IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

Enhancing the Learning Experience of Students by Early Prediction of Student Academic Performance using Machine Learning

Kalaiselvi G[1], Akshata[2], Hema S M[3], Iswarya M[4]
Assistant Professor, Department of Computer Science Engineering[1]
Final Year Students, Department of Computer Science Engineering[2,3,4]
Anjalai Ammal Mahalingam Engineering College, Kovilvenni, Tiruvarur, Tamil Nadu, India

Abstract: *Machine learning (ML) is expected to provide a variety of ways and effective tools to improve education in general in the future. Digital data tracks from different sources covering various aspects of students' lives are stored daily on most modern university and college campuses. However, it is very challenging to get a complete view of a student and combine that data, use this data to accurately predict a student's academic performance, and use such predictions to enhance students' positive engagement with universities or colleges. First of all, data from previous year's students (name, year, department, category, mark, etc.) such as academic performance and behavioral measurements are entered using online forms. Subsequent features were extracted and machine learning-based training was provided and machine learning-based taxonomy algorithms will be developed to predict educational performance. Based on the accuracy obtained through the analyzed and tested Machine Learning algorithms, will provide a set of recommendations for teachers to improve students' quality and learning ability.*

Keywords: Machine learning, Supervised approach, linear regression, GPA

I. INTRODUCTION

As the use of computers and the internet has spread around the world, the amount of data that can be analyzed has rapidly expanded. Data can be anything pertaining to the population, as well as scholarly data. Data can include demographic information, student academic information, and personal hobbies. In this, the new data will be emerging from time to time. The first step is to analyze the data. For humans, this is a difficult task. As a result, the computer arrives. Because it has more computing power than humans, it can analyze data more quickly and stores the data in a well-formatted digital format.

This is where machine learning first appeared. Machine learning is a component of Artificial Intelligence that allows computers to learn from their prior experiences automatically. In this case, the Machines are expressly programmed. As the name implies, it offers the computer the power to make it when it comes to learning, humans, and machines are very similar. Machine learning is based on the nature of the learning signal. There are two types of learning: supervised learning and unsupervised learning. learning without supervision. The focus of this research is on supervised learning. learning, and in particular predictive analysis. Whenever Predictions of future outcomes have been completed. The importance of analysis cannot be overstated. The variety of applications for which The field of predictive analysis is extensive. Predicting a student's academic success is critical since it can alert teachers to pupils who may beat risk. Drop out of the race, and forecasting can help more support for students who need to improve their grades and their academic achievements.

It is crucial to predict a student's academic achievement since it can warn teachers of students who may be at risk. Forecasting can assist you if you drop out of the event. More assistance for students who need to raise their GPA and their academic successes. This dataset should be seen as a tabular format containing student information (such as age, gender, academic records, and medical information). Several algorithms can be used to develop a model that produces results for this dissertation. Algorithms are utilized in a variety of ways. models that can be predicted. The focus of this research is on how linear the student's academic performance has regressed and the performance based on the dataset provided by the students.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

1.1 Performance Prediction

Improving the quality of educational processes in order to improve student performance is one of the most difficult issues. Instructors can adapt theirteaching methods to meet the needs of low-performingstudents while also providing additional support to meritorious pupils. Students may be able to use the prediction results to help them design a strategy. Individuals have a clear grasp of how well or poorly they will perform in a subject and can then plan accordingly. Take the appropriate steps. These techniques are utilized to gain a better knowledge of thestudent and his or her learning environment. Any educational institution around the world has a long-term goal of increasing student retention. There are many positive impacts of increased retention such as increased college reputation, ranking, better job opportunities for alumni, etc.

