Suleiman *et al.*
Journal of Electrical Systems and Inf Technol (2024) 11:41
https://doi.org/10.1186/s43067-024-00166-w

Journal of Electrical Systems and Information Technology

RESEARCH Open Access

Key factors influencing students' academic performance

Ibrahim Baba Suleiman[1*], Oluwasogo Adekunle Okunade[1], Emmanuel Gbenga Dada[2] and Uchenna Christiana Ezeanya[1]

*Correspondence:
ibsuleiman123@gmail.com

[1] Department of Computer Science, Faculty of Sciences, National Open University of Nigeria, Abuja, Nigeria
[2] Department of Computer Science, Faculty of Physical Sciences, University of Maiduguri, Maiduguri, Nigeria

Abstract

Academic achievement is a multifaceted outcome influenced by a multitude of factors spanning across educational, socioeconomic, and individual characteristics. Understanding the key determinants of students' academic performance is paramount for educators, policymakers, and institutions to enhance learning outcomes and facilitate targeted interventions. The absence of a cohesive framework for understanding academic achievement poses a barrier to effective interventions and policies, hindering educators and policymakers from implementing targeted strategies for student improvement. To address this challenge, it is crucial to conduct a thorough examination of the factors influencing academic performance, prioritizing their identification, exploring interactions, and examining their relative impact. Through this research, educational stakeholders can gain insights to develop evidence-based strategies that foster equitable learning outcomes and pave the way for academic success for all students. In this research, linear regression is chosen as the primary model due to its ability to model relationships between variables, making it suitable for assessing the impact of multiple factors on academic performance. Simulation results show that the proposed model achieved 98% accuracy in predicting the factors that are affecting students' academic performance. The findings derived from the linear regression analysis offer statistical insights into the key factors influencing students' academic performance. It also incorporates visual aids to enhance the clarity and interpretability of the results. This research interprets the implications of these key factors and discusses their broader significance in the context of students' academic performance.

Keywords: Data-driven education, Academic performance, Predictive modeling, Linear regression, Grade prediction, Student assessment

Introduction

The academic performance of students is a central concern in the field of education, with profound implications for their future success and societal well-being. Over the years, numerous studies have explored the factors that influence students' academic achievement, recognizing that it is a complex outcome shaped by a multitude of variables. Understanding these key determinants is essential for educators, policymakers, and institutions to develop effective strategies that can enhance learning outcomes, address disparities, and support students in reaching their full potential.

This journal topic examines the various factors that play a pivotal role in shaping students' academic performance. These factors encompass a wide spectrum, ranging from individual characteristics and behaviors to socioeconomic conditions, familial influences, educational settings, and broader societal contexts. By scrutinizing these elements, we aim to gain a comprehensive understanding of their intricate interactions and relative significance in determining academic success.

This collection of research articles and studies seeks to shed light on the multifaceted nature of academic achievement. It offers a platform for scholars, educators, and researchers to present their findings, share insights, and contribute to the ongoing discourse on the subject. By fostering a deeper understanding of the key factors influencing students' academic performance, this journal topic ultimately aspires to inform evidence-based practices and policies that can optimize educational outcomes for students across diverse backgrounds and contexts.

Academic performance among students is a critical metric for assessing the effectiveness of educational systems worldwide. However, the substantial variation in academic outcomes among students has raised a fundamental question: What are the key factors that significantly influence students' academic performance? This question represents a pressing problem within the realm of education as it highlights the need to understand, identify, and address the underlying determinants of academic success.

Despite extensive research in this area, a comprehensive and conclusive understanding of the key factors influencing students' academic performance remains elusive. The persistence of academic disparities among students from diverse backgrounds and educational settings underscores the complexity of the issue. Moreover, traditional methods of assessment, such as standardized testing and teacher evaluations, often provide limited insights and may not capture the full range of factors at play.

The lack of a clear and unified framework for understanding the multifaceted nature of academic achievement hinders the development of effective interventions and policies aimed at improving student outcomes. Educators, policymakers, and institutions face the challenge of navigating this intricate landscape without a well-defined roadmap, making it difficult to implement targeted strategies that address the unique needs of individual students.

To bridge this gap, it is imperative to conduct a comprehensive examination of the key factors influencing students' academic performance. This research seeks to identify and prioritize these factors, explore their interactions, and determine their relative impact. By doing so, we can equip educational stakeholders with the knowledge and insights needed to develop evidence-based strategies that can enhance learning outcomes, reduce disparities, and ultimately pave the way for academic success for all students.

Related works

Academic performance is a multifaceted aspect of a student's educational journey, influenced by various factors that come into play during their time in school. Among these factors, study hours, previous scores, past questions studies, and extracurricular activities stand out as crucial determinants of success in academia. In this study, we will look into the significance of each of these elements and how they collectively contribute to shaping a student's academic trajectory.

