known algorithms are C4.5 and Naive Bayes classifiers. Performance prediction is incomplete without visualization of the data. In addition to classification of data using decision trees and Naive Bayes algorithms it is also important to visualize the data so that educational sector get a visual aspect of the data in order to understand the patterns and can easily make decisions.

II. RELATED WORK

In [5] researchers applied the C4.5, ID3 and CART decision tree algorithms on engineering student's data to predict their performance in the final exam. The outcome of the decision tree predicted the number of pass students, fail students or who are likely to promote to next class and provided steps to improve the performance of the students who were predicted poor. In [7] Authors highlights important issues of higher education system such as predicting student's academic performance. For making analysis on the student data they have selected Decision Tree, Naive Bayes, Random Forest, PART and Bayes Network algorithms with the most important techniques like 10-fold, cross-validation, percentage split and training set. After performing analysis on different metrics using different data mining algorithms, they found Random Forest algorithm a best algorithm to analyze performance and to make a guideline for future improvement in student performance in education. In [8] data mining techniques are used to investigate the effect of socio-economic or family background on the performance of students using the data from one of the Nigerian tertiary institutions as case study. The analysis was done using Decision Tree algorithms. The data consists of 240 records of students. The students' first year cumulative grade point average was used for measuring the performance. Many Decision Tree algorithms were implemented to find the best models the data was used

to generate rule sets which can be used to analyze the effect of the socio-economic background of students on their academic performance. In [11] the accuracy of different decision tree algorithms is calculated.

The classification task is used to evaluate student’s performance with the help of their Attendance, Class test, lab work and Assignment marks. Classification model is developed using BFTree, C4.5 & CART which gives accuracy 67.07%, 66.83%, and 65.77% respectively.

III. METHODOLOGY

In Covid 19 pandemic duration offline classes switched to online mode along with student’s assessment too. Online teaching-learning impacts student’s performance positively and negatively [1] [2]. Therefore, in this section we have proposed “**Students Performance Evaluation System**” which is depicted in fig 1 and the flow of the system is as follows:

1. Student Data Collection
2. Data Preprocessing (Data Selection, Data Cleaning and Data Transformation) using Python.
3. Dividing the data into samples for training and testing purpose.
4. Training the model using training data set
5. Rules Generation and Pattern Discovery
6. Model Testing
7. Predictions based on rules and patterns
8. Statistical Data Analysis and Interpretation
9. Visualization using Matplotlib and Seaborn
10. Decision Making

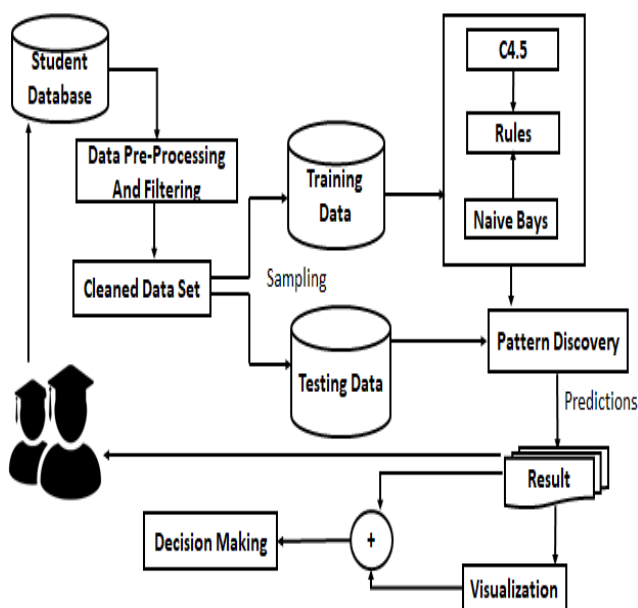


Fig 1: Architecture of Students Performance Evaluation System

Student’s performance is evaluated by internal and external assessment in higher education. Internal assessment of the students is determined using various parameters suggested by University and Institute. It is carried out by respective subject teacher on the basis of the performance given by the student in class test, surprised class test, presentations, viva, case studies, mid-term and end-term examinations, objective tests, practical knowledge and many more. University conducts external examination for external

evaluation. In every semester students should get minimum passing marks in both internal and external evaluation. Semester wise data is collected from our own institute for MCA course for the Academic period of 2018-19, 2019-20 and 2020-21. Dataset contains following fields.

Table-1: Dataset

Fields	Values
Branch	{MCA}
Semester	{I,II,III,IV,V,VI}
Grades	Pass, Fail, A.T.K.T. Distinction
Attendance	{Below 70%, As or Above 70%}
Assignments	{Submitted, Not Submitted}
Class Test	{Poor, Average,Good}
Mid Term	{Poor, Average,Good}
End Term	{Poor, Average,Good}
Presentation	{Yes, No}
Case Studies	{Yes, No}
Lab Work	{Yes, No}

Data is cleaned and divided into training and testing set for building the model using decision tree and Naive Bayes classifiers.

