Educational Administration: Theory and Practice
2025, 31(2), 202-212
ISSN: 2148-2403
https://kuey.net/ **Research Article**

Algometrics: A Data Driven Application for Academic Insights

Devi Naveen[1], Mohit Sanjeev Mahajan[2*], Neha R Deshpande[3], Rakshitha Raghuram[4]

[1]Assistant Professor, Dept of CSE BNM Institute of Technology Bangalore, India
[2*]Student, Department of CSE, BNM Institute of Technology Bangalore, India. Email: mohitmah50@gmail.com
[3,4]Student, Department of CSE, BNM Institute of Technology Bangalore, India.

***Corresponding Author:** Mohit Sanjeev Mahajan
*Email: mohitmah50@gmail.com

Citation: Mohit Sanjeev Mahajan, (2025), Algometrics: A Data Driven Application for Academic Insights, *Educational Administration: Theory and Practice*, 31(2), 202-212
Doi: 10.53555/kuey.v31i2.10520

ARTICLE INFO

ABSTRACT

Identifying students who are at risk early is of utmost importance for improving academic success, reducing dropouts and ensuring timely help for students who require attention with their courses. Students face challenges because of socio-economic conditions, lack of academic help, or inadequate preparation, this can lead to poor performance, repeated failures, or even withdrawal from their studies. Educational institutions very often struggle with effectively allocating resources to support these students, which makes data-driven approaches critical for performance analysis of students. This paper presents Algometrics, a web-based application that leverages machine learning and data analytics to predict student performance considering multiple permutations and combinations of academic and socio-economic factors. The tool processes datasets containing parameters such as race/ethnicity, parental level of education, families' monthly/annual income, gender, and individual subject scores in mathematics, reading and writing. The system has predictive modeling and integrated interactive visualizations, providing educators with insights into the performance trends of students enabling data-driven interventions. The integrated dashboard provides multiple analytical charts such as correlation heatmaps, score distribution trends, and demographic and socio-economic conditions in shaping student success. Algometrics assists institutions in identifying students at risk of failing or dropping out and underperforming through a scientific and data-driven approach. Thus, enabling targeted interventions, personalized learning strategies, and improved resource allocation.

Keywords: RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory), Predictive Analytics, Regression Analysis, Data Visualization, Big Data in Education.

Introduction

In a 2019 study, it was ascertained that 40% of school funding inequities occur because of improper distribution of teaching resources. When more than 20% of teachers leave school, instructional quality and learning in those schools suffers. But there is over 30% improvement observed in students' performance if the resources like study materials, teaching methods and time are allocated efficiently using data-driven strategies. [1].

Data driven projections are a helpful and necessary tool for schools to identify at-risk students, as well as allocate their teaching resources effectively to improve overall performance of students. Institutional efficiency will be restored when instructors use the student data insights to improve their teaching approaches by coming up with innovative ways (like extra classes or special teaching methods) to help school students attain individual best potential levels.

The Role of Early Intervention in Academic Success

Early intervention techniques along with the right support, like tutoring or counseling, has decreased a number of students who might leave their educational institutions to a great extent. So, student retention rates could increase by
30% with this approach and more students will get to complete their studies. Early intervention significantly improves the performance of pupils. Impact of teacher-student engagement in academics showed an inverse proportionality and reduced dropout rates by 40% [2].

Impact of Teacher-Student ratios and Resource Allocations

Teacher-Student ratio is an important parameter that determines academic success. Fewer than 14 pupils per instructor improves focus, personalized learning and students' performance by 20%. The research additionally determined that in many underfunded schools, shortage of teaching staff resulted in teacher-student ratio exceeding 1:40. This consequently limited individual attention and personalized learning. It concluded the study by highlighting the importance of predictive analytics in education and its role in dynamically adjusting teacher assignments to students that require more support, thereby increasing efficiency [3].

The Need for a Data-Driven Approach

Schools and Institutions largely rely on reactive measures and traditional methods of resource allocation and performance analysis. Intervention strictly after learning gaps does not help greatly. Data-driven models' approach can identify at-risk students before they fail a grade, enabling
timely preventions and interventions into mitigating academic gaps [4].