Study hours

The amount of time a student dedicates to studying plays a pivotal role in their academic performance. Effective time management and consistent study routines can significantly impact a student's ability to grasp and retain knowledge. Balancing study hours with other commitments is essential, as excessive or insufficient study time can have adverse effects.

A study published in 2019 by Traub et al. [1] delved into the connection between study hours and academic performance among college students. Their findings revealed a positive correlation, as students who invested more time in focused and structured studying tended to achieve higher GPA scores. Importantly, this relationship persisted even after adjusting for variables like prior academic performance and socioeconomic status. The study underscored the significance of study habits in shaping academic success.

In their study, Squires and Coates analyzed the relationship between study time and exam performance, taking into account the effects of age, gender, program of study, and ability level. They found that, even after controlling for these variables, students who spent more time studying tended to have higher exam scores. This finding suggests that study time is a significant factor in academic success, above and beyond the effects of demographic and ability-related differences. In other words, study time can be a key predictor of success, regardless of other individual characteristics [2].

Also, previous studies highlight the impact of study hours on academic performance. According to research, "increased study hours are positively correlated with higher academic performance indices, suggesting that dedicated study time significantly enhances students' learning outcomes." These studies examined the relationship between study hours and academic performance. Here are some key findings from the studies:

1 *Positive Correlation* There is a strong positive correlation between the number of hours students dedicated to studying and their academic performance indices. This means that, generally, students who spent more time studying tended to achieve higher grades [3].
2 *Optimal Study Hours* While more study hours were beneficial, study also identified an optimal range. Students who studied between 15 and 20 hours per week showed the most significant improvement in their performance. Beyond this range, the benefits plateaued, and excessive study hours could lead to burnout and decreased performance [4].
3 *Quality Over Quantity* Research also emphasizes that the quality of study time is crucial. Effective study techniques, such as active learning, spaced repetition, and regular breaks, were found to enhance the benefits of study hours [3].
4 *Individual Differences* Another factor is individual differences. Differences such as learning styles and personal circumstances play a significant role. Personalized study plans that cater to these differences were recommended for optimal results [4].

The results of the current meta-analysis showed that study time and academic performance are related. The association between study time and academic performance was significant for each of the three measures of academic performance and for both within-person and between-person study time data, indicating that the study time–academic

performance relation was reliable and generalizable [5]. This further supports the idea that study time is a key factor in academic success.

These recent studies collectively show the intricate relationship between study hours and academic performance. While dedicating substantial time to focused and structured studying can lead to improved academic outcomes, it is crucial to consider the potential implications for mental health and wellbeing. Furthermore, the timing of study hours and the importance of self-regulation emerge as vital considerations in understanding and optimizing academic performance. As education continues to evolve, these insights offer valuable guidance for both students and educators in their pursuit of academic excellence.

Previous scores

Student's prior academic achievements, such as test scores and grades, provide valuable insights into their strengths and weaknesses. These scores can serve as benchmarks for improvement and indicate areas that may require additional attention. A consistent upward trajectory in scores demonstrates a commitment to academic growth.

The significance of past academic achievement as a predictor of future performance has been consistently established. A study by Liu et al. [6] emphasized the strong predictive power of high school performance for college success, underscoring the continuity of academic achievement patterns.

A study by Koçak, Göksu, and Göktaş [7] reviewed multiple meta-analyses and found that previous academic performance, such as high school grades and GPA, significantly influences future academic success. They noted that these previous scores are strong predictors of students' performance in higher education.

Also, a research carried out by Kassarnig et al. [8] highlighted that previous academic performance, including high school GPA and entrance exam scores, plays a crucial role in predicting university students' success. They also found that class attendance and social network structures as well as socioeconomic status, and study habits also play crucial roles in determining academic success.

Another study by Arora et al. [9] also indicates how previous academic performance, including high school GPA and scores from nationwide university entrance exams, significantly influences college students' academic achievements. The study also mentioned other factors like gender, socioeconomic status, and study habits.

In their review in 2023, Al Husaini and Ahmad Shukor [10] discuss various factors affecting students' academic performance. One of the significant factors they highlight is the impact of students' previous scores. They found that previous academic performance is a strong predictor of future academic success. This includes high school grades, previous assessment grades, and internal assessment grades.

Another factor the study looked at was whether the course material was similar between the previous course and the current course. The study found that students who had previous courses that were similar to their current course had a stronger relationship between previous scores and academic performance. This suggests that if the material is similar, previous scores can be a good predictor of academic performance.

So, another recent study looked at the relationship between course difficulty and academic performance [11]. The study found that students who took more difficult courses

had a stronger relationship between previous scores and academic performance. This suggests that course difficulty may play a role in how previous scores influence academic performance.

