Journal of Engineering Education Transformations ,
Volume 38 , No. 3 , January 2025 , ISSN 2349-2473, eISSN 2394-1707

10.16920/jeet/2024/v38i3/24251

Evaluating Elements Influencing Student Engagement: An Analysis Utilizing Multiple Regression

R Saraswathi Meena[1], R Narmadha[2], P K Tejaswini[3], M Saravana Kumar[4]

[1,2,3,4] Department of Applied Mathematics and Computational Science,
Thiagarajar College of Engineering Madurai, India
[1] saraswathimeenacse@gmail.com,
[2] narmadhabts@gmail.com,
[3] tejaswini.pk.2004@gmail.com,
[4] saravananmsk1210@gmail.com

Abstract:"Exploring Student Engagement: A Comprehensive Multiple Regression Analysis" delves into the intricate interplay of factors influencing student engagement, focusing on variables such as class size, study hours, teaching style, class hours, and technology use. Conducted in a diverse educational setting with undergraduate students from urban and suburban areas, including a significant proportion of first-generation college students, this study employed extensive data collection and analysis. The results revealed significant correlations between these factors and engagement levels, offering actionable insights for educators. The findings underscore the critical influencers of engagement, facilitating targeted improvements in teaching practices and classroom environments. This approach equips educators with knowledge about effective interventions, enabling them to assess their impact and tailor strategies to individual student needs. By doing so, the study aims to enhance overall learning outcomes, offering a valuable framework for fostering student engagement through informed, evidence-based educational practices, thus significantly contributing to the academic success of diverse student populations.

R Saraswathi Meena
Department of Applied Mathematics and Computational Science,
Thiagarajar College of Engineering Madurai, India
saraswathimeenacse@gmail.com,

Keywords : education, learning outcomes, student engagement.

1. Introduction

In the ever-evolving landscape of education, the imperative to optimize student learning outcomes has propelled a dedicated exploration into the intricate realm of factors influencing student engagement. Recognizing the pivotal role of student engagement as a linchpin for effective teaching practices and the creation of conducive learning environments, educators are compelled to unravel the complexities inherent in this symbiotic relationship. Engaged students, characterized by heightened motivation, genuine interest, and profound investment in the learning process, not only epitomize the essence of educational success but also translate their commitment into improved academic performance and elevated retention rates.

This study aims to investigate the nuanced dynamics between various influential factors and levels of student engagement. Specifically, it seeks to discern the intricate interplay and relative significance of class size, teaching style, technology utilization, and class hours. The objectives are to provide educators with precise insights derived from robust statistical analysis, guiding them in making informed decisions about tailored teaching practices designed to foster and sustain student engagement.

With a specific lens on class size, teaching style, technology utilization, and class hours, this research emphasizes the unique combination of these elements. While existing literature has illuminated the varying impacts of these factors on student engagement, their synergistic effects and comparative importance remain elusive. The chosen methodology for this exploration is the rigorous application of multiple regression analysis, a statistical tool meticulously employed to pinpoint the factors exerting the most substantial influence on student engagement.

The study population comprises a diverse demographic of students from various educational backgrounds, providing a comprehensive context for analyzing the factors influencing student engagement. By delving into the complex relationships among class size, teaching style, technology integration, class hours, and student engagement, the study aspires to empower educators with insights tailored to diverse educational settings.

The practical implications of this research transcend the boundaries of individual classrooms, offering pragmatic guidance for educators grappling with budget constraints. The study underscores the benefits of smaller class sizes Still, it acknowledges the practical need for alternative strategies like active learning and technology integration, which offer viable ways to enhance student engagement despite financial constraints. Additionally, the study serves as a valuable repository of input for policymakers who are tasked with making critical decisions on issues such as class size limits, teacher training initiatives, and technology investments.

In essence, this study not only delves into the origins and motivations behind exploring factors influencing student engagement but also represents a proactive effort to fill existing gaps in knowledge. Through the meticulous lens of multiple regression analysis, it seeks not merely to uncover influential factors but to prioritize them, providing a nuanced understanding of their impact. Ultimately, the goal is to equip educators and policymakers with actionable insights, fostering a transformative approach to educational practices that not only enhances student engagement but paves the way for improved academic outcomes.

