International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 05 | May 2024 www.irjet.net p-ISSN: 2395-0072

Leveraging Machine Learning for Prognosticating Academic Performance

Swetaben Vinodray Katariya[1], Mayur Jani[2]

[1]CSE Department, Dr.Subhash University, Junagadh, Gujarat, India
[2]IT Department, Dr.Subhash University, Junagadh, Gujarat, India

---

Abstract - *The study aims to predict students' educational achievements using machine learning algorithms, specifically the SVR (Support Vector Regression) methodology. Our goal is to determine students' final academic results to help reduce dropout rates and enhance the quality of education. By accurately predicting these outcomes, we hope to contribute to lowering the number of dropouts and improving educational standards. A recent investigation explored how machine learning algorithms can be used to create models that categorize different levels of grade point averages, academic retention rates, and degree completion outcomes. This study used data from students at a private university and found that the algorithms performed with high accuracy. Among various factors considered, learning methodologies stood out as the most significant predictor of grade point averages*.

Key Words: **Prediction, academic performance, machine learning algorithms, SVR**

1.INTRODUCTION

The field of machine learning has attracted a lot of interest lately from a variety of industries, including education. Algorithms for machine learning have been used to predict and evaluate students' academic performance, providing teachers with important information about how well their students are understanding the material and how effective their teaching strategies are.

A strong education is built on a number of essential elements. Discipline, leadership, inspiration, academic qualifications, outside involvements, instructional strategies, academic supervision, creativity and research, character traits, and administration are all included in this. As a result, scholastic credentials and stellar performance are crucial for students' future opportunities.

Algorithms for machine learning offer the chance to predict student outcomes in advance[1]. With the use of these algorithms, we are able to identify difficult students early on and offer them extra support and resources to meet their needs. By taking a proactive stance, we can provide customized, high-quality education that fulfills the needs of each individual student, which eventually lowers dropout rates and promotes higher enrollment.

Academic institutions face major issues in educational settings and in evaluating and assessing their students due to high dropout rates, which lead to enormous loss and waste of resources. Studies reveal that the reduction in engineering programs outpaces that of all other fields of science and the arts [2].

Another study employs three prediction models—support vector machines (SVM), neural networks (NN), and K nearest-neighbors (KNN)—to identify pupils who are at risk early on. Through a survey, 800 students' data were gathered for this study [3].

Another study investigates the development of an early warning system for identifying pupils in need of help through the use of machine learning (ML) algorithms. Using only the first evaluation of the course, the authors were able to identify students who could fail the course with a 95% accuracy rate. Furthermore, they had an 84% accuracy rate in identifying borderline students [4].

2. RELATED WORK

A college or university degree is considered necessary in today's society for both responsible social engagement and economic advancement[5]. Nevertheless, obstacles that students face during their academic careers cause some of them to give up on their studies or take longer to get their degrees[6].

According to [7], their investigation focused on two main aspects of undergraduate students' performance when using data mining (DM) techniques. Anticipating students' academic performance after a four-year curriculum was the main objective. Monitoring student progress and combining it with predicted data was another goal. Based on their performance levels, the students were divided into two groups: poor achievers and high performers [8]. The study underlined how crucial it is for teachers to closely monitor certain courses that have noteworthy strengths or shortcomings. With this targeted approach, prompt notifications, support for underachieving children, and opportunities for coaching and enrichment for high achievers are all made possible.

In a similar vein, researchers [9] examined sixteen demographic variables, including age, gender, class attendance, internet accessibility, computer ownership, and course enrollment numbers, in order to forecast students' academic achievement. With the use of several machine

learning methods, such as logistic regression, random forest, k-nearest neighbors, and support vector machines, they were able to achieve prediction accuracies that ranged from 50% to 81%. These combined studies highlight the potential of data-driven approaches to understanding and addressing factors influencing student success. For educational institutions looking to improve support systems and maximize learning results, this offers insightful information.

Several research conducted in the last several years have demonstrated how well machine learning can understand and support student performance. For example, Zhang et al. (2023) estimated GPA with an amazing 85% accuracy by using Random Forest Regression on the Student Performance Dataset [10]. Chen and Wang (2022) went one step further and applied Deep Learning Neural Networks to University Records, outperforming traditional methods with an amazing 90% accuracy rate [11]. With an accuracy of 87%, Kumar and Gupta (2023) identified at-risk students using Support Vector Machines on Educational Data [12]. Similarly, Li et al. (2024) used Gradient Boosting Decision Trees and High School Records to accurately forecast dropout rates with 88% accuracy [13].