Past questions studies

Reviewing past exam questions and test materials can be a strategic approach to understanding the format, content, and patterns of assessments. This practice helps students become familiar with the types of questions that may appear and refine their problem-solving skills. Past questions studies can be a valuable tool for exam preparation.

Iliya and Musa conducted a study in 2017. The findings of the study reveal that more than half of the students admitted that using past questions helps them to understand a topic or subject more easily, while three-quarters reported that past questions helped them to remember facts and ideas more easily. A great majority of the students also indicated that using past questions during tests and exams had helped them to answer the questions and perform better [12]. The results of this study suggest that past question papers can be effective in helping students to understand, remember, and apply information more effectively, which ultimately leads to improved performance on tests and exams.

One recent study, published in 2022, found that students who studied past exam questions had higher academic performance than those who did not [13]. This was true even after controlling for other factors, such as prior academic performance and intelligence. The study found that past question studying was especially beneficial for students who were anxious about exams.

Extracurricular activities

Beyond academics, participation in extracurricular activities can influence a student's overall development. Involvement in clubs, sports, or community service can foster important skills such as leadership, teamwork, and time management. These skills can, in turn, have a positive impact on academic performance by enhancing a student's ability to juggle multiple responsibilities.

A recent study conducted by Anjum Shabiha [14] highlights several importance of students' participation in extracurricular activities. The study shows that students who participate in extracurricular activities are more likely to achieve higher academic achievement. He went further to stated that;

Students participating in extracurricular activities also have more self-confidence, teacher perception, and a positive attitude toward school. Students participating in extracurricular activities are less likely to drop out and are more likely to achieve higher academic achievement. Participation in extracurricular activities also reduces absenteeism and the late arrival of students. Policies like "No Pass No Play" developed by schools encourage students to keep their grades. This prevents them from failing or dropping out. Other school rules ensure that participants avoid drug and tobacco use so that they can be disciplined and focused. Extracurricular activities also enable students to become productive learners and adults. Participating in extracurricular activities makes a student capable of skills such as leadership, teamwork, organization, analytical thinking, and problem-solving, and time management. These activities help the student learn the

skill of multitasking. This allows them to discover their talents. This kind of commitment can also help them find an interest-based job or relate their experience entering the job market to real life. After-school activities also improve pupils' social skills. Students can meet new people and work with people from different backgrounds who share common skills and interests. We must be aware of the positive impact of extracurricular activities on the educational system.

Another study published in 2021, looked specifically at the relationship between sports participation and academic performance [15]. The study found that sports participation was positively associated with academic performance, but that this association was stronger for boys than for girls. Additionally, the study found that sports participation was associated with better mental health outcomes, such as lower levels of stress and anxiety. This suggests that sports participation may have different benefits for different genders.

Academic performance is shaped by a combination of factors, including study hours, previous scores, past questions studies, and extracurricular activities. Striking a balance among these elements is essential for holistic growth and success in both academic and personal realms. By recognizing the significance of these factors and leveraging them effectively, students can enhance their overall academic performance and achieve their educational goals.

Application of linear regression in academic performance prediction

Predicting academic performance has been a subject of interest for decades, with early efforts relying on traditional assessments and teacher judgments. However, the advent of data analytics and machine learning techniques has revolutionized this field. In recent years, linear regression, a well-established statistical method, has gained prominence for its simplicity and interpretability in predicting academic outcomes.

Linear regression, a linear modeling technique, has been widely employed to predict academic performance due to its suitability for continuous outcomes. Several studies have utilized linear regression to forecast students' grades, often using a combination of demographic, behavioral, and prior academic information as predictor variables.

Linear regression analysis is a valuable tool for exploring and understanding the key factors influencing academic performance. By quantifying the relationships between these factors and academic outcomes, educators and policymakers can make data-driven decisions to enhance the educational experience and success of students.

To explore the key factors influencing academic performance, linear regression is applied and an effective predictive model is built.

Materials and methods

A detailed account of the materials, data, and the procedures used for this study is analyzed in this section. The data sources, experimental procedure/set-up, data collection and processing as well as statistical methods and tools used are all explained below.

Dataset description

The data used in this research project are derived from secondary sources. The dataset was obtained from Kaggle.com, an online platform renowned for its role in the

Table 1 Dataset description

Name	Description	Value
Hours studied	The total number of hours spent studying by each student	It has integer value
Previous scores	The scores obtained by students in previous tests	(0–100)
Extracurricular activities	Whether the student participates in extracurricular activities	(Yes or No)
Sleep hours	The average number of hours of sleep the student had per day	It has an integer value
Sample question papers practiced	The number of sample question papers the student practiced	It has integer value
Performance index	A measure of the overall performance of each student	It has integer value

