# PREDICTING THE STUDENTS' PERFORMANCE IN COVID PANDEMIC SITUATION USING DATAMINING

*R. Gomathi*, S. Saranya, M. Suganthapriya*

**Abstract**

The performance of students' has to be predicted and how they have affected in education on COVID pandemic situation. A big challenge arises while students and faculty members are having classes and interactions through online mode. Effectiveness of education in online becomes very low and students' are not gaining knowledge in both academic and Socio-economical ways. Here, the implementation of a systematically review for analyzing both students' and faculty's performance by Data mining techniques is proposed to encourage students' achievements. The Support Vector Machine algorithm is appropriate for predicting students' performance. Support Vector Machine algorithm produces high prediction among other prediction algorithms.

**Keywords:** prediction, students' performance, support vector algorithm, data mining, educational data mining;

## I INTRODUCTION

Education Mining is used to predict the data from knowledge discovery of students' performance using various data mining techniques will be implemented by machine learning and statistics (For example., Colleges/Universities and Online Tutoring System).

Prediction of students' performance is based on knowledge discovery by classification of students' and predicting the student data by the association rules. It helps to have better decision on education. Prediction is performed by survey of students' education level. Classification is used to classify the students' based on performance in End Semester Examinations.

Online education modes (Google Meet, Google Classroom, Moodle, Edumodo, etc.,) maintain large data sets for monitoring students' performance. Analyzing the student's performance is based on End Semester Marks and to make precautions for the educational improvements are important.

The effectiveness of the online class and examinations should be discovered and to classify the students are discussed in the paper. The following Data Mining algorithms are Decision tree, Random Forest and Naive Bayes used to check the accuracy rate for classification. Data pre-processing is used to improve the accuracy rate of the prediction model.

The remaining of the paper is framed as 4 parts: Part [2] is the study of related works presented. Part [3] contains the data mining techniques and algorithms implemented on this study to visualize the representation of collected data sets and to explore the implementation and results of end semester examinations. In Part [4], future works are included. Finally, Part [5] produces the results (outcome) of this study.

## II Related Work

In 2020, pandemic situation changes both schools and college level educations to online mode and raised inequalities in education [1][[2]. Due to COVID, education faced online learning and has raised more expectations on quality of education. Prediction is to improve online experiences in teaching and teachers have less apprehensive in engaging with students.

Department of Computer Applications,
KGiSL Institute of Information Management, Coimbatore, Tamil Nadu
*Corresponding Author

The advancements in technology increase the technological investments in the field of education [1][2]. Evolution on web based and multimedia technologies has occurred and cost of online learning decreased. Time and space for online learning technologies have eliminated.

Prediction of students' performance can be done by knowledge discovery of drop outs from End semester examinations [2][3]. Cognitive skills of students' should be monitored. To analyze advanced learners cum slow learners becomes important.

## III RESEARCH METHODOLOGY

To achieve the aim of this study is to attain the pre-processing and clustering for predicting students' performance through grades and visualized in dissertation [2][3][4]. Normalization is the best method for pre-processing methods to involve datasets. Those datasets have been focused on clustering or prediction algorithms.

| Students Attributes | Data Type | Attributes' Details |
|---|---|---|
| ID | Ordinal | 1,2,3 4, …., 57(total were 55 students) |
| Age | Ratio | Between 20 and 30 |
| BCA | Nominal | Computer Applications,... |
| B.Sc. grade | Ratio | 3.25, 3.62, … |
| Course names | Nominal | Cloud Computing, Python Programming,... |
| Course Grades | Ordinal | A,B,C & D |
| Mentor Names | Nominal | Mentor 1, Mentor 2, … |

*Table – 1 Main Dataset attributes*

The above dataset should be summarized by the involvement of normalization [4][5][6]. That clustering and prediction algorithms recommend that making the instances values within specific ranges, either [0, 1] or [−□1, 1], since scaling to these ranges tend to give better results. In this project, the MinMaxScaler is used to calculate the instances with range [0, 1] and assume the range is [a, b]:

That clustering and prediction algorithms recommend that making the instances values within specific ranges, either [0, 1] or [−□1, 1], since scaling to these ranges tend to give better results. In this project, the MinMaxScaler is used to calculate the instances with range [0, 1] and assume the range is [a, b ]:

$$Y-Normalized<-([b-a]*(X-min(X))/(max(X)-min(X)))+aY-Normalized<-([b-a]*(X-min(X))/(max(X)-min(X)))+a$$

| Descriptive Statistics | Dataset1 | Dataset2 |
|---|---|---|
| Number of instances | 273 | 38 |
| Dependent Variables (DV) | Grades (All Courses Grades) | Grade (Dissertation Grade) |
| DV Mean | 3.39 | 3.44 |
| DV Median | 4 | 4 |
| DV Mode | 4 | 4 |
| Accuracy Baseline | P (4)□=□(136/234) * 100□= **58.1 %** | P (4)□=□(23/38) * 100□= **60.5 %** |

*Table – 2 Summary statistics for Dataset1 and 2*

**Data Mining Algorithms for Predicting Accuracy Eate: Decision Tree**

Decision tree is used for classification and Regression problems, and it acted as a supervised learning technique [7][8][9]. Classification problems are focused mainly in decision tree and it has Decision Node and Leaf Node.