2. Study Context

In the dynamic landscape of contemporary education, the relentless pursuit of effective teaching practices and heightened learning outcomes stands as a paramount concern for educators and policymakers alike. At the core of this collective endeavor lies the foundational concept of student engagement—an acknowledged linchpin for academic success and student retention. The intricacies of student engagement, spanning motivation, interest, and active participation, have catalyzed a dedicated exploration into the myriad factors that influence this critical aspect of the learning experience.

Fig.1 : Context of the study

Conducted in a diverse educational environment characterized by undergraduate students from urban and suburban areas, with a significant proportion of first-generation college students, this study explores the interplay of various factors influencing student engagement. The context-specific insights gained from this analysis contribute to a broader understanding of student engagement dynamics.

Student engagement, as a multifaceted construct, embodies the essence of a student's involvement and commitment to the learning process. Its facets, including motivation, interest, and active participation, converge to shape the trajectory of academic success and the likelihood of student retention. Recognizing the pivotal role that student engagement plays in shaping the educational experience, this study embarks on a comprehensive exploration, delving into the nuanced relationships between various factors and the levels of student engagement.

A. Methodological lens: multiple regression analysis

To contribute meaningfully to the ongoing discourse on student engagement, this study adopts multiple regression analysis as its methodological lens. This statistical tool, characterized by its ability to discern intricate relationships within complex datasets, is employed to scrutinize the interplay between pivotal elements of the educational environment and student engagement. The factors under meticulous consideration—class size, teaching style, technology use, and class hours—have been identified as key determinants, and the study aims to unravel the complexity surrounding their interactions and their relative significance.

The impetus driving this research emanates from the pragmatic challenges encountered by educators and policymakers in the quest to optimize educational practices within resource constraints.

Employing multiple regression analysis, this study meticulously examines the relationships between class size, teaching style, technology use, class hours, and student engagement. This methodological rigor provides a robust framework for understanding the complex dynamics at play, offering new statistical insights into the field.

The ideal of smaller class sizes, often lauded as a catalyst for improved student engagement, inevitably collides with real-world budget limitations. In response, this study, grounded in robust statistical analysis, aspires to transcend these challenges by offering pragmatic insights. Through evidence-based decision-making, educators can navigate resource constraints while fostering meaningful student engagement.

B. Extending impact: transcending classrooms to policy realms

Beyond individual classrooms, the study envisions its findings transcending into the broader realm of education policy. Policymakers, entrusted with decisions ranging from class size limits to teacher training initiatives and technology investments, are poised to benefit from a nuanced understanding of the intricate factors influencing student engagement. The study emerges as a valuable resource, offering insights that extend beyond immediate pedagogical concerns to inform strategic decisions at the policy level.

C. The collective commitment: refining educational practices

In essence, this study operates within the broader context of a collective commitment to refining educational practices. By peeling back the layers of student engagement through the intricate lens of multiple regression analysis, the research seeks to equip educators and policymakers with actionable insights. These insights, derived from a meticulous examination of diverse educational settings, contribute to the ongoing dialogue on how to systematically improve student learning outcomes.

Fig. 1 serves as a visual representation of the contextual framework of the study. It meticulously showcases various features under investigation, prompting an exploration into the unidentified factors potentially influencing the depicted variables.

The visual representation encapsulates the complexity of the educational ecosystem, providing a starting point for the in-depth analysis undertaken by this study.

In conclusion, this comprehensive exploration into the intricate dynamics of student engagement, underpinned by multiple regression analysis, offers a roadmap for transformative action in education. By elucidating the relationships between class size, teaching style, technology integration, class hours, and student engagement, this study empowers educators to make informed decisions. Moreover, it positions policymakers to craft strategic initiatives that resonate with the complexities of student engagement in diverse educational settings. As the educational landscape continues to evolve, this research serves as a beacon—a testament to the collective commitment to fostering enriched student engagement and, consequently, achieving improved academic outcomes in the 21st century and beyond.

3. Methods

In the intricate landscape of unraveling the dynamics influencing student engagement, this study adopts a comprehensive and multi-faceted approach, employing several advanced regression methods. These include Ordinary Least Squares (OLS), Multilinear Regression, Random Forest, Lasso, and Ridge Regression. This diverse methodological framework is designed to provide a robust and nuanced analysis of the factors affecting student engagement.

A. Research design

The study's research design is centered around a quantitative approach, utilizing regression analysis to explore the relationships between various independent variables (class size, teaching style, technology use, and class hours) and the dependent variable (student engagement). The primary objective is to identify and quantify the impact of these factors on student engagement, offering insights that can inform educational practices and policies.