IV. RELATED TECHNOLOGIES

1) C4.5:

The decision tree model is a technique specifically proposed for a class label ID3 and C4.5 algorithm which is referred as statistical classifier. C4.5 algorithm is developed by Ross Quinlan to generate a decision tree. It is an extension of Quinlan’s ID3 algorithm. This decision trees is used for classification. Information gain is used as a splitting criterion which accepts data with categorical or numerical values. In this case, the records will be classified into two classes yes or no. The $p(S, j)$ is denoted as the proportion of instances in S assigned to j^{th} class. Therefore, the entropy of attribute S can be calculated as:

$$\text{Entropy}(S) = -\sum_{j=1}^c p(S,j) * \log p(S,j)$$

Entropy is calculated for every attribute. The information gain for a training dataset T is defined as:

$\text{Gain}(S, T) = \text{Entropy}(S) - \sum_{v \in \text{values}(T_s)} |T(s,y)/T(s)| * \log p(S,j)$ where, (T_s) is denoted as the set of values of S in T , T_s is denoted as the subset of T and v is denoted as the subset of T in where the attribute S has a value of v .

2) Naive Bays:

Naive Bayes is very simple algorithm of probability based on the Bayes rules. It aims to build a model in the form of probability. It is an effective and easy processing method. The probability of classification data is denoted by $P(a_i|v_j)$, where a_i refers to the attribute i and v_j refers to the class label j . Therefore, for this probability classification has been calculated. For each class the highest probability of a_i is depended on v_j . The range of probability is always in between 0 and 1 [9][10]

$$V_{NB} = \arg.\max P(v_j) * \prod_{i=1}^n P(a_i|v_j)$$

3) Python:

For implementing proposed system Python programming is used. Python is an interpreted object oriented programming language with rich libraries to implement machine learning algorithms.



For plotting graphs and giving visualization effect like Matplotlib and Seaborn are used.

V. EXPERIMENTAL RESULTS

The main aim of the research is to evaluate the performance of the students in the semesters during covid-19 pandemic and compare it with previous semester performance. Research aims to check the accuracy and performance of machine learning algorithms too.

A. Model Performance

Models are constructed by researchers.

Two classifier models are used C4.5 and Naive Bayes under supervised learning based on their popularity. To accomplish our objective, we worked on each machine learning algorithm independently. For each one we have applied the set of testing and training data sets and optimized prediction accuracy for each set using a test driven methodology and when the algorithm works correctly we have evaluated the expected results. In case of machine learning the most important thing happen is that the algorithms can memorize the data and when we want to use it with a new dataset they might give a poor performance. This behavior is called as over fit. To avoid this problem, we have used the test driven methodology. Each data set is divided into two parts as follows [9][10].

- 1. Train Data Set:** It is used to train the machine learning algorithm for the learning process.
- 2. Test Data Set:** It is used to check whether algorithm is giving good performance or not.

All the classification algorithms follow the same principals for evaluation of models.

$$\begin{aligned} \text{Accuracy} &= (TN+TP)/(TN+FP+FN+TP) \\ \text{Precision} &= TP/(FP+TP) \\ \text{Sensitivity} &= TP/(TP+FN) \\ \text{Specificity} &= TN/(TN+FP) \end{aligned}$$

Table 2: Model performance and confusion matrix

		Predicted Class	
		No	Yes
Observed Class	No	TN	FP
	Yes	FN	TP

In supervised machine learning, there exists a true output and one predicted output generated by model for each data point. Therefore, the result for each data point can be assigned to following categories.

1. True Positive (TP): True output is positive and prediction is also positive.
2. True Negative (TN): True output is negative and prediction is also negative.
3. False Positive (FP): True output is negative and prediction is positive.
4. False Negative (FN): True output is positive and prediction is negative.

Based on the correctly and incorrectly instance classification the accuracy is measured as shown in the table 3. Set-1 contains data related to internal evaluation and Set-2 have data related to external evaluation.

