INTRODUCING THE STUDENT PERFORMANCE ASSESSMENT AND PREDICTION SYSTEM WITH CONSIDERATION OF PARENTS EDUCATION

[1]**MS Sarika Rathi**

[1]Asst. Professor, [1]Computer Science & Engineering Department,
[1]School Of Engineering & Technology MGM University, Aurangabad , India

Abstract: Today's generation is rapidly evolving world, education plays a pivotal role in shaping the future of individuals and societies. With the advent of advanced technologies, there is an increasing need to leverage these innovations to enhance the learning experience and ensure every student reaches their full potential. Machine learning, a subset of artificial intelligence, has emerged as a powerful tool that can revolutionize the education sector by providing valuable insights into student performance.This paper aims to introduce the Student Performance Assessment and Prediction System, a cutting-edge application of machine learning in education. By harnessing the power of data analysis and predictive modeling, this system aims to facilitate a comprehensive evaluation of student performance, thereby enabling educators to make informed decisions and tailor teaching strategies to individual needs.

The Student Performance Assessment and Prediction System capitalize on the vast amounts of data generated within educational institutions. Traditional assessment methods, such as exams and assignments, often fall short in providing a holistic view of student capabilities. However, by leveraging machine learning algorithms, this system can analyze a wide range of data points, including academic records, attendance, engagement levels, and even socio-economic factors, to derive actionable insights about each student's performance and potential.With this system, educators gain the ability to predict and identify potential areas of concern for students at an early stage. By analyzing historical data patterns and performance indicators, machine learning algorithms can generate accurate predictions about student achievement. This empowers teachers and administrators to implement targeted interventions, such as personalized tutoring, tailored educational resources, or additional support, to maximize student success.

Index Terms - Component, formatting, style, styling, insert.

I. INTRODUCTION

Students are very important part of an educational instituteand also for the country. In a crowded class a teacher can't monitor every students. So it becomes very difficult for a teacher to give attention to every student in the class equally. A classroom is filled with a lot of introvert and extrovert students. When started giving tuition to students , feel the importance of monitoring every student. A teacher shouldknow before a student falls behind. Student's academic data is the most important thing for this research, because it indicates most of the things about a student. Like how much he/she studies, what type of subjects he likes and subjects he doesn't like. An IQ test and a physiological test can also help a lot for this research. It could be know how much time he spends for studies and how much he spends for a hobby then we could understand what type of motivation he needs from his teacher.The main goal of this research is about knowing if the student is learning. A subject's result is depends on many things. Usually every subject have class test marks, attendancemark, assignment marks, presentation Marks, mid-termexamination marks, final examination marks. The summation of every test is equal to the result of the subject. It also dependson some other things like the physiological data of the student,IQ score, how much time he or she spend on studying etc.If a student is regular in the class and can carry a good remark in class test and mid-term, then he/she can perform well in the final examination. But what if when a student carrygood mark in mid-term and has poor performance in attendance or class test! Or what would happen in a case whenhe/or she has an excellent presentation skill but cannot performin the main examination! It is said that 'No one is perfect in this earth' and it is also applicable for a student. But we as parent believethat we can boost our perfectness to a maximum level according to our personal capacity. So primarily we would predict the performance of the final examination in this research according to students past event's report. In MachineLearning, K-Nearest Neighbors, SVC, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, Linear Discriminant Analysis algorithm can beapplied to predict the future result from some existing attributes of students. [1] Learning Analytics plays major role in improving educational system by focusing the different perspective e.g student perspective, teacher perspective and administrative perspective. Student's proper assessment, a clear understanding of educational problems, selecting, and planning proper interventions at the right time are few goals of learning analytics. The major goal is to achieve a

technological shift from traditional learning and teaching practices towards automation. Educational data mining is an emerging research area composed of a big set of psychological and computational approaches for providing a roadmap of how students learn. The Latest automated interactive learning tools like creative games, simulation-based applications, and intelligenttutoring systems have given ways to analyze and discover student's data and patterns those data contains. [2]

II.   LITERATURE SURVEY

The Student Performance Assessment and Prediction System can aid in identifying patterns and trends at a macro level. Educational institutions can leverage this information to enhance curriculum development, identify gaps in teaching methodologies, and optimize resource allocation. By understanding the factors that contribute to student success, schools and policymakers can implement data-driven strategies to improve overall educational outcomes.