II. LITERATURE SURVEY

To several research articles that are related to the thesis in order to specify the thesis as a well-structured thought. The following are the details of a few of the publications' conclusions. The linear regressiontechnique is described in detail in this research report. This method is used to forecast a student's academic achievement. In this research paper, a model named AugmentED is proposed to predict the academic performance of college students. These experimental results demonstrate that AugmentED can predict academic performance with quite a high accuracy, which helps to formulate personalized feedback for at-risk (or unself-disciplined) students [1]. This research paper proposed various predictive models trained on several ML and DL algorithms for predicting students' performance based on demographics variables, demographics and clickstream variables, and demographics clickstream and assessment variables [3]. This work aims to improve performance measure values by using ensemble methods and comparing them with previous studies that used the Jordan dataset to investigate student performance prediction. These results are obtained via the use of strategies and techniques, such as SMOTE, hyperparameter optimization, and cross-validation process, which demonstrate the dependability of the novel model [4]. This research paper shows overall results indicating that all predictive models derived from J48, NB,kNN, SVM, LR, and RF deliver better performance whenwe applied SMOTE independently to the imbalanced dataset [5]. In this research paper, the experiments show Linear discrimination analysis as the most effective approach to correctly predict the students' performance outcome in final exams. In this study, we have used the historic results of a course to evaluate the accuracy of students' success. It has tested several classification algorithms and found the linear discrimination analysis algorithm to be more accurate [6]. This paper implements the new applications of machine learning in teaching and learning by considering students' backgrounds and students' past academic scores and considering other attributes [7]. In this paper, the authors have proposed an ICGAN-DSVM algorithm to improve the prediction accuracy of students' performance. Results have revealed its effectiveness by comparing ICGAN and existing CGANs, heuristic-based MKL and typical kernel functions as well as between ICGAN-DSVM and related works [8]. In this paper, two predictive models have been designed namely students' assessment grades and final students' performance. The models can be used to identify the elements that influence MOOC students' learning outcomes. The result shows that both models gain feasibleand accurate results [9]. This paper proposes a student academic performance prediction model by applying threeclassification classifiers: KNN, Naive C4.5, and Bayes on the student data set of University A, which contains admission information and undergraduate GPA. The mainaim of this model is to help the college to select a candidate who has the potential for good academic performance. Theexperimental results show that the KNN algorithm is better than C4.5 and Naive Bayes [10].

III. METHODOLOGY

Collecting the data set required for the studyendeavor is the first step in the implementation process. The methodology is used on a dataset including information from the students. It can minimize our analysis by identifying the unique properties of the data collection. It can't be used for analysis. Following the data collection, the data is changed into the format thatis desired. This is how it works. Data pre-processing iswhat it's called. It is the most crucial stage. In order to extract the needed information from the raw data, the higher the rate of raw material pre-processing accuracy,the better. The higher the rate of accuracy of acceptabledata, the better.

After pre-processing the data, the following step is to detect and eliminate any incomplete or irrelevant data from the dataset in order to achieve correct findings. Getting rid of unnecessary data is known as the Data Wrangling.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

Following that, it can use any of the algorithms, such as linear, to solve the problem. Naive Bayes Standard, regression, Support vector machine for better classification and decision tree algorithms, classification.

In this paper, the linear regression algorithm isused to implement the solution. It must also select a training set from the dataset and identify the resultproperties that are relevant which determines the outputand begins the classification process.

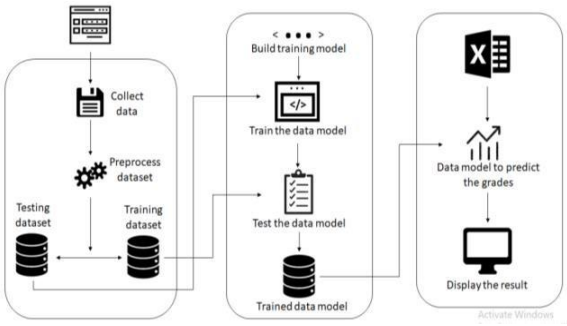

FIGURE 1: System Architecture

3.1 Algorithm

The project is implemented using a variety of algorithms. This employs linear regression in this project. Even though they are all used to forecast the dependent variable based on the independent variable. They differ in how they implement the independent variables.

3.2 Linear Regression

One of the machine learning methods is linearregression. It is based on supervised learning, which is awell-known and simple to understand method. Even bysomeone who isn't very knowledgeable about machine algorithms for learning, linear regression is exactly whatit sounds like. Regression is performed. It establishes thetwo parties' working connection. By fitting a regressionline to the data, two variables can be identified. There are two variables that are reliant on each other. An independent variable is a variable that is unrelated to the dependent variable. It is necessary to ensure that there is aconnection between the two. Before modeling, determine the dependent and independent variables. The degree to which the variables are related can be measured. The scatterplot can be used to determine this. The linear regression line is represented in the formof Y=a*X+b x

- Y-Dependent Variable x
- a-slope
- X-Independent Variable
- b-Intercept

The error rate between the predicted and true values can be reduced by using the best-fit regression line to the data. There are two types of linear regression. One of them is only one independent variable is used in simple linear regression. Multiple regression is the second type of regression that is employed A way of predicting the futureis linear regression. Multiple regressions are used in this type of analysis. This employs independent variables, which we are currently using for the purpose of the thesis.