Table 2 Actual dataset

S/N	Hours studied	Previous scores	Extracurricular activities	Sleep hours	Sample question papers practiced	Performance index
0	7	99	Yes	9	1	91
1	4	82	No	4	2	65
2	8	51	Yes	7	2	45
3	5	52	Yes	5	2	36
⋮	⋮	⋮	⋮	⋮	⋮	⋮
9996	7	64	Yes	8	5	58
9997	6	83	Yes	8	5	74
9998	9	97	Yes	7	0	95
9999	7	74	No	8	1	64

data science and machine learning communities. It serves as a hub for data enthusiasts, offering a wealth of datasets and hosting various data-related projects. Kaggle provides an extensive collection of resources, including datasets, code repositories, and collaborative tools, making it a valuable platform for data exploration, analysis, and the advancement of data-driven research and solutions. The study population consists of ten thousand (10,000) students enrolled in a school. The Student Performance Dataset was designed to examine the factors influencing academic student performance. Each record of the dataset contains information about various predictors and a performance index. The dataset is fully described in Table 1.

As illustrated in Table 2, the dataset consists of ten thousand (10,000) students enrolled in a school including their corresponding values for study hours and other activities.

Experimental settings

In this section, we elucidate the experimental settings that underpin our study on predicting students' academic performance through the application of linear regression. These settings encompass critical elements that define the environment and conditions in which our research was conducted:

Data source

The data used in this research project are derived from secondary sources. The dataset was obtained from Kaggle.com, an online platform renowned for its role in the data science and machine learning communities. It serves as a hub for data enthusiasts, offering a wealth of datasets and hosting various data-related projects. It encompasses a comprehensive data on students, including their academic records, and study habits. This dataset serves as the foundation for our predictive modeling.

Data preprocessing

Before applying the linear regression model, we rigorously preprocessed the dataset to ensure data quality and integrity. These preprocessing steps are implemented using open-source data mining software like Jupyter Notebook, which provides an interactive environment for coding and visualizing data. Jupyter Notebook supports various programming languages like Python and R, which have extensive libraries for data preprocessing and analysis, such as Pandas, Scikit-learn, and NumPy.

During the preprocessing, the ordinal data(Extracurricular Activities) are nominalized using numeric scales, that is, the values in column "Extracurricular Activities" changed to numerical values. This is to enable the program to run successfully during implementation.

Feature selection

Feature selection was a crucial aspect of our experimental settings. We carefully selected a subset of independent variables (predictors) from the dataset, considering their anticipated influence on academic performance. These features encompassed prior academic achievements, as well as study habits.

Training and testing data split

To evaluate the model's predictive performance, we divided the dataset into two distinct sets: the training set and the testing set. The training set was used to train the linear regression model, while the testing set was kept separate for model validation and performance assessment. Twenty (20) percent of the data was used for testing the model, and the remaining eighty (80) percent was used for training the model.

The independent variables are: Hours Studied, Previous Score, Extracurricular activities, Sleep Hour, and Sample Question Paper Practiced, and the dependent variable (target variable) is Performance Index.

The independent variables were selected because each one is relevant and has a strong correlation with the target variable (Performance Index). Thus, it will ensure a robust and interpretable Linear Regression model.

Evaluation metrics

To assess the model's performance, we used some evaluation metrics, mainly mean squared error (MSE), R-squared, and mean absolute squared error (MSME). These metrics provided insights into the model's accuracy and predictive power.

Performance metrics

Once the model is trained, it is essential to assess its performance on the test data to ensure its reliability and effectiveness [16]. The following evaluation metrics were used to measure the performance of our model: R-squared, mean squared error, and root mean squared error.

- *R-squared* R-Squared (R^2 or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, r-squared shows how well the data fit the regression model (Frost, 2019). Ranging from 0 to 1, a higher R-squared value indicates that the model can better account for the observed variance, with a value of 1 suggesting a perfect fit.
- *Mean squared error (MSE)* Mean squared error (MSE) calculates the average of the squared differences between the predicted and actual values [17]. It serves as a measure of the quality of a model by quantifying the extent to which the predictions deviate from the actual data points. A lower MSE indicates a better model performance, as it signifies the proximity of the predicted values to the actual values, while a higher MSE indicates larger errors and poorer model performance.
- *Mean Absolute Error (MAE)* Mean absolute error measures the average of the absolute differences between the predicted and actual values [17]. Unlike MSE, MAE is less susceptible to the influence of outliers, as it does not square the errors.

Proposed model

Our proposed model is based on a linear regression approach, which assumes a linear relationship between the dependent variable (Performance Index) and independent variables (factors influencing academic outcomes). The general form of the linear regression model can be expressed as:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn \ldots\ldots\ldots\ldots \text{Eq. (1)}.$$

Where:

Y is the Performance Index.