However, it is crucial to recognize the ethical considerations associated with the implementation of such a system. Safeguarding student privacy and ensuring the responsible use of data must be prioritized. Clear guidelines and regulations should be established to protect sensitive information and ensure that the system operates within ethical boundaries.

As the usage of computers and internet is everywhere, theavailability of data that can be analysed rapidly increased. Data can be anything related to population, academic data ofstudents, and interests of people. I can see that new data emerging from time to time. Analysing the data is the difficult task for humans. So here comes the computer whichcan analyse the data more efficiently than humans because itstores the data digitally in a well-formatted way. This is where the machine learning emerged. Machine learning is the branch of Artificial Intelligence that provides ability to automatically learn from past experiences. Here themachines do get programmed explicitly. As the name suggests, it gives the ability to the computer that makes humans and machines look alike in the aspect of learning. Onthe basis of the nature of the learning signal, machine learning is classified into supervised learning andunsupervised learning. This study focuses on supervised learning, more specifically on predictive analysis. Wheneverthe predictions of future outcomes are done, predictive analysis plays an important role. The range of applications ofpredictive analysis is very vast. Predicting student's academic performance is very important because it can indicate the teachers about the students who are possibly to drop out from the course and prediction can provide additional assistance to the students who need to improve their academic performance. [3]

Machine learning lies in the computer science field butis different from basic computing algorithms which are usedfor problem solving. In the process of machine learning, the algorithms are designed in a way which allows the system orcomputer to evaluate the data inputs, create training sets andproduce the required range specified output using statistical estimation [4].

Education Data mining,which is considered to be the subset of Data mining focuses on finding meaningful insights from the data in the educationenvironment. Universities across the globe are facing challenges of students dropping out at the very early stage ofthe program. Research shows that the high drop ratio is between the first and the sophomore year. Reasons for this could be the wrong choice of the program due to the pressurefrom family or friends, financial issues; poor performance; psychological factors or even the distance the student has to travel every day to take the program. To identify these problems, an early intervention mechanism is required so that these students can be identified based on theirperformance in their early semesters and proper counseling or support can be given. [5] The main challenge of Institutions is to deeply analyze their performance in terms of student performance, teaching skills and academic activities. Class Performance and Sessional Marks are important factors for analyzing and predicting student class result. There are many Data Mining Techniques like Decision trees, neural networks, Nearest Neighbor; Naive Bayes etc are being used in Educational Data Mining. [6]The students' performance level may be affected by many factors related to father's job, gender, and their average marks in the previous years. The early prediction of student performance may help in improving the educational process[7].

Classifiers are represented by different kinds of models. And the variation ofalgorithms is great in some times for inducing classifiers from data. Some popular classification algorithms are K -Nearest Neighbors classifier, Naïve Bayes, Neural networks, Decision Tree Algorithms (ID3, C4.5, and Random Forest), and SupportVector Machine (SVM). [8]

Ibrahim and Rusli (2007) [9] have compared the performance of three different algorithms like linear regression, decision tree and artificial neural network. They have used CGPA and demographic variables for measuring the student performance. Finally, they have got the prediction accuracy over 80 and linear regression method performs outstanding for short data samples.

Pauziah Mohd Arsad, et al. [10] used the method of Artificial Neural Network for predicting student academic performance. Their work used cumulative grade points as the standard of measurement. The results of students from the first semester were taken as the initial input variable making it the independent variable and marks of the eighth semester grade points as the output the dependent variable. The correlation coefficient R and the mean square errorwere used to measure the performances of the models.

