ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

Evaluating Predictive Models for Student Test Scores: A Machine Learning Perspective

K.C.Bhanu[1], Dr.P.Uma Maheswari Devi[2]

[1]Research Scholar, Department of Commerce and Management Studies
Adikavi Nannayya University ,Rajahmundry
[2]Associate Professor , Department of Commerce and Management Studies
Adikavi Nannayya University ,Rajahmundry
bhanu1605@gmail.com,umadevi_4@yahoo.com

ABSTRACT:

Student performance evaluation, which provides informative data about each student's learning outcomes and overall academic development, is an essential part of education. Predicting student test results has grown in importance as a research area with the potential to revolutionise educational practises and assist personalised learning approaches. By accurately forecasting student performance, educators can identify those who are at risk of failing and use early intervention strategies. Additionally, tailored training is made possible through test result prediction, which enhances the learning process overall. This is done by tailoring teaching methods and course content to each student's needs. Machine learning and statistical methods are most typically utilised to assess the collected data and develop prediction models. Regression models, for instance, can be used to find connections between the anticipated test score (dependent) and variables like past test results or other characteristics (independent). This strategy makes it feasible for early intervention, personalised instruction, and informed decision-making, all of which enhance student learning outcomes.

KEYWORDS: teaching strategy, prediction models, performance evaluation, regression models.

INTRODUCTION:

One area of this profession that is getting greater consideration and research is the forecasting of student test results. Educators and researchers are developing models and strategies to predict student success on educational assessments in order to enhance educational systems, individualise training, and enable targeted interventions. The capacity to forecast student test scores has the power to fundamentally alter the way we approach education and raise learning outcomes for students.

Early intervention and personalised support are possible when children who may be at risk of underachievement are identified. By identifying the factors that influence student performance and adapting instructional strategies, tools, and curriculum to meet specific needs, teachers can create learning experiences that are more effective and individualised. In

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

addition to aiding in decision-making and planning for education, test prediction can also help allocate resources more effectively.

A range of data sources and analytical techniques are used to forecast student test results. Researchers typically take into account contextual variables, such as demographic information, prior academic achievement, socioeconomic conditions, and other contextual characteristics, to develop prediction models. Advanced data analysis tools, such as statistical methods and the continual use of machine learning methodology, are used to uncover patterns, correlations, and prediction indicators within the data. These models are created to generate accurate estimates of student test scores prior to tests.

OBJECTIVE:

• To construct a reliable prediction model that accurately predicts future test results based on the stated variables.

• To study the relationship between the students' performance and parental education.

LITERATURE REVIEW:

[1] In their essay "Predicting students' final exam scores from their course activities," Ashenafi, Riccardi, and Ronchetti looked at how to predict students' final exam results based on their course activities. The study examines the potential use of a variety of activities, such as participation, tests, and homework assignments, to determine a student's general success in a course. In order to estimate how well students would perform on their final exams, the study provides a prediction model that uses these course assignments. [2] Yavuz's research, "A study on variables that affect class scores of primary education students in placement test," focuses on identifying the elements that influence these kids' placement test results. Understanding factors that affect pupils' performance is the aim. The study employs data analysis to get insights into the many variables that could affect students' class marks, giving policymakers vital information to improve the effectiveness of placement exams and improve academic outcomes. [3] Namoun and Alshanqiti assess the use of data mining and learning analytics techniques in their structured review of the literature titled "Predicting student performance using data mining and learning analytics techniques." The study contributes to our understanding of how these tactics can enhance instructional practises and support decision-making processes concerning student performance. [4] In their article titled "Data mining approach for predicting student performance," Osmanbegovic and Suljic discuss a study that focuses on data mining to estimate the students' progress. The study studies the use of data mining algorithms for assessing student data in order to estimate academic outcomes. The study looks at the methodology used and presents the results of anticipating student achievement. [5] This paper, "Predicting student performance using advanced learning analytics," was presented by Daud et al. This highlights how advanced analytics have the ability to provide informative data on student performance and to enable customised interventions to improve learning outcomes.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

DATA SET:

As a secondary source of information, the data was gathered from the Kaggle website. Data analysis was done using a machine learning approach and a Python Jupyter Notebook. Gender, Ethnic Group (this part covers the pupils' backgrounds in terms of culture or ethnicity), Parental Education (Details about the parent's educational experience and how it may affect the academic success of the child) Lunch Type (This area defines the preferred lunch to determine the student's socioeconomic level), Test preparation (let's the learner know whether they need any further help), Sports (which indicates whether the student is interested in sports), Parent Marital Status (which indicates the parent's marital status), First kid (asks if the student is the family's first-born kid), Number of Siblings (the number of the student's siblings), Transport Means (the method used by the student to get to and from school) Weekly Study Hours (Students' weekly study hour), Math Score (Score on students' math test), Reading Score (Score on students' reading test), and Writing Score (Score on students' writing test).

DATA ANALYSIS:

There are 14,758 males and 14,928 women who are officially recognised as such, according to the count. This information reveals that the dataset has a relatively balanced gender representation, with a slightly higher number of girls than boys.