Table 3: Algorithm Performance
(Compiled By Researchers)

Accuracy		
Series	Decision Tree	Naive Bayes
Set-1	93.35	97.40
Set-2	94.64	97.45
Average	93.99	97.42

Set-1	93.35	97.40
Set-2	94.64	97.45
Average	93.99	97.42

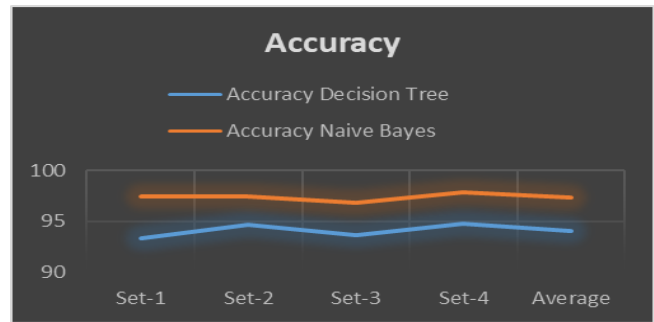


Fig 2: Algorithm Prediction

From the above table 3 and graph shown in figure 2 it has been observed that, the average prediction accuracy of decision tree algorithm is less than the prediction accuracy of Naive Bayes. Naive Bayes gives better performance than c4.5.

B. Students Performance

Student performance is measured based on their internal evaluation and external evaluation as depicted in Fig 3 and Fig 4.

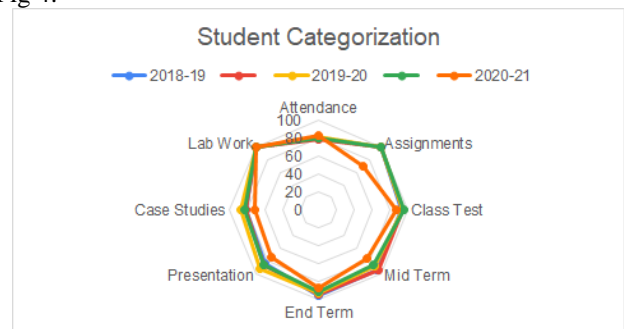


Fig 3: Student Categorization Based on Internal Evaluation

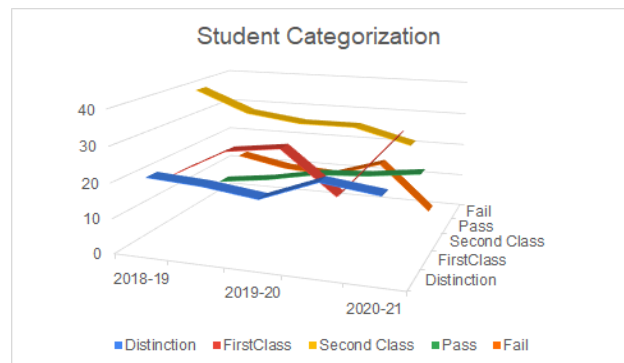


Fig 4: Student Categorization Based on External Evaluation

Student performance is evaluated by using the proposed architecture. Performance is improving in every semester. But during Covid -19 pandemic student gives poor performance in internal evaluations which is 50% of total evaluation. As compared to off line theory examinations student performed well in online MCQ examination,

perhaps they unable to express themselves in other internal evaluation parameters due to lack of knowledge, technical difficulties, miss interpretations of concepts, psychological effect of Covid-19 and many more.

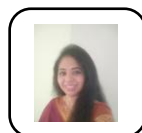
VI. CONCLUSION

This paper defines the proposed Architecture of Student's Performance Evaluation System by implementation of C4.5 and Naive Bayes using python along with the visualization. With the rapid growth in development of various technologies like Artificial Intelligence, Machine Learning, Deep Learning and Big Data Analytics, traditional technologies cannot fit any more and hence we will use advance technologies like python for implementing C4.5 decision tree and Naive bayes probabilistic classifier algorithms. Visualization is done using Matplotlib and Seaborn libraries. Student performance is measured based on various parameters during internal and external examinations. Based on the results it has been also proved that student's performance is decreased during Covid-19 pandemic. Naive Bayes gives accurate results than C4.5 algorithm. In future works, proposed architecture system will be incorporated with advanced algorithms like Forest Tree, Artificial Neural Network to increase the performance of the system.