Teir and Halees (2012) [11] have applied data mining techniques for extraction of knowledge from educational domain to improve the performance of graduate students i.e. low grades. In their exercise, considered the data of college of science and technology. Finally, experimental results stated that they could have been succeeded to affected problem of low grades of students.

Brijesh Kumar Bhardwaj and Saurabh Pal [12] saidin their research that, parent's profession, living environment,physiological issue are also important behind student performance. But collecting this data are too difficult especially physiological data because we need a certified physiological specialist.

Kovacic [13] explored the possibilities of identifying successful and unsuccessful students based on the enrollmentdata. This study entirely focused on the socio-demographic variables and the study environment that may impact the dropout of the students from the distance education program.Data mining techniques such as Feature Selection, Classification and Regression algorithms are employed to identify the success and unsuccessful students. The empiricalresults showed that the ethnicity, course programme and the course block are the most important factors in classifying thesuccessful and uns0075ccessful students. Also, research shows that in comparison with the other techniques, Classification and Regression technique produced better results.

Jadhav and Channe [14] conducted two experiments to predict the student's final mark. The first experiment comparedclassification algorithms using three datasets in which the results achieved with better accuracy when all available data are taken into account versus filtering. In the second experiment, they concluded that the best accuracy can be obtained by applying classification model for both numerical and categorical data.

Nagy, Aly & Hegazy [15] proposed a "Student Advisory Framework" using the classification and clustering algorithms. The main aim of this research work is to provide a consultation to the first-year students to pursue certain education track where he/she may succeed. This model can considerably reduce academic failure as the students get a pre-consultation and a chance to decide their successful track at the very early stage. Naïve Bayesian, K-nearest neighborhood, Decision trees, Support Vector Machine, Linear Regression and rule induction techniques are used to classify the students and cluster them into groups to focus on the tracks.

III.   CASE STUDY

B The dataset used in this case study is student data set. In every educational system there is a lot of data stacking, thousands of transactions every year. You can imagine how difficult it would be if you had to analyze the data manually one by one

In the field of education, predicting student performance is crucial for educators and institutions to provide timely support and interventions. Educational institutions are continually seeking ways to identify students who may be at risk of academic challenges to provide timely support and interventions. For educational institutions to ensure that students receive the necessary support and interventions to succeed academically, this case study is suitable. Here I explore how a machine learning model was developed and deployed to predict student performance based on various factors such as demographics, previous academic performance, and attendance. The dataset used for this study includes information on students' demographics, study habits, and previous academic performance. . In this study the database used is taken from Kaggle website , the data collected from the transaction as much as 1000 studnets and with 08 different attributes are there.

3.1 Data Mining:-

Data mining is a process that uses statistics, mathematics, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases. Data mining is a part of knowledge discovery data which is a useful, unknown, and hidden information extraction process from data.

Gather relevant data about students, including academic records, demographic information, attendance, previous exam scores, socioeconomic factors, and any other variables that may influence student performance etc.

Dataset Description:

The dataset used for this case study contains the following attributes:

1. Gender: Student's gender (Male/Female).
2. Parental Level of Education: Highest education level of the student's parents.
3. Ethnicity: Student's ethnicity.
4. Lunch: Which type of lunch standard/free/reduced
5. Test Preparation: Whether the student completed test preparation courses (Yes/No).
6. Maths Score:- Maths score
7. Reading Score:- Reading score of student
8. Writing Score:- Writing score of student

3.2 Data Preparation:-

Clean the data by removing any outliers, handling missing values, and normalizing or standardizing the features to ensure all variables are on a similar scale. Here in this dataset it does not contain any missing values or duplicate records.