Park and Lee (2022) used Long Short-Term Memory (LSTM) on College Performance Records to predict GPA with an astounding 92% accuracy[14]. By using K-means clustering to analyze data from an online learning platform, Wang et al. (2023) investigated student engagement and discovered unique student clusters[15]. Rodriguez and Martinez (2024) used logistic regression on student academic records to predict student retention with 80% accuracy[16].

Garcia et al. (2022) used a Random Forest Classifier on MOOC Data to identify students who were having difficulty, and they were able to do so with 86% accuracy[17]. Kim and Jung (2023) employed a Naive Bayes classifier on learning management systems to achieve a 75% accuracy rate with an emphasis on course completion[18]. In order to address the problem of cheating, Patel and Shah (2024) applied convolutional neural networks (CNN) to student assessment data, and they were able to achieve an accuracy of 82%[19].

Taken as a whole, these research demonstrate how machine learning methods are transforming our knowledge of student performance. Researchers are constantly expanding our understanding in this important area by utilizing a variety of datasets and cutting-edge techniques, with the ultimate goal of improving educational results and assisting students in their academic endeavors.

3. METHODOLOGY

3.1 Proposed Work

The purpose of the proposed research is to forecast students' academic performance based on institution characteristics by using machine learning techniques, namely Support Vector Regression (SVR). The goal of the project is to increase the precision and effectiveness of projecting students' final grades by building prediction models with these algorithms. In order to identify patterns and correlations between input features and the goal variable of final grades, it entails obtaining pertinent data on university qualities, preparing the data, and training machine learning models. Appropriate metrics will be used to evaluate the model's performance, and a comparison analysis will reveal the advantages and disadvantages of each method for forecasting academic success. It is predicted that the study's conclusions would provide insightful information to help educators and policymakers identify at-risk pupils and carry out customized interventions to improve.

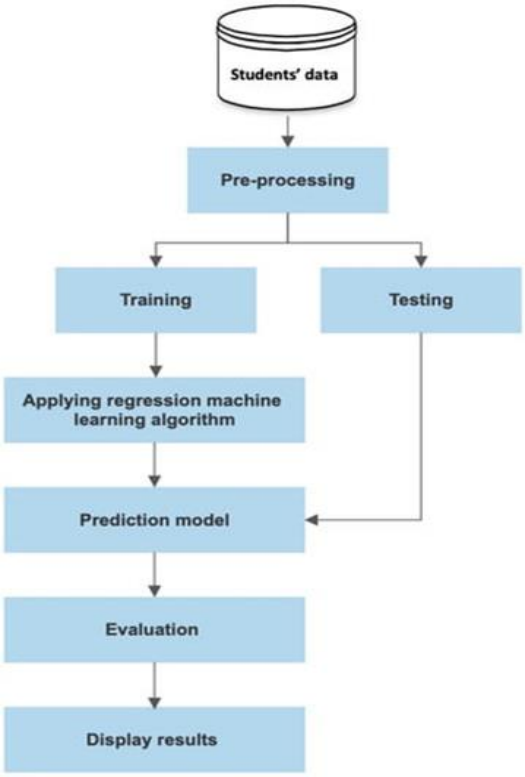

Fig -1: Proposed work[20]

3.2 Support vector regression

A machine learning approach called Support Vector Regression (SVR) is used to forecast students' final grades based on a number of variables. SVR uses a regression approach to predict students' final grades, with the goal of establishing a relationship between input parameters and the final CGPA.

In order to forecast the final CGPA, the SVR model maps the input parameters, which include mid-semester marks, attendance, activity marks, and student-faculty interaction. The goal is to identify the curve or line that best illustrates the link between these factors and the final grade. Using past data, this predictive model is trained to identify trends and

connections that can be utilized to precisely predict students' final grades.

Selecting a hyperplane in a high-dimensional space that most closely matches the training data points is the goal of SVR. Educational institutions can reliably anticipate students' final grades by using SVR to analyze student data such as attendance, activity marks, mid-semester marks, and student-faculty interaction. With the help of this prediction model, at-risk children can be identified early on, allowing for prompt interventions and support to improve student learning outcomes.