Literature Review

Data-backed analyses using model training-testing and predictive analytics has gained importance, aiming to identify at-risk students and improve academic outcomes through early intervening techniques. The following discussions compare and review related solutions and studies contributed to this particular field.
Advancements in data analytics has enhanced the ability to forecast student performance and detect possibilities of dropouts. Even then, challenges persist. Various predictive models primarily have reliance on numerical data, which leads to heavy negligence of external factors like parental education and backgrounds, which are equally affecting a student's performance. It is crucial to take the factors into consideration for making accurate predictions. The absence of real-time school resource allocation mechanisms hinders the practical application of these models in dynamic educational settings. There is also the concerns regarding scalability and generalizability of these models, which make it difficult to apply them across diverse educational contexts.
In 2021 study, O trellis et al. applied a pair of models with machine learning approach to predict student performance in higher education. Their method included combination of unsupervised learning like K-Means clustering and supervised learning like RF (random forest). This approach was applied to CSE curriculum at Thessaly University, Greece. Initially students were separated based on educational metrics using K-means clustering and 3 distinct clusters were formed at the end. RF was applied within each cluster to predict TDD (time to degree) completion. One of the clusters got 0.87 as its prediction co-efficient. Subsequently this study revealed its limitations as its focus was confined only to a single curriculum. The diversity in its application to other fields was curbed. There was also the factor of K-means clustering's sensitivity to initial conditions, which questioned the mode's robustness and interpretability. [5]
A Tarika et al., in their 2021 study explored the data forecasting domain in academia using the machine learning techniques. Their study aimed to forecast Morocco's high school students' performance during the pandemic. The models used were Linear regression, Decision trees and RF (random forest) that were trained and tested to gain highest accuracy in predicting grades in core subjects only. However, the setbacks revealed themselves as their approach wasn't general. It was specific to a particular geographic region and its school students' datasets. Additionally, only the marks were the primary factors considered for making predictions. This led to proof of how their study overlooked a lot of important factors which could have proved their approach to be effective. [6]
In 2010, O. O. J. et al. implemented K-means clustering algorithm to group students based on their academic performance. The research uses data from a Nigerian institution. The model separates the students into multiple performance groups, which helps in better academic planning and also decision-making. Clustering models provide important information to the educators by highlighting the distribution of student performance across clusters. The research only relied on GPA for clustering and oversimplifies the student performance. It also does not consider the socio-economic backgrounds that play a major part in influencing academic success [7].
In 2022, M Yağcı et al. forecasted under-graduate students' final exams by applying machine learning and statistical methods on midterm grades. This study's goal was to establish a framework for analytics and early intervention systems. The study evaluates the combination of the ensemble model: RF, deep learning model: Neural networks and supervised learning model: Support Vector Machines algorithms. The result showcased

that RF along with Neural Networks gave the highest accuracy, highlighted the significance of midterm grades and department information forecasting performance. Again, this study uses only midterm marks and performance to forecast final performance, ignoring other socio-economic demographics [8].

The Decision tree model produced accuracy of 94.44% which is the highest, while predicting student performance based on data collected from assessments, class tests, midterm and final exams. H.M. Rafi Hasan et al., 2019 study presents Decision tree that identifies key predictors for final prediction such as midterm and final exams. The study neglects external factors that have influence on the performance of the student in an eagle's eye perspective [9].

The paper by Fidelia A. Orji et.al., 2022 considered some extremely important but easily overlooked factors that affect performance of students such as intrinsic/extrinsic motivation, self-confidence etc,. The study involved about 1000 students and their motivational attributes. The research used the Random Forest model and it resulted in 95% accuracy. While motivational attributes are important, they alone cannot determine the future performances of students [10].

The research conducted by S. Li et.al., 2021 included an enormous database containing 83,000 students' performance data spanning over a decade from 2007 to 2019. The study was conducted by combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) to predict student performance. The major weakness involved is the old data that may be highly biased due to changes observed in the education system and processes [11].

Another study conducted by M. H. Rahman et.al., 2017 considered many parameters to predict student performance. It used a combination of Naïve Bayes, KNN, Decision Tree, and ANN. The method resulted in 84.3% accuracy. However, the paper lacks universality because the study involved a very small database of 480 students. The paper also lacked proper analysis of the model interpretability thus failing to explain how different parameters affected student performance[12].

Research conducted by M. R. Rimadana, et.al., 2019 focused on the role of time management and preparation time. The study took into consideration of combination of five models: SVM (Support Vector Machine), Decision tree, and Random Forest. The SVM model had accuracy of 80% which was highest score. However, since the study only involved 125 students marks it introduces bias into the study. It also overlooks other vital factors and overfocuses on time management [13].

Realinho et al. (2022) conducted an experiment to investigate student dropout. The study involved adoption of ML models on 4,424 students' data. This paper also took socio-economic parameters into consideration. The study however, did not look into the behavioural factors like attendance and motivation. The accuracy of the study too failed to be significant [14].

Table 1: Comparative Study of various models.

Year	Reference	Technique used	Results
2017	[21]	Students were categorized into performance clusters.	Clusters: "High" (46.67%), "Medium" (45%), and "Low" (8.33%) based on GPA intervals (2.20–4.0).
2017	[12]	Behavioural features and attendance were key predictors of performance, with ensemble filtering using ANN yielding the best results, highlighting their strong influence.	Naïve Bayes: 74.0% K-Nearest Neighbour: 74.3% Decision Tree: 77.5% Artificial Neural Network (ANN) with Ensemble Filtering: 84.3%
2019	[9]	Decision Tree Classifier, K Nearest Neighbour (KNN)	Decision Tree Classifier: 94.44%, K-Nearest Neighbour (KNN): 89.74%
2019	[13]	Linear Support Vector Machine, Neural Network, Naive Bayes, RF (random forest), Decision tree	Linear SVM: 80% accuracy, Decision Tree: 68% accuracy, Random Forest: 77.8% accuracy, Neural Network: 75.2% accuracy, Naive Bayes: 73.6% accuracy
2020	[16]	Improved Random Forest	Random Forest: Accuracy - 93%, True Positive Rate (TPR) - 0.96, False Positive Rate (FPR) - 0.031, Precision - 87.2%, Recall - 96%, F1-Score - 0.93
2022	[15]	Random Forest, MLP (multi-layer perceptron), Logistic Regression, Naïve Bayes, SVM (support vector machine)	Random Forest: Accuracy - 97.4%, Multilayer Perceptron (MLP): Accuracy - 95.7%, Support Vector Machine (SVM): Accuracy - 94.8%, Logistic Regression: Accuracy - 92.1%, Naïve Bayes: Accuracy - 82.6%