In given dataset '48.20%'male students and '51.80%' female students.Firstly, let's figure out the performance of each field for male and female.
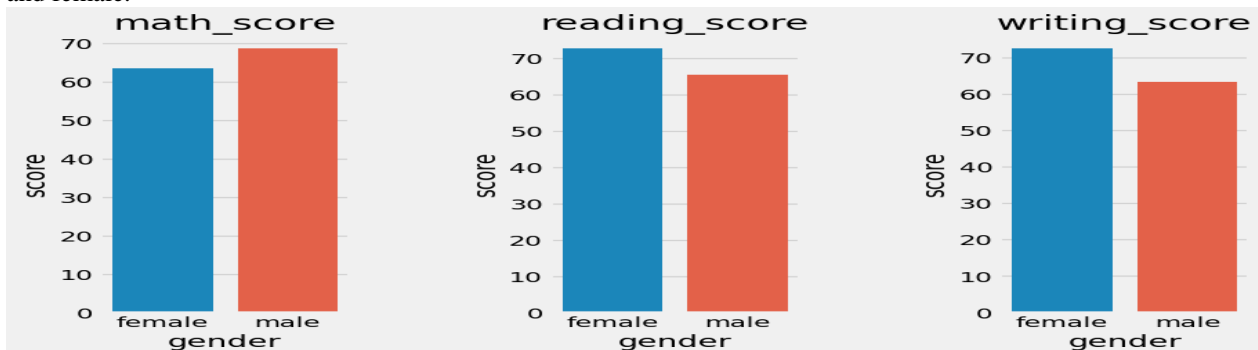

Fig. 1 Performance of male and female on each field

Here it can see that male has better performance on math field, but worse on reading and writing. Secondly, see the performance of ethnicity.
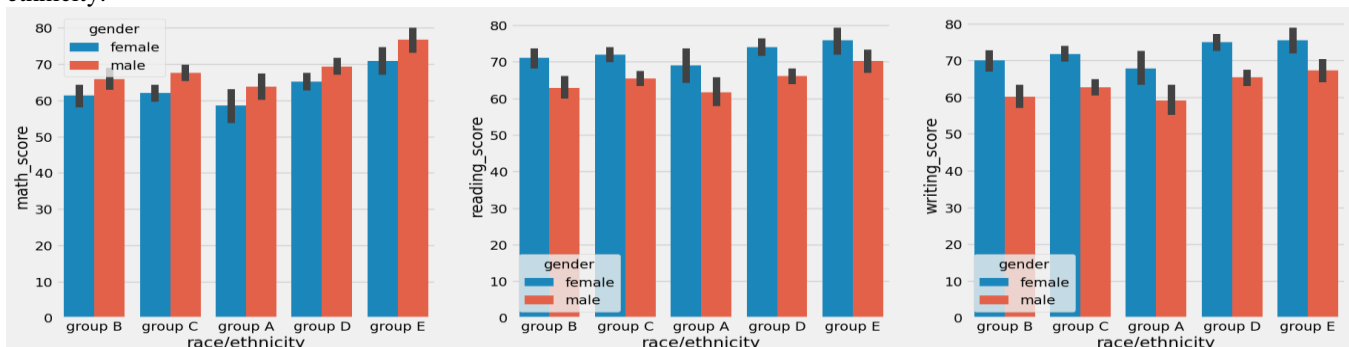

Fig. 2 Performance of male and female on ethnicity

Here it can see that male has better performance on ethnicity field, but worse on reading and writing. Means female has better performance on reading, writing. Obviously, group E has best performance for all the fields, and group A is the worst.

IV.   CLUSTERING ALGORITHUM

Clustering can be defined as the process of organizing objects in a database into clusters/groups such that objects with in the same cluster have a high degree of similarity, while objects belonging to different clusters have a high degree of dissimilarity. Clustering is an unsupervised learning approach in which the raw data are classified in a way so as to discover the inherent patterns that exist in data. Clustering partitions data points into smaller number of clusters such that objects are very similar within a cluster and very discordant to objects in other clusters.Clustering techniques fall into a group of un directed data mining tools. The goal of undirected data mining is to discover structure in the data as a whole.