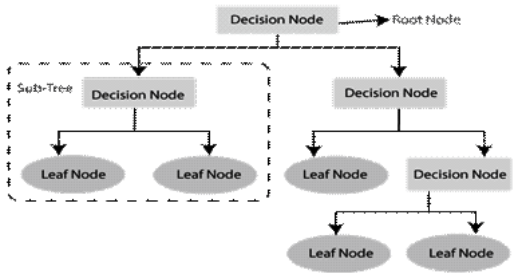


*Figure – 1 Decision Tree in Classification.*

Pruning, Dependent variables, Input variables and split at each node are the methods used to classify the problems with a complete decision tree.

**Random Forest**

In a random forest tree, bagging classifier or hyper parameters are used in decision tree [7][8][9]. Random forest adds additional random search in decision tree while growing itself. Random forest solves regression and classification problems.

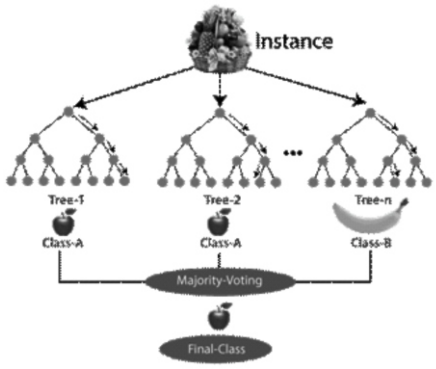


*Figure – 2 Random Forest Classifier*

**Naïve Bayes**

Naïve Bayes is a classification technique uses Bayes theorem for making assumptions on independence in predictions [7][8][9]. It focuses on unrelated features of present class is present in any other class. Filtering spam, classifying the documents and sentiment prediction etc., are done by Naïve Bayes algorithm.

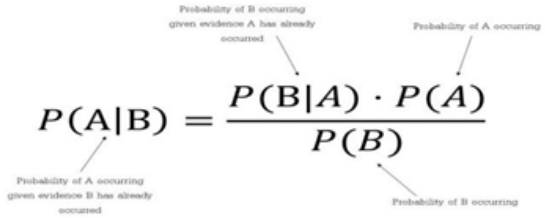


*Figure – 3 Naïve Bayes Classifier.*

**Accuracy and Experimental results based on Naïve Bayes Classifier**

| Classification Algorithms | Accuracy | F1-Score |
|---|---|---|
| **Logistic Regression** | 84.50% | 0.6237 |
| Naïve Bayes | 81.11% | 0.6105 |
| Stochastic Gradient Descent | 82.23% | 0.5880 |
| K-Nearest Neighbours | 83.66% | 0.6924 |

*Table – 3 Accuracy and Experimental results*

**Experimental results of Algorithms in Classification**

| Algorithm | Grading for First-level | Grading in Binary(P/F) |
|---|---|---|
| Decision tree classifier (J48) | 68.82% | 92.38% |
| Random forest classifier | **74.52%** | **94.17%** |
| Naive Bayes classifier | 67.36% | 87.43% |

*Table – 4 Overall Experimental Results.*

## IV CONCLUSION AND FUTURE WORK

In this research work, predicting the students' performance based on Decision tree, Random Forest and Naive Bayes with their final marks and grades is done. Pre-processing in data sets produced accuracy rate for classification. Overall, better accuracy rate have been achieved.

In future, various methods and algorithms may be implemented for giving better accuracy rate. More data sets can be involved for prediction.

## REFERENCES

[1] N D Lynn, and A W R Emanuel, Using Data Mining Techniques to Predict Students' Performance. a Review Published under licence by IOP Publishing Ltd IOP Conf. Ser.: Mater. Sci. Eng. 1096 012083 https://www.intechopen.com/chapters/71573

[2] Michael Nachbar, Executive Director, Eight Predictions for Education in 2021 – Article on Jan 08, 2021.

[3] Lubna Mahmoud Abu Zohair, Prediction of Student's performance by modelling small dataset size, International Journal of Educational Technology in Higher Education volume 16, Article number: 27 (2019),https://rdcu.be/czjXx.

[4] Ferda Ünal, Data Mining for Student Performance Prediction, Data Mining - Methods, Applications and Systems. DOI: 10.5772/intechopen.91449

[5] Educational Data Mining & Students' Performance Prediction , Amjad Abu Saa International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5, 2016

[6] Yadav, S.K. and Pal, S., 2012. Data mining: A prediction for performance improvement of engineering students using classification. World of Computer Science and Information Technology Journal (WCSIT). (ISSN: 2221-0741), Vol. 2, No. 2, 51-56, 2012

[7] Annisa Uswatun, Khasanah and Harwati, A Comparative Study to Predict Students' Performance Using Educational Data Mining Techniques, Published under license by IOP Publishing Ltd.,

[8] Pooja Thakar , Anil Mehta, Manisha, Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue International Journal of Computer Applications (0975 – 8887) Volume 110 – No. 15, January 2015

[9] Nilesh V. Ingale, et.al , Survey on Prediction System for Student Academic Performance using Educational Data Mining. https://www.turcomat.org/ index.php/ turkbilmat/issue/view/48