Comparative Analysis

The web application applies data analytics and machine learning to analyse student performance and determine at-risk students based on socio-economic and academic indicators. It analyses datasets with prominent features like parental levels of education, their monthly income, race or ethnicity group, academic scores in writing, reading, mathematics, lunch scheme (standard / reduced), test preparation level. These indicators are significant as they are the general influences that impact student outcomes. Academic scores in core subjects are direct measures of student proficiency and progress, while demographic indicators assist in understanding inequality and optimizing interventions. Through preprocessing of the input dataset, the system performs data analysis and provides in-depth insights, which will assist educators in identifying students who are about to fail in their exams. This will allow institutions to provide targeted academic support and implement data-driven initiatives to improve student performance and retention. Refer to Fig 1: Algometrics Flowchart.

The web platform makes use of data analytics and machine learning techniques to examine student performance to identify at-risk individuals by evaluating both socio-economic and academic indicators. The variables decided to be used are essential for two main reasons. First, of them being the general influence that impacts student outcomes. Second, they give a comprehensive understanding of the other factors influencing student success. Core subjects-related performance gives a direct measure of student aptitude, while demographic indicators highlight potential inequalities or hindrances to an individual's capabilities. This enables educators to intervene and fill in the appropriate needs. Through meticulous preprocessing of the input (student dataset), Algometrics performs data analysis

and generates actionable insights, assisting educators in recognizing students and providing them with timely support. This web tool is a proactive academic support that implements data-driven initiatives using evidence-based strategies to enhance retention rates and overall academic outcomes of every student.

The application predicts the students' immediate term marks using the socio-economic demographics and the marks of the students from earlier exams conducted. Since the prediction of marks is given as numeric values, regression models were opted over classification models. From the literature review conducted, it was found that all the studies conducted leveraged the power of a single model. To cover this gap, a multi-model approach was followed to build the application. The models considered to build the application are as follows:

Based on current term marks scored by students, along with their socio-economic and demographic factors, the marks of the immediate next term are predicted by this application. Regression models were finalised over classification models as they produce predictions in the form of numerical values. A thorough literature review of existing works and a comparative study of previously used models (refer to Table 1: Comparative Study of various models) revealed that there is heavy reliance on use of single model for such academic predictions. Addressing this limitation became crucial, which led to the adoption of a multi-model strategy in developing the application. To bridge this gap, the following combinations of models were considered. Mean absolute error makes it easy to interpret errors, mean squared error is useful for penalizing significant deviations, and the Coefficient of determination evaluates model performance by indicating how closely the predictions fit the actual data.

Linear Regression + Random Forest

Linear Regression, when compared based on MAE, R2 score and MSE, outperformed RF (random forest) in terms of predicting accuracy across the entire evaluation metrics. It had a lower MAE (3.317 vs. 4.835), higher R2 score (0.933 vs. 0.853), and covered more variance in data, making it a better fit. Additionally, the MSE score of Linear regression (17.666 vs. 38.635) confirmed that it has fewer large errors. The major achievement was when both the models performances were averaged using the ensemble approach as there was significantly better scores. MAE reduced to 1.804, R2 increased to 0.977, MSE lowered to 5.218, showcasing the advantage of combining two models.

Random Forest + RNN

The next pair of models chosen was combination of RNN with Random Forest, where Random Forest was outperformed yet again. RNN had extremely low MAE (0.47) and MSE (0.00349), compared to RF's MAE (3.718) and MSE (22.027). Their R2 scores, however are close (0.9218 for RNN vs. 0.9167 for RF). After averaging these models, R2 score improved to 0.9277 and MSE reduced to 19.129, indicating better overall performance in making predictions.

RNN + LSTM

RNN and LSTM, being the last pair of models chosen, performed better when averaged (R2 score - 0.885, MSE score - 0.00512, 0.0058 - MAE score). Although LSTM achieves slightly lower MAE (0.044), higher R2 score (0.9308), which is better than RNN's MAE and R2 Score (0.085 and 0.7751 respectively). This evidently proves that averaging or combining models helps in reducing errors, in turn improving prediction accuracy.