MULTIVARIATE ANALYSIS:

Pictorial representation of various factors:

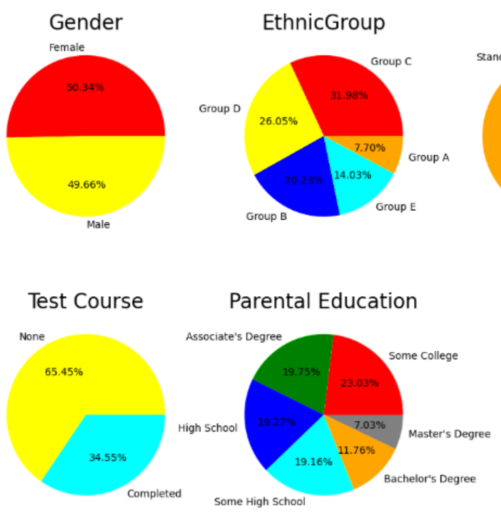

Figure 1: Graphical representation of independent factors.

The pie chart shows the general overview, which includes:

• There are an equal number of male and female students; • Students from the Group C ethnic group are considerably more numerous; and

• Students who take the normal lunch are more.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

• The number of students not taking exam preparation courses is high.

• Parental education at'some college' exceeds associate's and master's degrees.

BIVARIATE ANALYSIS:

Comparison between math, reading and writing score based on parental education.

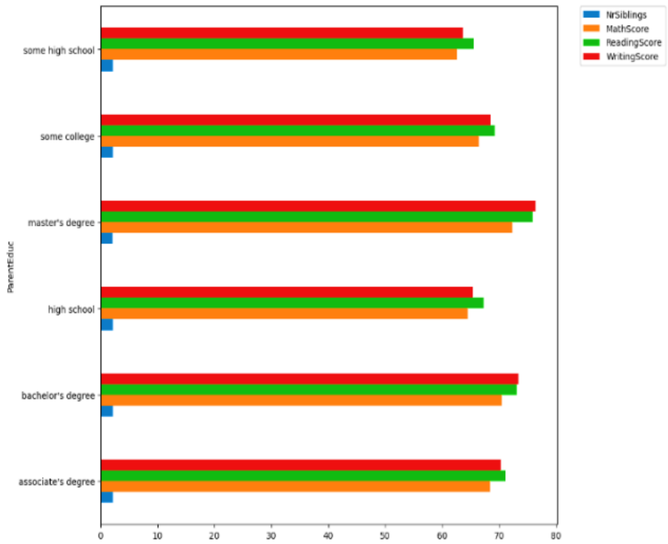

Figure 2: Comparison between math, reading and writing score

• Students with master's degrees perform exceptionally well across the board; students with bachelor's degrees have reading and writing scores that are somewhat higher than those of students with master's degrees.

• There are always an equal number of siblings.

• In all three scenarios, parents with a high school diploma have the fewest points.

• Bachelor's and master's degrees are roughly equivalent to associate degrees.

DESCRIPTIVE STATISTICS:

	MathScore	ReadingScore	WritingScore
Count	30641.000000	30641.000000	30641.000000
Mean	67	69	68
Standard Deviation	15.361616	14.758952	15.443525
Minimum	0	10.00	4.00
25 Percent	56.00	59.00	58.00
50 Percent	67.00	70.00	69.00
75 Percent	78.00	80.00	79.00
Maximum	100.00	100.00	100.00

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

• The information in the following table is comparable to one another and ranges from 66.37 and 69.55.

• There is a modest reduction in the standard deviation between 14.7 and 15.44.

• The minimal math score is zero, whereas the minimum writing and reading scores are 10, respectively.

Only the math score is analysed using the six model. This leads to the conclusion.

MODELS	ERROR	MATH SCORE
LINEAR REGRESSION	MEAN SQUARED ERROR	28.9822
	MEAN ABSOLUTE ERROR	4.3065
	R SQUARED	**0.877189**
DECISION TREE REGRES SOR	MSE	3305585
	MAE	4.6150
	R2	**0.857797**
RANDOM FOREST REGRE SSOR	MSE	30.2078
	MAE	4.3884
	R2	**0.871995**
GRADIENT BOOSTING RE GRESSOR	MSE	28.2739
	MAE	4.2509
	R2	**0.880190**
LASSO REGRESSOR	MSE	35.7654
	MAE	4.7877
	R2	**0.848445**
K-NEIGHBORS REGRESSOR	MSE	0.0093
	MAE	0.0022
	R2	**0.999961**

ASCENDING ORDER OF THE MODEL

	Model Name	R2_Score
0	Linear Regression	0.877189
1	Lasso	0.848445
2	K-Neighbour Regression	0.999961
3	Decision Tree	0.857797
4	Random Forest Regression	0.871995
5	Gradient Boosting	0.880190

Table1: Ascending order of the model based on the R2_Score.