PSEUDOCODE

1. Start
2. Read .csv files
3. Print df. head()
4. Print df. columns. tolist()
5. Select x=df. iloc [:, :1]
   y=df. iloc [:, :2]
6. Split x, y, test_size, random state

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

7. Apply Linear Regression
8. Train model.fit (X_train, Y_train)
9. Display coefficient of determination, intercept, slope
10. Test model_predict (X_test)
11. Print the predicted data (Y)
12. Print the actual data
13. np. array (Y_test)
14. Prediction= [8.5]
15. Display prediction data model. predict (prediction)
16. Stop

IV. IMPLEMENTATION AND RESULTS

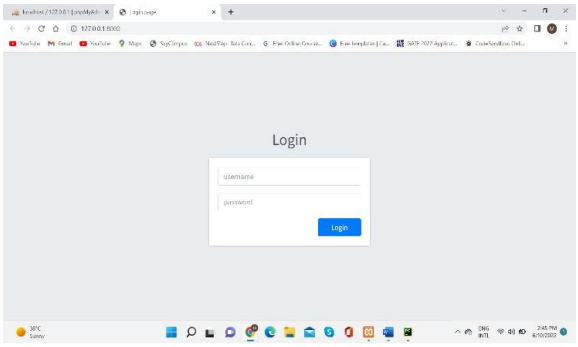

FIGURE 2: Login page

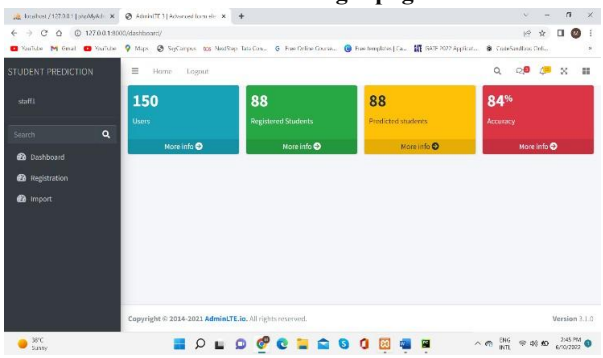

FIGURE 3: Dashboard

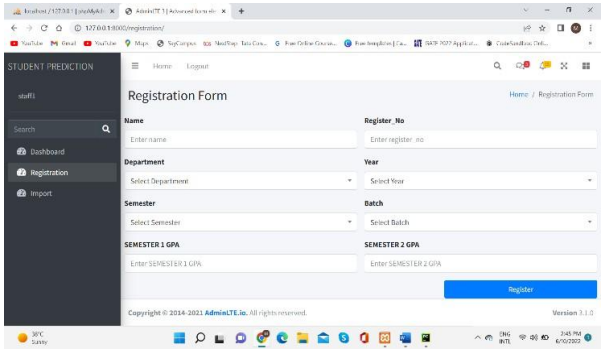

FIGURE 4: Registration Form

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

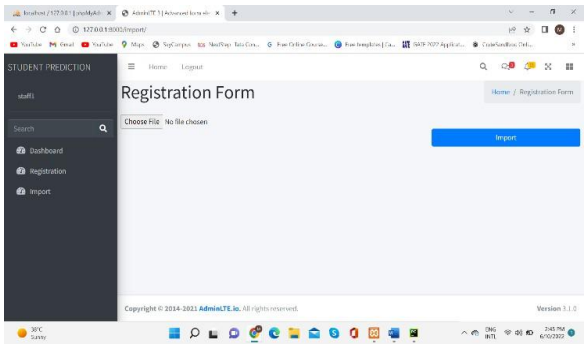

FIGURE 5: Importing .csv files

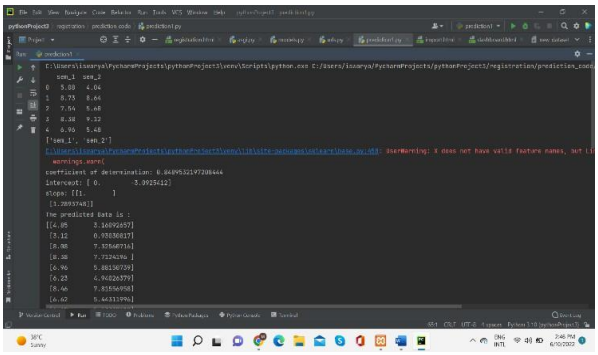

FIGURE 6: Output

V. CONCLUSION

The proposed system, predicts the academic performance of the students subsequently facilitating the instructors for timely intervention to persuade students to increase their study engagement and improve their performance. This research work presented a student performance prediction model based on the supervised learning technique of Linear Regression. This system will be helpful in many educational institutions to improve the performance of the students. These should draw on performance data to identify different student profiles in near real-time, which could be used to promptly adapt content to fit the students' needs, and for both students and educators, this is a useful tool for assessing student progress.

FUTURE WORK

Furthermore, to expand our data sets in the future, with a focus on accuracy and results. To upgrade our model in the future and provide some additional features for academic and industrial applications. To use the model in a firm to anticipate or analyze employee performance on a yearly or monthly basis, the corporation can then advise employees about their shortcomings to boost productivity. This research will also assist us in a variety of fields. This may serve as a warning to students, instructors, and parents who are concerned about their daughter's or son's academic standing. This model's limitation is that it will only function for this curriculum. The qualities' dependencies will vary depending on the curriculum. In addition, both input and output data will be collected, and create a website to forecast performance. Students and teachers can quickly acquire guidance from a website to boost their performance.