Linear Regression and Random Forest combination yielding the highest performance, indicated a strong predictive capability. This model pair achieved an overall R^2 score of 0.97, and the R^2 score serves as a key evaluation metric, measuring the proportion of variance in student marks. A model's effectiveness in capturing the underlying patterns in student performance is proven by its value's closeness to 1. So, the chosen pair proves the robustness in accurately forecasting academic outcomes. Refer to Table 2: Research Analysis of three pairs of models for an in-depth insight into the performance metrics of each pair of models.

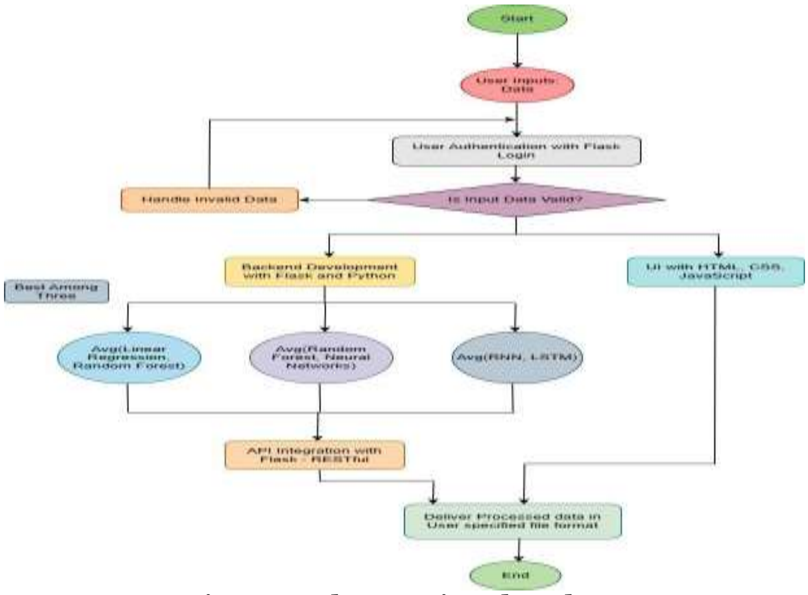


Figure 1. Algometrics Flowchart

Table 2: Research Analysis of three pairs of models

Model	MSE	R2 Score	MSE
Random Forest Linear Regression Average of both the models	4.835 3.31719 1.80485	0.85392 3.31719 0.97748	38.635 17.66633 5.21817
RNN Random Forest Average of both the models	0.04757 3.71807 3.45057	0.92180 0.91670 0.92770	0.00349 22.02790 19.12953
LSTM RNN Average of both the models	0.04404 0.08573 0.05845	0.93080 0.77510 0.88521	0.00309 0.01003 0.00512

Methodology

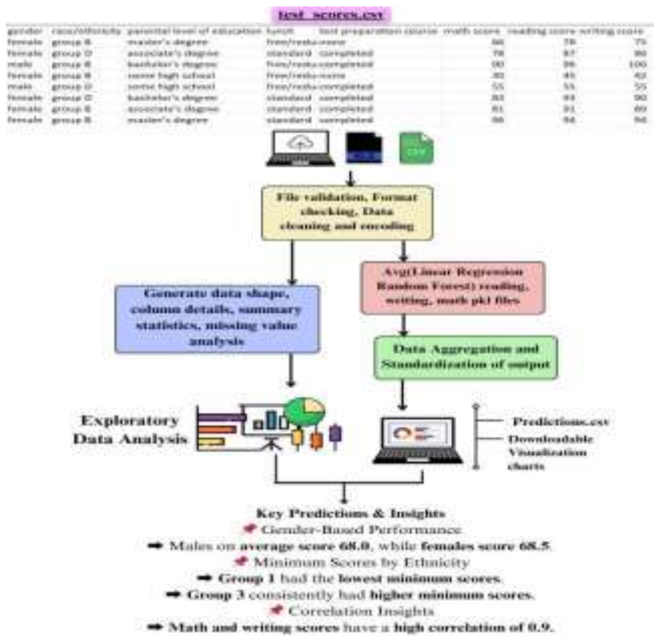


Figure 2. Methodology

Dataset

The system uses a dataset consisting of diverse features to train the model. The various attributes include parental level of education, lunch scheme, test preparation level, race, gender and academic scores in math, reading and writing. These features serve as critical indicators to analyze students and identify at-risk individuals.

Input

The web application accepts datasets provided by the user. The file submission interface accepts data in the .csv and .xls file formats. This data then undergoes preprocessing before being analyzed. The following code is used to handle the file inputs:

```
def upload():
if request.method == 'POST':
file = request.files['fileInput'] if file:
filepath
=os.path.join(app.config['UPLOAD_FOLDER'], file.filename)
file.save(filepath)
flash('File successfully uploaded!', 'success')
```

Preprocessing

After the dataset has been uploaded by the user, a complete Exploratory Data Analysis (EDA) is conducted on the dataset. The EDA contains basic details like dataset shape, missing values and summary statistics. The dropna() function from the pandas library is used to remove missing values from a DataFrame or Series. Thereby cleaning datasets by dropping rows that contain null values.

```
insights = insights.dropna()
```

Data Encoding

The python library *scikit-learn is imported in the script and LabelEncoder specifically for* mapping categorical values to predefined integers. This is done to maintain simplicity during computation and to avoid issues when categorical variables do not have an ordinal relationship. The following columns are encoded before training the model: gender, race/ethnicity, parent's level of education, lunch, test preparation course.