4.1 **K-meansClusteringAlgorithm**

K-means is a well-known unsupervised, iterative, partitioning learning algorithm in the field of data mining. It is used in solving various clustering problem especially for large datasets. The algorithm has two separate parts. In first part, Knumber of centers are selected randomly. K is initially fixed. In second part, every data object is taken to the closest center . In order to choose the initial value of K, self-organising map can be used.
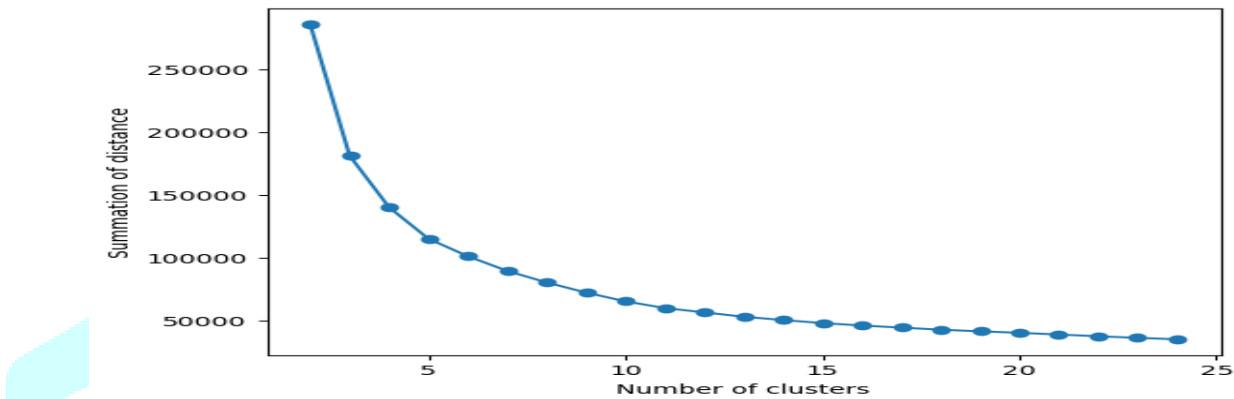