- β0 is the intercept (the constant value).
- β1, β2, β3, β4, β5 are the regression coefficients for each of the input features (Hours Studied, Previous Scores, Extracurricular Activities, Sleep Hours, Sample Question Papers Practiced).

In applying this formula, we have:

Performance Index (Y) = β0 + β1 * Hours Studied + β2 * Previous Scores + β3 * Extracurricular Activities + β4 * Sleep Hours + β5 * Sample Question Papers Practiced.

To determine the exact values of these coefficients based on our data, we use can statistical software or programming libraries (e.g., Python's scikit-learn, R) to perform a multiple linear regression analysis. This will help us to find the optimal coefficients that minimize the error between the predicted Performance Index and the actual Performance Index values in your dataset.

Linear regression

Linear regression, a linear modeling technique, has been widely employed to predict academic performance due to its suitability for continuous outcomes. Several studies have utilized linear regression to forecast students' grades, often using a combination of demographic, behavioral, and prior academic information as predictor variables.

Linear regression provides a straightforward, transparent, and interpretable method for forecasting academic performance, rendering it a valuable tool, particularly during the initial exploration phase. Nevertheless, its efficacy diminishes when confronted with intricate, nonlinear relationships and may not encompass all pertinent factors influencing academic success. Researchers ought to judiciously weigh these merits and drawbacks when selecting a modeling approach for predicting academic performance.

Architecture of proposed model

In this section, we outline the structure of our model tailored for forecasting students' academic performance through the utilization of linear regression, a statistical technique frequently applied in predictive endeavors of this nature. The fundamental elements of our model architecture include:

Data preprocessing Prepare the data for analysis through data preprocessing steps, this includes:

- *Handling missing data* Strategies on removal of missing data.
- Encoding categorical variables: Convert categorical variables into numerical format if necessary. This is done for the "*Extracurricular Activities*"
- *Data splitting* Divide the dataset into training and testing sets for model validation.

Model selection Linear regression serves as the chosen predictive modeling technique, as its appropriateness lies in forecasting continuous target variables such as academic performance, owing to its simplicity, interpretability, and efficiency. It provides significant insights into the associations between independent variables and outcomes, rendering it a valuable instrument for comprehending and predicting academic success.

Model training Train the linear regression model using the training dataset. The model estimates the coefficients (β values) that best fit the data, minimizing the sum of squared residuals between the predicted values and the actual academic performance scores.

Model evaluation The performance of the linear regression model was measured using various evaluation metrics, such as R-squared (R2), mean absolute error (MAE), and root mean square error (RMSE). These metrics provide insights into how well the model predicts academic performance.

Validation and testing This is a very important step. It enables us to validate the model's performance on the testing dataset to ascertain its generalization to new data. This process aids in evaluating the model's predictive accuracy and safeguards against overfitting to the training data.

Prediction and application Once the linear regression model is validated and optimized, it is applied to make predictions on new or future academic performance data. Consequently, educational institutions, educators, and other policymakers are able to make informed decisions with respect to students' needs.

Flowchart of proposed model

The below flowchart outlines the selection of input and the procedure to get the output (performance index). The flowchart illustrates the first step as a selection of inputs that are parameters, processing of these inputs and completing training, testing and prediction for accurate and precise output. The flowchart in Fig. 1 illustrates the sequential steps carried out to generate a prediction (performance index).

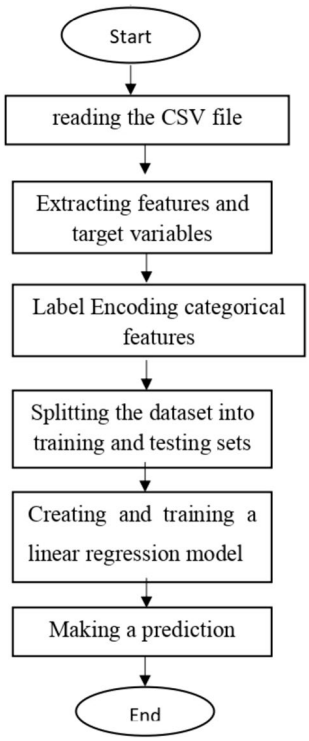

Fig. 1 Flowchart of proposed model

Data presentation and analysis

The aim our study is to systematically investigate and identify the primary determinants that have a significant impact on students' academic success. From our study, the following results were obtained.

Coefficients and intercept From the dataset presented, we obtained the coefficients and the intercept of the linear regression equation. These coefficients represent the relationship between each input feature and the target variable. The intercept represents the predicted value when all input features are zero.