Examples:
Gender: Male => 0, Female => 1
Level of Parental Education: High School => 0,
Associate's Degree => 1, Bachelor's Degree => 2

Model Training

The heatmap was generated to identify correlation patterns among student performance variables. Students who excel in one subject, like mathematics, tend to score good in reading and writing subjects as well, indicated by positive correlations (0.82, 0.81, 0.95). Negative or weak correlations arise in factors like gender, race/ethnicity, parental education and lunch schemes. This strongly suggests minimal to no impact of those factors. Gender and reading scores have negative correlations (-0.19) and writing scores (-0.25), while parental education and lunch status show slight negative correlations with all three scores. This shows that socio-economic factors have a high influence on performance. Highly correlated variables serve as strong predictors of academic success, while negative correlations or inverse relationships identify the potential factors that might be a risk to performance outcomes.

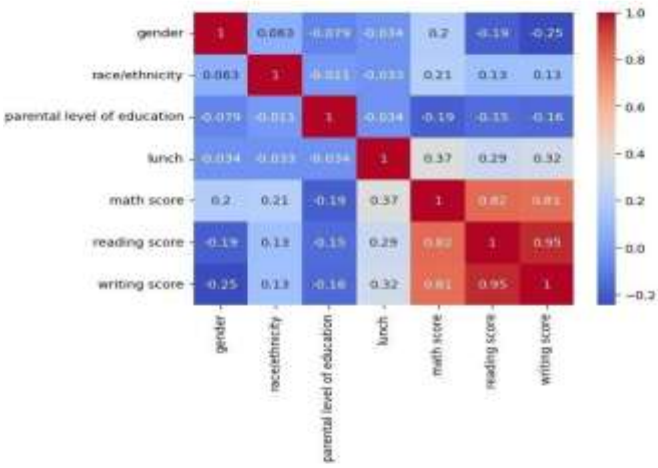


Figure 3. Heatmap

Visualizations
The performance trends and other analytics are displayed through an interactive dashboard. The dashboard comprises five distinct charts that incorporate academic and socio-economic parameters. Using these analytics, faculty can identify key trends and patterns, facilitating data-driven decision making. Therefore helping institutions in allocating resources and also paying attention to at-risk students and also ensuring timely intervention and support. The following charts are included in the dashboard.

a) Average Score by Gender (Pie Chart)
The average score by gender pie chart analyses the mean of math, reading and writing scores by gender. The chart gives a proportionate indication of mean scores and a quick overview of the gender gap in performance.

The gender-based performance analysis helps in understanding areas of imbalance in performance. Teachers can then design focussed programs to bring balance in performance.

b) Parental Education Standard Deviation of Scores (Line Chart)
The line chart in the dashboard portrays the standard deviation of marks obtained in math, reading and writing by parental education levels. It announces the spread in the performance of the students and quantifies the extent of variation in the data.
This chart gives insights into the academic homogeneity. Greater variability indicates uneven academic achievement according to parental level of education. Therefore, it is essential to detect such trends to then establish specialized academic support systems.

c) Minimum Scores by Race/Ethnicity (Bar Chart)
Minimum Scores by Race/Ethnicity depicts the lowest scores in math, reading and writing scores by race/ethnicity. It indicates the lowest-performing groups. This enables institutions to support demographic groups that require immediate attention.

Analyzing data based on minimum scores by ethnicities helps in identifying underperforming minority groups. Hence the institutions can focus more of their resources to improve the performance of these groups. It also helps in creating culturally responsive learning plans that cater to the diverse needs of students.

d) Average Math Score by Test Preparation Course (Donut Chart)
The effectiveness of test preparation classes on mean mathematics scores is shown by the doughnut chart. It signifies the differences in scores between students who enrolled in a test preparation course and students who did not.
This case demonstrates the efficacy of the test preparation program in schools. Also highlights to the institution to recommendation of these courses to at-risk students.

e) Correlation Heatmap (Heatmap Matrix)
The heatmap represents the correlation coefficients of all the demographic and academic variables. It portrays the strength and nature of relationships between the variables.
Heatmaps are highly essential for determining correlations among different variables. Strong math and writing score correlations, for example, may indicate that an improvement in one subject will directly improve all other subjects. Such data allow institutions and educators to adopt special interdisciplinary learning methods.