REFERENCES

[1]. L. Zhao et al., "Academic Performance Prediction Based on Multisource, Multi-feature Behavioral Data," in IEEE Access, vol. 9, pp. 5453- 5465, 2021, DOI: 10.1109/ACCESS.2020.3002791.

[2]. M. S. Ram, V. Srija, V. Bhargav, A. Madhavi and G. S. Kumar, "Machine Learning-Based Student Academic Performance Prediction," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 683-688,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

[3]. M. Adnan et al., "Predicting At-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models," in IEEEAccess, vol. 9, pp. 7519-7539, 2021,

[4]. H. Sahlaoui, E. A. A. Alaoui, A. Nayyar, S. Agoujil, and M. M. Jaber, "Predicting and Interpreting Student Performance Using Ensemble Models and Shapley Additive Explanations," in IEEE Access, vol. 9, pp. 152688-152703, 2021,

[5]. S. D. A. Bujang et al., "Multiclass Prediction Model for Student Grade Prediction Using Machine Learning," in IEEE Access, vol. 9, pp. 95608-95621, 2021,

[6].H. Gull, M. Saqib, S. Z. Iqbal, and S. Saeed,"Improving Learning Experience of Students by Early Prediction of Student Performance using Machine Learning," 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-4,

[7]. Sravani and M. M. Bala, "Prediction of Student Performance Using Linear Regression," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5,

[8]. K. T. Chui, R. W. Liu, M. Zhao, and P. O. De Pablos, "Predicting Students' Performance with School and Family Tutoring Using Generative Adversarial Network-Based Deep Support Vector Machine," in IEEE Access, vol. 8, pp. 86745-86752, 2020,

[9].R. Alshabandar, A. Hussain, R. Keight and W. Khan, "Students Performance Prediction in Online Courses Using Machine Learning Algorithms," 2020 International Joint Conference on Neural Networks(IJCNN), 2020, pp. 1-7.

[10]. Du, X. (2019). A Prediction Model for High School Students' Academic Performance in College Based on Machine Learning.

[11]. Boran Sekeroglu, Kamil Dimililer, and Kubra Tuncal. 2019. Student Performance Prediction and Classification Using Machine Learning Algorithms. In Proceedings of the 2019 8th International Conference on Educational and Information Technology (ICEIT 2019). Association for Computing Machinery, New York, NY, USA, 7–11 DOI: https://doi.org/10.1145/3318396.3318419

[12]. P. M. Moreno-Marcos, T. Pong, P. J. Muñoz-Merino, and C. Delgado Kloos, "Analysis of the Factors Influencing Learners' Performance Prediction With Learning Analytics," in IEEE Access, vol. 8, pp. 5264-5282, 2020, DOI: 10.1109/ACCESS.2019.2963503.

[13]. Veeraiyan, R., and Ramakrishnan, S., 2020. [online] Ijaema.com. Available at: [Accessed 14 September 2020].

[14]. Ade, R., 2020. Students Performance Prediction Using Hybrid Classifier Technique In Incremental Learning | International Journal Of Business Intelligence And Data Mining. [online] Inderscienceonline.com. Available at: [Accessed 14 September 2020].

[15]. Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., and Erven, G., 2020. Educational Data Mining: Predictive Analysis of Academic Performance of Public-School Students in The Capital of Brazil.

[16]. Jain and S. Solanki, "An Efficient Approach for Multiclass Student Performance Prediction based upon Machine Learning," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1457-1462, DOI: 10.1109/ICCES45898.2019.9002038.

[17]. W. Nuankaew and J. Thongkam, "Improving Student Academic Performance Prediction Models using Feature Selection," 2020 17th International Conference onElectrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 2020, pp. 392-395, DOI: 10.1109/ECTI-CON49241.2020.9158286.

[18]. Burman and S. Som, "Predicting Students Academic Performance Using Support Vector Machine," 2019 Amity International Conference on Artificial Intelligence (AICAI), 2019, pp. 756-759, DOI:10.1109/AICAI.2019.870126

[19]. Yağcı, M. Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learn. Environ. **9,** 11 (2022). https://doi.org/10.1186/s40561-022-00192-z

[20]. Sekeroglu, Boran & Dimililer, Kamil & Tuncal, Kubra. (2019). Student Performance Prediction and Classification Using Machine Learning Algorithms. ICEIT 2019: Proceedings of the 2019 8th International Conference on Educational and Information Technology. 7-11. 10.1145/3318396.331841