REFERENCES

1. Impact of Impact of Pandemic COVID-19 on Education in India Pravat Kumar Jena, *International Journal of Current Research (IJCR)* , Vol-12, Issue-7, Page-12582-12586 (2020) DOI-<http://journalcra.com/article/impact-pandemic-covid-19-education-india>
2. Impact of Coronavirus Pandemic on Education Edeh Michael Onyema1* Dr. Nwafor Chika Eucheria2 Dr. Faith Ayobamidele Obafemi 3 Shuvro Sen4 Fyneyface Grace Atonye5 Dr. Aabha Sharma6 Alhuseen Omar Alsayed7
3. Using Machine Learning Techniques for Sentiment Analysis, Final Project, school of engineering, Oscar Llombart (2017),
4. Classification Model of Prediction for Placement of Students, Ajay Kumar Pal ,S.Pal I.J.Modern Education and Computer Science, 2013, 11, 49-56 Published Online November 2013 in MECS (<http://www.mecs-press.org/>) DOI: 10.5815/ijmecs.2013.11.07
5. Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification, Surjeet Kumar Yadav,S.Pal World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 2, 51-56, 2012
6. Data Mining Techniques in EDM for Predicting the Performance of Students, Ajay Kumar Pal ,S.Pal, International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 02– Issue 06, November 2013
7. Evaluation of Data Mining Techniques for Predicting Student's Performance, Mukesh Kumar, Prof. A.J. Singh, I.J. Modern Education and Computer Science, 2017, 8, 25-31 Published Online August 2017 in MECS (<http://www.mecs-press.org/>) DOI: 10.5815/ijmecs.2017.08.04
8. Investigating the Effect of Students Socio-Economic/Family Background on Students Academic Performance in Tertiary Institutions using Decision Tree Algorithm * Adeyemo, A.B. & Kuyoro, S. O. www.actasatech.com
9. <https://ibm.ent.box.com/s/atp4ezwvo5jr27zpxlu4987ercep2am>
10. <https://spark.apache.org>
11. STUDENT'S PERFORMANCE ANALYSIS USING DECISION TREE ALGORITHMS Abdulsalam Sulaiman Olaniyi , Saheed Yakub Kayode , Hambali Moshood Abiola , Salau-Ibrahim Taofeekat Tosin , Akinbowale Nathaniel Babatunde , Anale. Seria Informatică. Vol. XV fasc. 1 – 2017.

AUTHORS PROFILE



Dr. Poonam Sawant, is an Assistant Professor of Computer Application in Sinhgad Institute of Management and Computer Application, Pune, India. She has pursued her B.Sc. degree from Mumbai University, MCA from Shivaji University, M.Phil in Computers from Shivaji University and

Ph.D in Computers from Savitribai Phule Pune University. Her research interests are AI & Machine Learning, NLP, Big Data Analytics and Web Mining. Her Papers are published in peer review journals including International Journal of Innovative Research in Computer and Communication Engineering, International Journals of Multidisciplinary Research Academy (IJMRA), International Journal of Emerging Technology and Advanced Engineering, IPASJ International Journal of Computer Science (IJCS) , IJEAT and many more.



Prof. Sachin Suresh Gupta, is an Assistant Professor of Computer Application in Sinhgad Institute of Management and Computer Application, Pune, India. He has Pursued his B.Sc.(Electronics), PGDCS and MCA degree from Amravati University, Amravati, India. His Research Interest Areas are Python, Big

data, Mobile Development, Data Analytics and Programming Technology. His Papers are published in Peer review journals including International Journal of Innovative Research in Computer application and Management, International Journal of Information Systems, International Conference of Commerce and Management.



Prof. Yogesh Sharma, has completed Bachelor of Computer Application, (BCA) from National College Vidisha affiliated with Barkatullah University, Bhopal, Madhya Pradesh in year 2004. Further he has proceeded to pursue Master of Computer Application (MCA) from School of Computer Science

Departmental Institute of Devi Ahilya Vishwavidyalaya (DAVV), Indore, Madhya Pradesh in 2007. He has been associated with Sinhgad Technical Education Society's Sinhgad Institute of Management and Computer Application (SIMCA) as Assistant Professor for MCA course since October 2012. He has having 11+ years of teaching experience with a proven track record of guiding and motivating students. Currently he has engaged into teaching subjects like JAVA Programming, Mobile Application Development using "Android", Advance JAVA, C programming with DATA Structure, Web Technology, and Advanced Internet Technology. Till date he has 05 papers in National/International Conference / Journals. He has attended various FDP/Seminars/Workshops.



Dr. Anamika Singh, Director Sinhgad Institutes of Management & Computer Application(SIMCA), an alumna of IIMA, Post Doctoral Fellowship in Management from Lincoln University, Malaysia and a PhD guide at SPPU, is associated with the esteemed Institute 'Sinhgad Technical Education Society' since

2010. She has over 16 years of academic experience. Dr. Singh is a true academician and a passionate teacher at her roots and believes that acumen comes with the power of knowledge. She has very vast teaching portfolio including subjects such as Basics of marketing, Marketing Management, Consumer Behaviour, Integrated Marketing Communication, Soft Skills etc. On the leading chair she acts like a great motivator, a dynamo and woman of action. She is a science graduate, an M.B.A. in marketing and PhD in the field of Marketing and Advertising. She has numerous national and International papers to her credit. She is developing few case studies in Strategic Management.