Predictions
The prediction step uses pre-trained models to determine students' marks in all three subjects (mathematics, reading, writing). The models involved in this step are Linear Regression and Random Forest. The predictions of the models are then averaged to obtain better accuracy. The predicted marks columns are concatenated to the original dataset. The faculty can easily download this as an excel file and inspect the results.

```
final_pred = pd.DataFrame(pred)
final_pred = pd.concat([original_data, final_pred], axis=1)
final_pred.to_csv("static/Predictions.csv", index=False) return final_pred
```

Report Generation
Along with the dashboard, the application further provides a summary report that includes key insights, trends and patterns observed in the analysis. This helps the institution
gain valuable knowledge and make better informed decisions with respect to their resource allocation and in helping their students perform better. They enforce data-driven syllabus and curriculum practices and processes.

Experimental Results

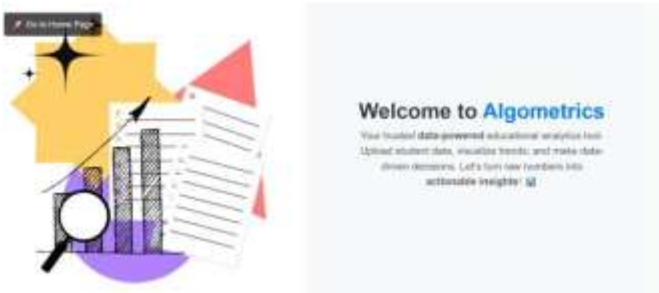


Figure 4. Landing Page

In the above figure, landing page gives a short overview of Algometrics. It gives information about the system's purpose and core functionalities.

Figure 5. Algometrics Features

The Features section of Algometrics (refer Fig 5) provides details about the current features and functionalities that are available for the user to try out. It also informs the users about the various data visualizations and personalized insights provided.

Figure 6. About Us

The About Us section of the website (refer Fig 6) talks about the vision and mission and the team members behind building Algometrics and highlighting its importance in enhancing student success.

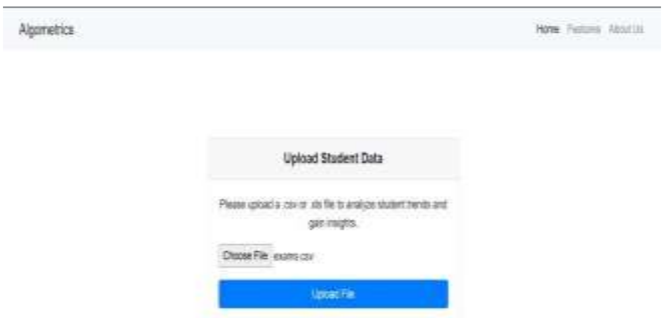


Figure 7. Uploading Dataset

Interface for users to upload their student performance data (in .csv or .xls format), which will be analysed by the system. Refer Fig 7: Uploading Dataset.

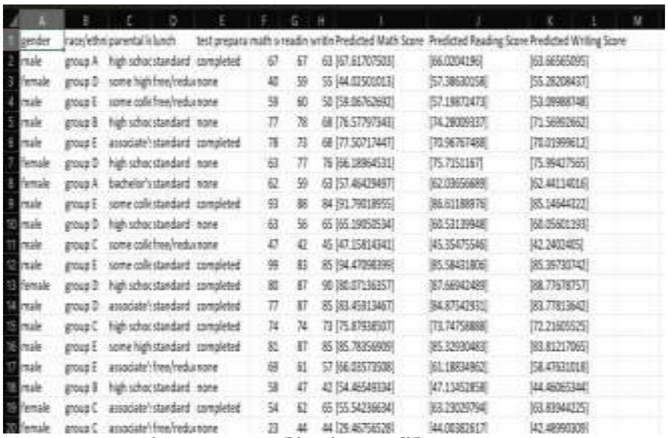
Figure 8. Exploratory Data Analysis

The dashboard is explained in this section and key insights derived from the dashboard are showcased here (refer Fig 8).

Figure 9. Algometrics Dashboard

In Fig 9, dashboard contains multiple interactive and downloadable charts that display the real-time performance metrics, trends and predictions made.

Algometrics Dashboard provides a complete interactive interface displaying visualizations like pie charts, bar graphs and heatmaps. Highlighting average scores, performance variations by gender and parental education and many more, the dashboard reflects the entire correlations of all the factors affecting the student's performance. Thus enabling teachers to identify at-risk students and implement timely interventions. The dashboard on a whole helps in monitoring academic trends effectively and enhance student success.