Fig. 3 k means clustering

on above fig. elbow point value is 8 so cluster is 8.From now on, I have find out the correlation between the performance of students and features. Let's plot pie chart to see whether parents education level can affect the performance or not.
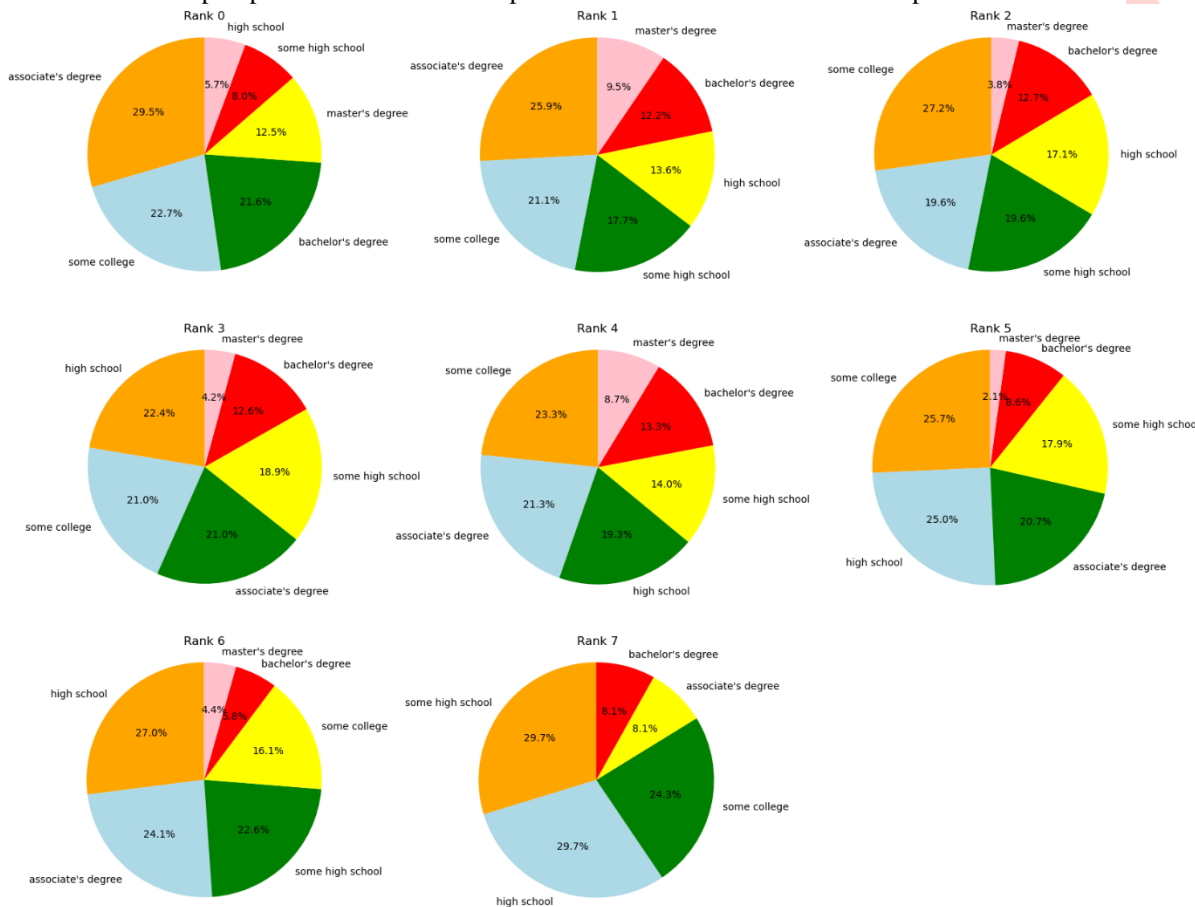

Fig 4 pie chart for parent's education level and cluster

Let's define the high degree of education. Parents having bachelor or master degree are high-level educated. So I have focused on these two terms.As pie chart were shown above, it can easily understand the ratio of high-degree education. For the rank0, its ratio is around 32%. In addition, there are no differences between rank1 to rank3, and the ratio are around 15~17%. Finally, the ratio is only 8% in rank7. I have calculated the average score of each rank before, so can say that parent's education affect the score but not obviously, because there are still 70%~80% parents without high education degree.

After that Over 50% of students in rank0 completed the test preparation course, and normally there is about 70~80% students in rank7 hadn't finished course. It is say that preparation course can help students get better score. Also the same trend as "pre". Students who had lunch before test got better score. That is, it's hard to get good performance without eating.

V.   CONCLUSION

In conclusion, student performance prediction is a valuable tool that can provide valuable insights and benefits to both students and educational institutions. By analyzing various factors and patterns, such as past academic records, socio-economic backgrounds, and personal attributes, predictive models can offer valuable predictions about students' future academic achievements.Therefore, while student performance prediction can provide valuable insights and support in promoting academic success, it should always be combined with qualitative assessments, teacher expertise, and a holistic understanding of the students' needs and aspirations. By integrating predictive analytics with a comprehensive approach to education, we as parent can maximize the potential of student performance prediction and create an environment that fosters student growth, achievement, and well-being.

Parents' education level may affect the performance of students, but not the important one. Finishing preparation course is benefitial. Having lunch is important to students, and it is also the most significant one. Gender has no correlation with the score. In summary, if students want to have good performance, they should have enough nutrient and make effort to prepare the test.