Using Python's scikit-learn and our results for the intercept and regression coefficients for each of the respective input features "*Hours Studied, Previous Scores, Extracurricular Activities, Sleep Hours, and Sample Question Papers Practiced*" are:

- $\beta 0 = -34.09$
- Hours studied ($\beta 1 = 2.852$)
- Previous score ($\beta 2 = 1.018$)
- Extracurricular activities ($\beta 3 = 0.586$)
- Sleep hour ($\beta 4 = 0.48$)
- Past Questions Practiced ($\beta 5 = 0.201$)

Prediction Once the model is trained, we can reliably make predictions on the test data. We supply values for the predictors and then make a prediction (Predicted performance Index). If;

- Hours studied is 10
- Previous score is 90
- Extracurricular activities: 1
- Sleep hour is 7
- Past Questions Practiced is 5

Our predicted performance Index is 91.1. This can be clearly seen in Fig. 2

```
X = RESULT.drop(columns = "Performance Index")
y = RESULT["Performance Index"]

from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()

X["Extracurricular Activities"] = le.fit_transform(X["Extracurricular Activities"])

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = (20/100))

from sklearn.linear_model import LinearRegression

model = LinearRegression()
model.fit(X, y)
print(f"Performance Index is:")
model.predict([[10, 90, 1, 7, 5]])
```

```
Performance Index is:
Out[49]: array([91.05913632])
```

Fig. 2 Prediction (performance index)

```
In [23]: from sklearn.linear_model import LinearRegression
         from sklearn.metrics import r2_score,mean_squared_error,mean_absolute_error, accuracy_score

         model = LinearRegression()
         model.fit(X, y)
         model.fit(X_train, y_train)
         SCORE = model.predict(X_test)
         mean_squared_error(SCORE, y_test)
         mean_absolute_error(SCORE, y_test)
         r2_score(SCORE, y_test)
         print(f" the R-squared is:")
         r2_score(SCORE, y_test)

          the R-squared is:

Out[23]: 0.9882381335376964
```

Fig. 3 R-Squared score

```
In [24]: from sklearn.linear_model import LinearRegression
         from sklearn.metrics import r2_score,mean_squared_error,mean_absolute_error, accuracy_score

         model = LinearRegression()
         model.fit(X, y)
         model.fit(X_train, y_train)
         SCORE = model.predict(X_test)
         mean_squared_error(SCORE, y_test)
         mean_absolute_error(SCORE, y_test)
         r2_score(SCORE, y_test)
         print(f" the Mean squared error is:")
         mean_squared_error(SCORE, y_test)

          the Mean squared error is:

Out[24]: 4.217148735880705
```

Fig. 4 Mean squared error

```
In [25]: from sklearn.linear_model import LinearRegression
         from sklearn.metrics import r2_score,mean_squared_error,mean_absolute_error, accuracy_score

         model = LinearRegression()
         model.fit(X, y)
         model.fit(X_train, y_train)
         SCORE = model.predict(X_test)
         mean_squared_error(SCORE, y_test)
         mean_absolute_error(SCORE, y_test)
         r2_score(SCORE, y_test)
         print(f" the Mean Absolute Error is:")
         mean_absolute_error(SCORE, y_test)

          the Mean Absolute Error is:

Out[25]: 1.6283644640783481
```

Fig. 5 Mean absolute error

Model performance metrics

Once the model is trained, and used to make predictions on the test data, Evaluation metric. The following evaluation metrics were used to measure the performance of our model. These metrics are: R-squared, mean squared error (MSE), R-squared, and root mean squared error (RMSE).

As shown in Fig. 3, the r2 from our project work was 0.98.

The mean squared error from pour project work was 4.22. This is shown in Fig. 4.

The mean absolute error from our project work was 1.62. This is shown in Fig. 5.

Discussion of results

From our study, we used specific variables and obtained the coefficients associated with those variables. The summary of the findings from our study are stated below.

Study habits play a significant role

Study habits like including study time, study location, and most especially study hours were found to have a substantial impact on academic performance. Students who time to time in conducive environments tend to achieve better grades.

Figure 6 helps to show the relationship between study hour and performance of a student in school.

From Fig. 6, it is evident that there is a positive correlation between study hours and the performance index. This means that as the number of hours spent studying increases, the performance index also tends to increase.

This relationship shows that students who dedicate more time to studying are likely to achieve higher performance levels. This positive trend indicates that consistent study sessions contribute to better understanding and retention of the material, leading to improved academic performance (performance index).

The graph in Fig. 6 shows a linear trend where the performance index rises steadily as study hours increase. This pattern reinforces the idea that time investment in studying pays off in terms of higher academic achievement.