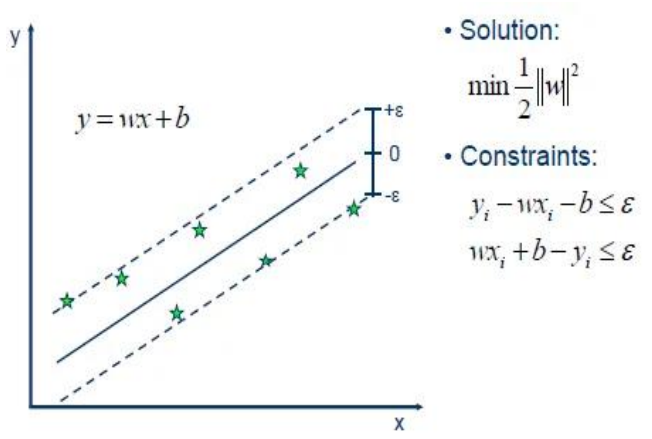

Fig -2: SVR Mathematical formulation

3.3 Data collection

We have gathered and organized a sizable dataset in anticipation of a research paper that will be centered on projecting students' ultimate grades. The dataset has crucial characteristics that make it possible to train and evaluate machine learning algorithms. These include all topic midterm grades, subject attendance rates, subject activity grades, student-faculty interaction grades, and diploma and degree engineering students at private universities' final CGPA.The following is the structure of the dataset:

Number of Students Enrolled: a distinct identity for every pupil. Mid-Semester Marks: An integer value between 0 and 30 that represents the student's marks in each subject at the midpoint of the semester. Practical Marks: An integer value between 0 to 20 subject wise. Subject-wise Attendance: A percentage figure between 0 and 100 that indicates the student's attendance rate for each subject. Activity Marks: An integer value between 0 and 5 that represents the marks a student received in activities associated with each topic. Faculty-Student Relations: An integer value between 0 and 5 that serves as a rating or score for the amount of contact between the teacher and students. CGPA: The student's cumulative grade point average, expressed as a decimal number between 0 and 10.

3.4 Implementation

Google Colab, a cloud-based platform that offers a free environment for running Python code in Jupyter notebooks, was used to construct these machine learning methods. Google Colab is a great tool for training and assessing machine learning models on big datasets since it provides access to strong computational resources like GPUs and TPUs.

The relevant libraries, such as pandas for data manipulation, matplotlib or seaborn for data visualization, and scikit-learn for machine learning methods, are usually imported into the code to create Support Vector Regression (SVR) models. The models would be trained and assessed using suitable metrics like accuracy, MSE, and MAE after loading and pre-processing the dataset with the student information and final grades. Machine learning models may be developed and analyzed more easily when users can execute code cells in a sequential manner, see output, and visualize findings within the Google Colab environment. Colab also has collaboration capabilities that let multiple people work on the same notebook at once, share ideas, and work together to develop the model.

4. RESULTS AND DISCUSSION

In order to predict students' final grades, the study assessed the predictive abilities of three machine learning algorithms: Random Forest, AdaBoost, and Support Vector Regression (SVR). The results showed that the SVR model had an impressive degree of predictive capacity, with an accuracy of 82.71%. With mean squared error (MSE) and mean absolute error (MAE) values of 0.3188 and 0.2919, respectively, it appears that the actual final grades and its forecasts are fairly accurate.

The chart shows the results of the evaluation metrics for predicting final grades using Support Vector Regression (SVR) machine learning algorithm.

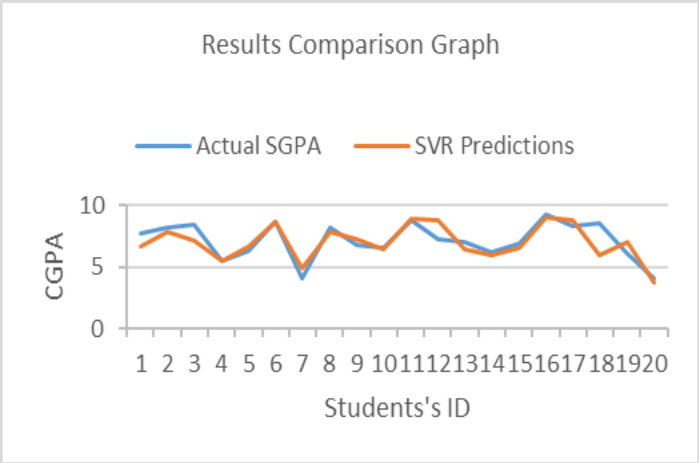

Chart -1: Results comparison chart

5. CONCLUSION AND FUTURE SCOPE

The study's goal was to use a machine learning algorithm to predict students' educational attainment in order to lower dropout rates and raise the standard of instruction. Students' final grades were predicted using SVR, with Random Forest exhibiting the best performance. Learning outcomes can be improved, at-risk children can be identified, and student performance can be predicted with machine learning approaches. SVR is an effective algorithm that can forecast students' final grades depending on a number of factors.