	A	B	C	D	E	F	G	H	I	J	K
1	gender	race/ethni	parental le	lunch	test prepara	math s	readin	writin	Predicted Math Score	Predicted Reading Score	Predicted Writing Score
2	male	group A	high schoc	standard	completed	67	67	63	[67.61707503]	[66.0204196]	[63.66565095]
3	female	group D	some high	free/redu	none	40	59	55	[44.02501013]	[57.38630158]	[55.28208437]
4	male	group E	some colle	free/redu	none	59	60	50	[58.06762692]	[57.18872473]	[53.09988748]
5	male	group B	high schoc	standard	none	77	78	68	[76.57797343]	[74.28009337]	[71.56992662]
6	male	group E	associate'	standard	completed	78	73	68	[77.50717447]	[70.96767488]	[70.01999612]
7	female	group D	high schoc	standard	none	63	77	76	[66.18964531]	[75.7151167]	[75.99427565]
8	female	group A	bachelor's	standard	none	62	59	63	[57.46429497]	[62.03656689]	[62.44114016]
9	male	group E	some colle	standard	completed	93	88	84	[91.79018955]	[86.61188976]	[85.14644322]
10	male	group D	high schoc	standard	none	63	56	65	[65.19050534]	[60.53139948]	[68.05601393]
11	male	group C	some colle	free/redu	none	47	42	45	[47.15814341]	[45.35475546]	[42.2402405]
12	male	group E	some colle	standard	completed	99	83	85	[94.47098399]	[85.58431806]	[85.39730742]
13	female	group D	high schoc	standard	completed	80	87	90	[80.07136357]	[87.66942489]	[88.77678757]
14	male	group D	associate'	standard	completed	77	87	85	[83.45913467]	[84.87542931]	[83.77813642]
15	male	group C	high schoc	standard	completed	74	74	73	[75.87938507]	[73.74758888]	[72.21605525]
16	male	group E	some high	standard	completed	81	87	85	[85.78356909]	[85.32930483]	[83.81217065]
17	male	group E	associate'	free/redu	none	69	61	57	[66.03573508]	[61.18834962]	[58.47631018]
18	male	group B	high schoc	standard	none	58	47	42	[54.46549334]	[47.13452858]	[44.46065344]
19	female	group C	associate'	standard	completed	54	62	65	[55.54236634]	[63.23029794]	[63.83944225]
20	female	group C	associate'	free/redu	none	23	44	44	[29.46756528]	[44.00382617]	[42.48990309]

Fig 10: Predictions file output

Algometrics produces as one of its results a downloadable CSV file (refer Fig 10) containing model-generated predictions. It, after utilizing ensemble technique that

combines Linear Regression and Random Forest models, and achieving accuracy, forecasts students' performance across various academic areas. This methodology significantly enhances prediction precision in subjects like math, reading and writing simultaneously mitigating overfitting and errors, ultimately enhancing overall dependability on the model.

Altogether, Algometrics is a tool having capability to conduct not just Exploratory Data Analysis (EDA), that provides intricate insights into distribution of data and highlights key factors such as gender, parental education, and test preparation efforts that influence student outcomes. This platform also provides downloadable comprehensive visualizations such as pie charts, bar graphs, and correlation heatmaps, uncovering all the critical trends and relationships between critical variables. With the Predictions.csv accessible, this web tool ultimately equipping educators to make carefully informed, data-backed decisions for the betterment of their students' academic future.

Conclusion

This study looks at the factors that influence students' performance such as family background, preparatory efforts and socio-economic background. Even if gender impact in performance is minimal, differences that are noticeable are more based on parental education and ethnicity suggesting broader socio-economic conditions. Test preparation courses contribute to improvement in scores, though they are not the only factor that decide success. These findings' insights can help educators identify at-risk students, in-turn helping institutions to allocate their resources (teachers, teaching hours, classrooms, labs) more effectively and implement targeted measures to improve overall academic achievements.

Future Scope

Further research can aim to concentrate more on incorporating factors such as study routines and participation patterns to enhance prediction accuracy. Inclusion of a much diverse range of students from various educational backgrounds would help model's ability to adapt across different institutions. Additionally, real-time data analytics and customized learning techniques could help in more personalization of support for specific set of students. By continuously refining the predictive models and leveraging advanced statistical data-model techniques, this system can serve as a tool for educators to enhance student learning experiences and guarantee better success rates of students.