Figure 7 is a correlation matrix of performance index prediction. The matrix shows the correlation between all variables involved (Hours studied, Previous Score, Extracurricular activities, sleep hours and sample question papers practiced). The correlation values range from −1 to 1, with 1 being a strong positive correlation, −1 being a strong

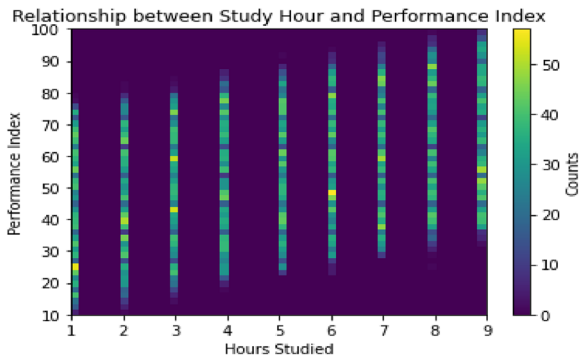

Fig. 6 Relationship between study hour and performance index

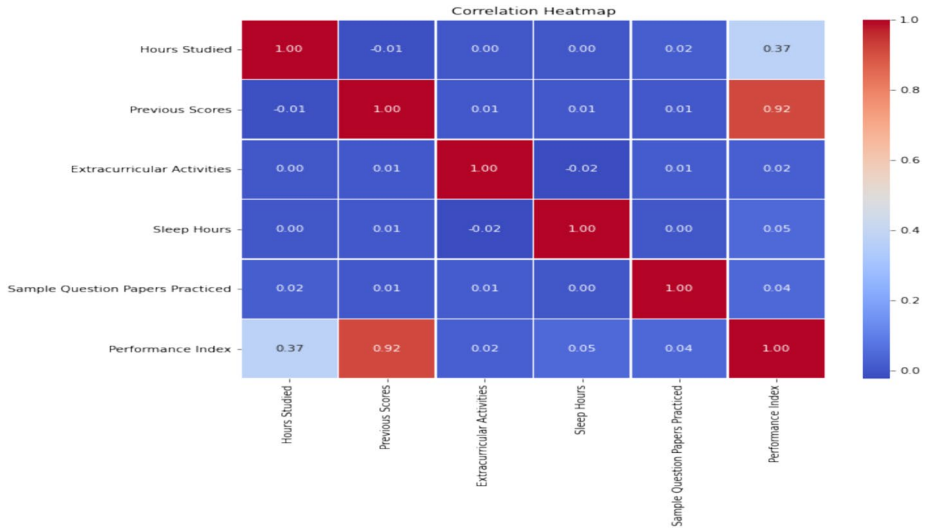

Fig. 7 Correlation matrix

negative correlation, and 0 being no correlation. The matrix is color-coded, with dark-red colors representing stronger correlations. Based on the matrix, it appears that there is a strong positive correlation between Hours studied and academic performance (Performance Index) indicating that the time spent studying has a great impact in student's performance.

Explanation of each factor

Study hours

- *Correlation with Performance Index* 0.37
- *Interpretation* There is a moderate positive correlation between Study Hours and Performance Index. This means that increasing study hours moderately improves performance.

Previous scores

- *Correlation with Performance Index* 0.92
- *Interpretation* There is a very strong positive correlation between Previous Scores and Performance Index. This indicates that previous academic performance is a highly reliable predictor of current performance. Students with higher previous scores tend to have significantly higher performance indexes. In simple terms, a students' with good academic history will perform well.

Extracurricular activities

- *Correlation with Performance Index* 0.02

- *Interpretation* This is the correlations between Extracurricular Activities and Performance Index. It indicates an extremely weak positive relationship, suggesting that the factor (Extracurricular Activities) has almost no impact on performance. This is a near-zero correlation, and by implication, changes in this factor do not significantly influence the performance index.

Sleep hours

- *Correlation with Performance Index* 0.05
- *Interpretation* This is the correlations between Extracurricular Activities and Performance Index. Just like Extracurricular Activities, it indicates an extremely weak positive relationship, suggesting that Sleep Hours has almost no impact on performance. This implies that changes in this factor do not significantly influence the performance index.

Sample questions practiced

- *Correlation with Performance Index* 0.04
- *Interpretation* There is a very weak positive correlation between Sample Questions Practiced and Performance Index. This suggests that practicing sample questions has a minimal effect on improving performance.

From our correlation matrix, we can deduce the following

- *Previous Scores* have the most significant impact on Performance Index, with a very strong positive correlation of 0.92.
- *Study Hours* also positively affect Performance Index, but to a lesser degree, with a moderate correlation of 0.37.
- *Extracurricular Activities, Sleep Hours and Sample Questions Practiced* have negligible effects on the Performance Index, with correlations close to zero.

Policy and program recommendations

Based on the findings, recommendations include the implementation of programs that promote effective study habits, parental involvement, and time management skills. Policymakers should consider strategies to reduce the impact of socioeconomic disparities on academic outcomes.

Limitations and future directions

Linear regression assumes a linear relationship between the features and the target variable. If the actual relationship is more complex, the model might not capture it accurately. Insights into cases where the model performs poorly can highlight scenarios in which a more sophisticated model or additional features might be necessary.