Looking at these graphs, there doesnt seem to be any discernable correlation for the types of lunch they had. There is however a massively clear correlation that females outperform the males in every single category.Generally speaking, females do better than males, with the exception of Maths, where males slightly excel.Out of the ones that took the Maths test, someone taking a standard lunch was more likely to achieve a higher result whether male or female.Group C's ethniicty values with a standard lunch meant that they were far more likely to get a higher score when compared with their writing scores. I think it goes without saying that there are very clear correlations. Gender and maths test with the lunch values as a hue is a fantastic graph to point out how clear this correlation really is.

REFERENCES

[1] H.M. Rafi Hasan, Mohammad Touhidul Islam, AKM Shahariar Azad Rabby, Syed Akhter Hossain ,"Machine Learning Algorithm for Student'sPerformance Prediction " IEEE 45670 , 10th ICCCNT 2019, July 6-8, 2019 , IIT - Kanpur,Kanpur, India

[2] Engr. Sana Bhutto, Dr. Isma Farah Siddiqui, Dr. Qasim Ali Arain, Maleeha Anwar, " Predicting Students' Academic PerformanceThrough Supervised Machine Learning", 978-1-7281-6899-9/20//$31.00 ©2020 IEEE, 2020 International Conference on Information Science and Communication Technology

[3] Boddeti Sravani, Myneni Madhu Bala, "Prediction of Student Performance using LinearRegression ",*2020 International Conference for Emerging Technology (INCET)Belgaum, India. Jun 5-7, 2020,* 978-1-7281-6221-8/20/$31.00 ©2020 IEEE

[4] https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning, accessed at 14.05.2019.

[5] Sujith Jayaprakash, Sangeetha Krishnan, Jaiganesh V, "Predicting Students Academic Performance usingan Improved Random Forest Classifier", 2020 International Conference on Emerging Smart Computing and Informatics (ESCI) AISSMS Institute of Information Technology, Pune, India. Mar 12-14, 2020, 978-1-7281-5263-9/20/$31.00 ©2020 IEEE

[6] Pratiyush Guleria[#1], Niveditta Thakur[*2], Manu Sood[#3], "Predicting Student Performance Using DecisionTree Classifiers and Information Gain" , 2014 International Conference on Parallel, Distributed and Grid Computing, , 978-1-4799-7683-6/14/$31.00©2014 IEEE

[7] V. S. Warke and R. S. Kamath "Data Mining Approach for the Analysis of Performance Based Appraisal System of Selected Teachers in Kolhapur City," no. Iv, pp. 1–6, 2016.

[8] Ihsan A. Abu Amra , Ashraf Y. A. Maghari , Students Performance Prediction Using KNN andNaïve Bayesian, *2017 8th International Conference on Information Technology (ICIT),* 978-1-5090-6332-1/17/$31.00 ©2017 IEEE

[9] Z. Ibrahim and D. Rusli, "Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression", In 21st Annual SAS Malaysia Forum, 5th September 2007, Shangri-La Hotel, Kuala Lumpur, 2007.

[10] Pauziah Mohd Arsad, Norlida Buniyamin and Jamalul-lail Ab Manan."A Neural Network Students' Performance Prediction Model (NNSPPM)" 2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Malaysia.

[11] M. Abu Tair, Alaa M. ElHalees, "Mining educational data to Improve Students' performance", International Journal of Information and Communication Technology Research, pp. 140-146. 2012.

[12] Brijesh Kumar Bhardwaj, Saurabh Pal. "Data Mining: A prediction for performance improvement using classification", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 2011.

[13] Kovacic, Z.J. (2012). "Predicting student success by mining enrolment data".

[14] S. D. Jadhav and H. P. Channe, "Comparative Study of K-NN , Naive Bayes and Decision Tree Classification Techniques," vol. 5, no. 1, pp. 2014–2017, 2016.

[15] Nagy, H. , Aly, W. , Hegazy, O. (2013). "An Educational Data Mining System for Advising Higher Education Students". World Academy of Science, Engineering and Technology, Open Science Index 82, International Journal of Computer and Information Engineering, 7(10), 1266 - 1270.