References

1. J. Willis, K. Krausen, R. Caparas, and T. Taylor, "Resource allocation strategies to support the Four Domains for Rapid School Improvement," 2019. Available: https://files.eric.ed.gov/fulltext/ED602981.pdf
2. Organisation for Economic Co-operation and Development (OECD), Class size and student-teacher ratios," OECD. Available: https://www.oecd.org/en/topics/class-size-and-student-teacher-ratios.h tml
3. K. Diaz, C. Fetti, G. Torres-Garda, and N. M. Crisosto, "The Effects of Student-Teacher Ratio and Interactions on Student/Teacher Performance in High School Scenarios," Mathematical and Theoretical Biology Institute, Los Alamos National Laboratory, Aug. 2003. Available: https://mcmsc.asu.edu/sites/default/files/2024-09/MTBI%202003%20 Student-Teacher%20Ratio%20Report.pdf
4. Hanover Research, "Early Alert Systems in Higher Education," Hanover Research, Nov. 2014. Available: https://hanoverresearch.com/wp-content/uploads/2017/08/Early-Alert-Systems-in-Higher-Education.pdf
5. O. Iatrellis, I. K. Savvas, P. Fitsilis, and V. C. Gerogiannis, "A Two-phase Machine Learning Approach for Predicting Student Outcomes," Int. J. Educ. Data Min., vol. 8, no. 2, pp. 112-120, 2021. Available: https://doi.org/10.1016/j.jedm.2021.02.001
6. A. Tarika, H. Aissab, and F. Yousefc, "Artificial Intelligence and Machine Learning to Predict Student Performance during the COVID-19," Data Science, 2021, pp. 123-129.
7. O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, "Application of k-Means Clustering Algorithm for Prediction of Students' Academic Performance," Int. J. Comput. Sci. Eng., vol. 8, no. 2, pp. 140-145, 2010. Available: https://www.idss.iocspublisher.org/index.php/jidss/article/view/109
8. M. Yağcı, "Educational Data Mining: Prediction of Students' Academic Performance Using Machine Learning Algorithms," J. Educ. Data Min., vol. 22, no. 4, pp. 345-350, 2022. Available: https://www.academia.edu/ 99037405/Decision_Analytics_Journal?uc-sb-sw=63617647
9. H. M. Rafi Hasan, A. K. M. Shahariar Azad Rabby, M. T. Islam, and S. A. Hossain, "Machine Learning Algorithm for Student's Performance Prediction," Int. J. Mach. Learn. Appl., vol. 7, no. 2, pp. 100-110, 2019. Available: https://doi.org/10.1016/j.ijmla.2019.03.004
10. F. A. Orji and J. Vassileva, "Machine Learning Approach for Predicting Students' Academic Performance and Study Strategies Based on Their Motivation," J. Comput. Educ., vol. 6, no. 1, pp. 31-47, 2022. Available: https://doi.org/10.1007/s40692-022-00212-6
11. S. Li and T. Liu, "Performance Prediction for Higher Education Students Using Deep Learning," Educ. Data Mining, vol. 28, no. 5, pp. 112-118, 2021. Available: https://doi.org/10.1016/j.edm.2021.01.009
12. M. H. Rahman and M. R. Islam, "Predict Student's Academic Performance and Evaluate the Impact of Different Attributes on the Performance Using Data Mining Techniques," J. Educ. Technol., vol. 34, no. 6, pp. 212-218, 2017. Available: https://doi.org/10.1016/j.jest.2017.01.003
13. M. R. Rimadana, S. S. Kusumawardani, M. S. Ferdinand, and P. I. Santosa, "Predicting Student's Academic Performance Using Machine Learning and Time Management Skill Data," J. Educ. Perform. Predict., vol. 12, no. 4, pp. 67-73, 2019. Available: https://doi.org/10.1016/j.jep.2019.06.001

14. V. Realinho, J. Machado, L. Baptista, and M. V. Martins, "Predicting Student Dropout and Academic Success," 2022. Available: https://www.mdpi.com/2306-5729/7/11/146
15. G. B. Brahim, "Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features," 2022. Available: https://link.springer.com/article/10.1007/s13369-021-06548-w
16. S. Jayaprakash, S. Krishnan, and J. V., "Predicting Students' Academic Performance Using an Improved Random Forest Classifier," 2020. Available: https://scholar.google.co.in/scholar?q=S.+Jayaprakash,+S.+Krishnan,+and+J.+V.,+%22Predicting+Students%27+Academic+Performance+Using+an+Improved+Random+Forest+Classifier,%22+2020&hl=en &as_sdt=0&as_vis=1&oi=scholart
17. A. S. Aljaloud, D. M. Uliyan, A. Alkhalil, M. A. Elrhman, A. F. M. Alogali, Y. M. Altameemi, M. Altamimi, and P. Kwan, "A Deep Learning Model to Predict Student Learning Outcomes in LMS Using CNN and LSTM," 2022. Available: https://colab.ws/articles/10.1109%2Faccess.2022.3196784
18. L. M. A. Zohair, "Prediction of Student's Performance by Modelling Small Dataset Size," 2019. Available: https://educationaltechnologyjournal.springeropen.com/articles/10.11 86/s41239-019-0160-3
19. B. Sravani and M. M. Bala, "Prediction of Student Performance Using Linear Regression," 2020. Available: https://www.researchgate.net/publication/343405696_Prediction_of_Student_Performance_Using_Linear_Regression
20. L. Al-Alawi, J. Al Shaqsi, A. Tarhini, and A. S. Al-Busaidi, "Using Machine Learning to Predict Factors Affecting Academic Performance – The Case of College Students on Academic Probation," 2023. Available: https://link.springer.com/article/10.1007/s10639-023-11700-0
21. S. P. Borgavakar and A. Shrivastava, "Evaluating Student's Performance Using K-Means Clustering," 2017. Available: https://pubs.aip.org/aip/acp/article/3175/1/020085/3339042/Performa nce-prediction-of-students-data-by-using-k
22. Y. Baashar, G. Alkawsi, A. Mustafa, A. A. Alkahtani, Y. A. Alsariera, A. Q. Ali, W. Hashim, and S. K. Tiong, "Toward Predicting Student's Academic Performance Using Artificial Neural Networks (ANNs)," 2022. Available: https://www.mdpi.com/2076-3417/12/3/1289