In view of this, it is advisable to obtain results that can be represented in terms of percentage of accuracy, interested individuals or groups who intend to carry out further work on this project are advised to use other machine learning algorithms and use more than one source of data. One can also use other platforms such as WEKA and SPSS to implement the algorithms for prediction, thereby improving the level of understanding of the students and teachers.

Conclusion

This study on has shed light on the intricate web of variables that influence students' success in their educational pursuits. Our findings underscore the central role of prior academic performance as well as study hour as a robust predictor of future achievement, emphasizing the cumulative nature of learning. Additionally, the significance of effective study habits and time management cannot be overstated, as students who cultivate disciplined routines and strategic study approaches are better positioned for academic excellence. Furthermore, extracurricular engagements and revision of previous papers have emerged as influential factors, highlighting the importance of intensive study.

It must be noted that balancing extracurricular activities with academic commitments is a key challenge, necessitating a thoughtful approach to time allocation.

This study also highlights the capacity of data-driven methodologies to influence the trajectory of education's future. Through the utilization of a linear regression model, we have unveiled the complex interconnections among variables such as study routines, involvement in extracurricular activities, prior academic achievements, sleep patterns, and performance metrics. These revelations go beyond mere numerical statistics, offering valuable guidance to stakeholders, enabling them to make informed decisions, and facilitating impactful interventions.

Author contributions
All authors contributed collectively to the conceptualization, design, data collection, analysis and interpretation of the study. All authors read and approved the final manuscript.

Funding
This study was self-funded and no external sources of funding was utilized in the research process.

Availability of data and materials
The dataset generated/analyzed during this study is available on KAGGLE repository. Kaggle provides an extensive collection of resources, including datasets which makes it a valuable platform for data exploration, analysis, and the advancement of data-driven research and solutions. The link to this dataset is: https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 1 January 2024 Accepted: 22 September 2024
Published online: 30 September 2024

References

1. Traub J, Bajwa K, Agbaw E, Hussein S, Aakre MH, Ferullo T (2019) Exploring the relationship between study hours, sleep, stress, and academic performance of college students. J Instr Psychol 46(1):27–36
2. Squires T., Coates LG (2017) Effects of study time on exam performance: an analysis of multilevel data. J Educ Psychol.

3. Nonis SA, Hudson GI (2010) Performance of college students: impact of study time and study habits. J Educ Bus 85(4):229–238. https://doi.org/10.1080/08832320903449550
4. Baliyan SP, Khama D (2020) How distance to school and study hours after school influence students' performance in mathematics and English: a comparative analysis. J Educ e-Learn Res 7(2):209–217
5. Chung GK, Kyung HR (2015) Effects of study time on exam performance: study time and students' academic performance: a meta-analysis. J Educ Psychol.
6. Liu J, Conger RD, Wu Y (2020) The future significance of high school grade point average for college success. J Appl Sch Psychol 36(2):158–176
7. Kocak O, Goksu I, Goktas Y (2021) The factors affecting academic achievement: a systematic review of meta analyses. Int Online J Educ Teach (IOJET) 8(1):454–484
8. Kassarnig V, Mones E, Bjerre-Nielsen A et al (2018) Academic performance and behavioral patterns. EPJ Data Sci 7:10. https://doi.org/10.1140/epjds/s13688-018-0138-8
9. Arora N, Singh N (2017) Factors affecting the academic performance of college students. I -manager's J Educ Technol 14(1):47–53
10. Husaini AY, Shukor A, Syufiza N (2023) Factors affecting students' academic performance: a review. Res Mil 12:284–294
11. Zingoni AK, Bates CJ (2022) Course difficulty as a predictor of academic success in college-level online science, technology, engineering, and mathematics (STEM) courses. Int J Learn Technol 13(5–6):358–374
12. Iliya ME, Musa RM (2017) The effectiveness of past question papers in secondary school mathematics. J Acad Res Educ.
13. Khalil Y, Bangud C (2022) The impact of past exams as a learning strategy on learning outcomes: an experimental study. African J Res Math, Sci Technol Educ 26(1):15
14. Shabiha A (2021) Impact of extracurricular activities on academic performance of students at secondary level. Int J Appl Guid Couns 2:7–14. https://doi.org/10.26486/ijagc.v2i2.1869
15. An J, Lee KH (2021) Sports participation and academic achievement: the impact of playing sports on academic performance. SAGE Open 11(2):1–18
16. Varoquaux G, Colliot O (2023) Evaluating machine learning models and their diagnostic value. In: Colliot O (ed) Machine learning for brain disorders (Chapter 20). Humana, New York, NY, p 2023
17. Orji FA, Vassileva J (2023) Modeling the impact of motivation factors on students' study strategies and performance using machine learning. J Educ Technol Syst 52(2):274–296. https://doi.org/10.1177/00472395231191139

Publisher's Note