B. Data collection methods

Data for this study were collected from a diverse population of students across different educational settings. A stratified random sampling technique was employed to ensure that the sample is representative of the broader student population, capturing various demographic characteristics such as age, gender, socioeconomic status, and academic background. Surveys and observational methods were used to gather data on the independent variables and measure student engagement levels. The survey included standardized questions designed to assess the quality and extent of student engagement, while observational data provided additional context and validation.

C. Analysis procedures

The analysis procedures involved the application of multiple regression techniques to analyze the collected data. Each method was chosen for its unique strengths and ability to provide different perspectives on the data.

D. Ordinary least squares (ols) regression

OLS regression was utilized as a foundational technique to model the relationship between the dependent and independent variables. The primary goal of OLS is to minimize the sum of squared differences between observed and predicted values, resulting in coefficients that signify the strength and direction of relationships. The simplicity and interpretability of OLS make it a valuable tool for initial analysis.

$$minimize \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (1)$$

Where,

y_i is the observed value of the dependent variable.

$\hat{y}_i$ is the predicted value of the dependent variable.

n is the number of observations.

2) Multilinear regression

Building on OLS, multilinear regression allows for the inclusion of multiple independent variables simultaneously. This method provides a more comprehensive understanding of how each factor influences student engagement when considered in conjunction with other variables. It helps to identify the combined effects and potential interactions between factors.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \qquad (2)$$

Where,

y is the independent variable.

β_0 is the y intercept.

$\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients of independent variables.

$x_1, x_2, \ldots, x_n$ are the independent variables.

ε is the error term.

Random Forest

Random Forest, a powerful ensemble learning method, was employed to capture the complex, non-linear relationships between the variables. This method uses multiple decision trees to improve prediction accuracy and handle large datasets with high dimensionality. It is particularly useful for identifying the most significant predictors of student engagement.

4) Lasso Regression

Lasso (Least Absolute Shrinkage and Selection Operator) regression was applied to enhance model selection and regularization. This method penalizes the absolute size of the regression coefficients, effectively shrinking some coefficients to zero, thereby selecting a simpler and more interpretable model. Lasso is beneficial for handling multicollinearity and improving model generalizability.

$$minimize \left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right) \qquad (3)$$

Where,

λ is the penalty term (regularization parameter).

β_j are the coefficients of independent variables.

p is the number of predictors.

5. Ridge Regression

Ridge regression, another regularization technique, was used to address multicollinearity by penalizing the square of the coefficients. Unlike Lasso, Ridge does not set coefficients to zero but instead shrinks them, reducing model complexity and improving robustness. This method helps to balance the trade-off between bias and vari

$$minimize\left(\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\lambda\sum_{j=1}^{p}\beta_j^{\,2}\right) \qquad (4)$$

Where,

λ is the penalty term (regularization parameter).

are the coefficients of independent variables.

p is the number of predictors.

By employing these diverse methods, the study ensures a thorough and nuanced analysis of the factors influencing student engagement. Each technique offers unique insights, contributing to a comprehensive understanding of the individual and collective impact of class size, teaching style, technology use, and class hours on student engagement.

In the subsequent sections, the study will present the findings from these analyses, offering detailed interpretations and practical recommendations for educators and policymakers. The multi-faceted approach adopted in this research aims to provide actionable, easily understandable insights, fostering a more targeted approach to optimizing student engagement within the realistic constraints of real-world resource limitations.

4. Data Analytics

In the realm of educational research, the implementation of robust data analytics is instrumental in uncovering the nuanced factors influencing student engagement. This section delves into the comprehensive data collection process, elucidating the various student-related variables gathered through Google Forms for meticulous analysis.

1. Emails: Emails of students are collected as part of the dataset. Although they might not directly impact the study's core focus on student engagement, they serve as a unique identifier, ensuring accurate tracking of responses.

2. Names: The names of students are gathered to maintain a record of respondents. While not contributing to the study's substantive analysis, they play a crucial role in tracking and organizing data. It is acknowledged that names will be dropped later for the sake of anonymity and streamlined analysis.

3. Gender: Gender information is collected to demonstrate its unrelated nature to the study's observations. It serves as evidence that gender is not a variable required for the study, and accordingly, it can be dropped during the analysis phase.