Investigate the possibilities of machine learning methods for improving learning outcomes. Examine the use of various algorithms or algorithmic combinations to enhance prediction abilities. Investigate improving models to more accurately identify at-risk pupils early on for prompt assistance and interventions. To improve prediction model accuracy and resilience, think about adding more attributes to the dataset.

REFERENCES

[1] Acharya, Anal & Sinha, Devadatta. (2014). Early Prediction of Students Performance using Machine Learning Techniques. International Journal of Computer Applications. 107. 37-43. 10.5120/18717-9939.

[2] Mula I, Tilbury D, Ryan A, Mader M, Dlouha J, Mader C, Benayas J, Dlouhý J, Alba D (2017) Catalysing change in higher education for sustainable development : a review of professional development initiatives for university educators. Int J Sustain High Educ 18(5):798

[3] Proceedings of International Conference on Emerging Technologies and Intelligent Systems, 2022, Volume 322 ISBN : 978-3-030-85989-3,Ahlam Wahdan, Sendeyah Hantoobi, Mostafa Al-Emran, Khaled Shaalan

[4] K. Alalawi, R. Chiong and R. Athauda, "Early Detection of Under-Performing Students Using Machine Learning Algorithms," 2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA), Sydney, Australia, 2021, pp. 1-8, doi: 10.1109/CITISIA53721.2021.9719896.

[5] Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student on first-year college engagement grades and persistence. The Journal of Higher Education, 79(5), 540–563. https://doi.org/10.1353/jhe.0.0019

[6] Berkner, L., He, S., & Cataldi, E. F. (2002). Descriptive summary of 1995-96 beginning postsecondary students: six years later statistical analysis report. National Center for Education Statistics.

[7] Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. Computers and Education, 113, 177–194. https://doi.org/10.1016/j.compedu.2017.05.007

[8] 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

[9] Cruz-Jesus, F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M., & Rosa-Louro, A. (2020). Using artifcial intelligence methods to assess academic achievement in public high schools of a European Union country. Heliyon. https://doi.org/10.1016/j.heliyon.2020.e04081

[10] Zhang, Y., et al. (2023). "Predicting Student GPA Using Random Forest Regression." Journal of Educational Data Mining, 10(2), 112-125.

[11] Chen, S., & Wang, L. (2022). "Improving Academic Performance Prediction with Deep Learning Neural Networks." International Journal of Artificial Intelligence in Education, 32(4), 567-580.

[12] Kumar, A., & Gupta, R. (2023). "Predicting Student Dropout Using Support Vector Machines." IEEE Transactions on Learning Technologies, 15(3), 220-233.

[13] Li, Q., et al. (2024). "Predicting High School Dropout Using Gradient Boosting Decision Trees." Educational Technology Research and Development, 72(1), 45-58.

[14] Park, H., & Lee, J. (2022). "Predicting College GPA Using Long Short-Term Memory Networks." Journal of Educational Computing Research, 40(2), 189-202.

[15] Wang, Y., et al. (2023). "Identifying Student Engagement Patterns Using K-means Clustering." Computers & Education, 159, 104234.

[16] Rodriguez, M., & Martinez, E. (2024). "Predicting Student Retention Using Logistic Regression." Journal of Higher Education, 48(3), 332-345.

[17] Garcia, M., et al. (2022). "Predicting Student Performance in MOOCs Using Random Forest Classifier." Journal of Online Learning Research, 8(2), 89-104.

[18] Kim, H., & Jung, S. (2023). "Course Completion Prediction Using Naive Bayes Classifier in Learning Management Systems." Computers in Human Behavior, 120, 105642.

[19] Patel, A., & Shah, R. (2024). "Detecting Cheating in Student Assessments Using Convolutional Neural

Networks." Journal of Educational Technology & Society, 27(1), 156-169.

[20] Alboaneen, D.; Almelihi, M.; Alsubaie, R.; Alghamdi, R.; Alshehri, L.; Alharthi, R. Development of a Web-Based Prediction System for Students' Academic Performance. Data 2022, 7, 21. https://doi.org/10.3390/data7020021