INFERENCE:
A rating of 0.999961 shows that the K-Neighbour's Regressor can forecast the target variable with a very high degree of accuracy since it can explain about 99.9961% of the variation in the target variable using the provided attributes. The model is doing a fantastic job of

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

identifying the fundamental patterns in the data, showing a strong relationship between the properties of the input & the result.

An R2 score of 1.0 indicates a perfect fit, and in the result, this type will be taken into consideration as an option. The model's performance should be assessed using additional assessment metrics, even though the score of 0.999961 is quite close to 1.0. If the model is too complex for the provided data, overfitting may occur.

CONCLUSION:

The exam score forecasting provides personalised training and learning strategies. By being aware of each student's unique needs, teachers can adapt their teaching methods and develop original course materials to optimise learning for all students. This personalised strategy can enhance students' general academic development and provide more effective content interaction. One of the statistical and machine learning techniques used in test score prediction is regression modelling. With the aid of these techniques, educators can look at a variety of factors that could affect test results, including historical performance, demographic characteristics, and other relevant factors. By seeing connections and patterns, educators can use data to improve teaching tactics and optimise student learning results.

Therefore, by enabling early intervention, individualised instruction, and sensible decision-making, test score prediction has the potential to alter education. By applying data analysis and prediction models, educators may raise student performance, identify areas that require improvement, and create a more productive and welcoming learning environment. The ultimate goals are to maximise student growth and make sure that every student reaches their potential.

References:

1. Al-Shabandar, R., Hussain, A., Laws, A., Keight, R., Lunn, J., & Radi, N. (2017). ‘Machine learning approaches to predict learning outcomes in Massive open online courses’, *Proceedings of the International Joint Conference on Neural Networks*.
2. Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. International Journal of Educational Technology in Higher Education, 17(1).
3. Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. AMd. (2021). Multiclass prediction model for student grade prediction using machine learning. IEEE Access, 9, 95608–95621.
4. Coman, C., Țîru, L. G., Meseșan-Schmitz, L., Stanciu, C., & Bularca, M. C. (2020). Online teaching and learning in higher education during the coronavirus pandemic: Students’ perspective. Sustainability, 12(24), 10367.
5. Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students’ academic failure in introductory programming courses. Computers in Human Behavior, 73, 247–256.
6. Jayashree, G., & Priya, C. (2019). Design of visibility for order lifecycle using datawarehouse. International Journal of Engineering and Advanced Technology, 8(6), 4700–4707.

7. Krawczyk, B., Minku, L. L., Gama, J., Stefanowski, J., & Woźniak, M. (2017). Ensemble learning for data stream analysis: A survey. Information Fusion, 37, 132–156.
8. Landset, S., Khoshgoftaar, T. M., Richter, A. N., & Hasanin, T. (2015). A survey of open source tools for machine learning with big data in the Hadoop ecosystem. Journal of Big Data, 2(1).
9. Liu, Z., Yang, C., Rüdian, S., Liu, S., Zhao, L., & Wang, T. (2019). Temporal emotion-aspect modeling for discovering what students are concerned about in online course forums. Interactive Learning Environments, 27(5–6), 598–627. https://doi.org/10.1080/10494820.2019.1610449
10. Mourdi, Y., Sadgal, M., El Kabtane, H., & Berrada Fathi, W. (2019). A machine learning-based methodology to predict learners' dropout, success or failure in MOOCs. International Journal of Web Information Systems, 15(5), 489–509.
11. Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. International Journal of Computer Sciences and Engineering, 6(10).
12. Shida, N., Osman, S., & Abdullah, A. H. (2019). Students' perceptions of the use of asynchronous discussion forums, quizzes, and uploaded resources. International Journal of Recent Technology and Engineering, 8(2S9), 704–708.
13. J. L. Hoffman and K. E. Lowitzki, "Predicting college success with high school grades and test scores: Limitations for minority students," Review of Higher Education, vol. 28, no. 4, 2005, doi: 10.1353/RHE.2005.0042.
14. M. M. Ashenafi, G. Riccardi, and M. Ronchetti, "Predicting students' final exam scores from their course activities," Proceedings - Frontiers in Education Conference, FIE, vol. 2015, Dec. 2015, doi: 10.1109/FIE.2015.7344081.
15. M. YAVUZ, "A study on variables that affect class scores of primary education students in placement test," İlköğretim Online (elektronik), vol. 9, no. 2, pp. 705–713, 2010, Accessed: Jul. 24, 2023. [Online]. Available: http://search/yayin/detay/104220
16. Daud, M. D. Lytras, N. R. Aljohani, F. Abbas, R. A. Abbasi, and J. S. Alowibdi, "Predicting student performance using advanced learning analytics," 26th International World Wide Web Conference 2017, WWW 2017 Companion, pp. 415–421, 2017, doi: 10.1145/3041021.3054164.
17. E. ; Osmanbegovic and M. Suljic, "Data Mining Approach for Predicting Student Performance," Economic Review: Journal of Economics and Business, vol. 10, no. 1, pp. 3–12, 2012, Accessed: Jul. 24, 2023. [Online]. Available: http://hdl.handle.net/10419/193806