4. Age: Age data is gathered to explore its potential impact on individual student scores. Subsequent analysis reveals that age does not significantly influence scores, leading to its elimination from the dataset.

5. Marks Gained: Marks gained by each student, representing semester marks out of 700, stand as the central output variable. This pivotal metric serves as the cornerstone for assessing academic performance and its relationship with other variables.

6. Class Strength: The number of students present in each class is collected, aiming to predict its impact on individual scores. This variable explores the potential influence of class size on academic performance.

7. Study Hours: Individual study hours per day are gathered to assess their impact on individual scores. This variable provides insights into the relationship between study time and academic achievement.

8. Courses Applied: Students are queried about their enrollment in additional courses beyond the standard college curriculum. This information aims to discern whether pursuing additional courses impacts academic scores.

9. Rating of Study Materials: Ratings of study materials provided by the college are collected. This variable investigates whether the perceived quality of

study materials correlates with individual scores.

10. Technology: Information about the technology used for communication in class is collected. Categorical variables such as 'blackboard,' 'presentation,' and 'online course materials' are converted into dummy variables for preprocessing, enabling a nuanced analysis of their impact on understanding and scores.

11. Teaching Style: The teaching method employed, whether project-based or lecture-based, is specified. Categorical variables are later converted into dummy variables for preprocessing, enabling an exploration of how teaching style influences student understanding and academic performance.

12. Hours: Data on the study hours per day of individuals is gathered to assess its impact on individual scores. This variable offers valuable insights into the relationship between study time and academic achievement.

The richness of this dataset, encompassing demographic details, academic performance metrics, study habits, and perceptions about the learning environment, positions it as a comprehensive resource for analyzing the multifaceted factors shaping student engagement.

A. Data preprocessing

1) Data Cleaning

An essential and thorough stage in the dataset's preprocessing, data cleaning has a significant impact on the accuracy and dependability of ensuing studies. This complex procedure aims to detect, correct, and improve the dataset's overall quality, covering all the necessary elements to guarantee its integrity.

Handling Missing Values: Addressing missing data is imperative to uphold the dataset's integrity. Employing suitable imputation techniques, such as mean or median imputation, becomes essential when dealing with incomplete data points. Mean imputation, as per equation (2), involves replacing missing values with the average of the available data, thereby ensuring a representative estimate. This method proves valuable in maintaining the dataset's completeness and preserving the statistical properties of the variables.

$$Mean = \frac{\sum_{i=1}^{n} x_i}{n} \tag{5}$$

Similarly, median imputation, outlined in equation (5), offers a robust alternative, particularly in scenarios involving skewed data or outliers. By replacing missing values with the middle value of the dataset, it accommodates both odd and even sample sizes. This adaptive approach ensures a more accurate representation of central tendencies, enhancing the dataset's suitability for analysis.

$$Median(X) = \begin{cases} X\left[\frac{n+1}{2}\right], if\ n\ is\ odd \\ \frac{X\left[\frac{n}{2}\right] + X\left[\frac{n}{2}+1\right]}{2}, if\ n\ is\ even \end{cases} \tag{6}$$

Elimination of Duplicate Entries: Preserving the authenticity of the dataset necessitates a meticulous process of identifying and eliminating duplicate entries. Duplicate entries can introduce bias and distort analyses, making their removal crucial for maintaining the uniqueness of each data point. This step contributes to the robustness of subsequent analyses by ensuring that each observation is distinct and carries unique information.

Validation of Categorical Responses: Enhancing data accuracy involves a comprehensive validation process for categorical responses, including age values, email addresses, and other pertinent information. This rigorous validation verifies the correctness of entries, fortifying the overall reliability of the dataset. Inaccuracies in categorical data could potentially introduce errors in subsequent analyses, underscoring the importance of this validation step.

Fig. 2 serves as a visual representation of the outlier detection process through a boxplot comparing class strength and study hours. The identification of a singular data point lying outside the whiskers signals a potential outlier in the dataset. This outlier prompts further investigation to discern its implications within the broader context of the study.

Outliers, due to their potential to significantly impact statistical analyses and model outcomes, demand careful consideration. Understanding and addressing these anomalies is crucial for ensuring the robustness of subsequent analyses. The visualization presented in Fig. 2 serves as a preliminary step in outlier detection, guiding the subsequent steps in scrutinizing and, if necessary, mitigating the impact of outliers on the study's findings.

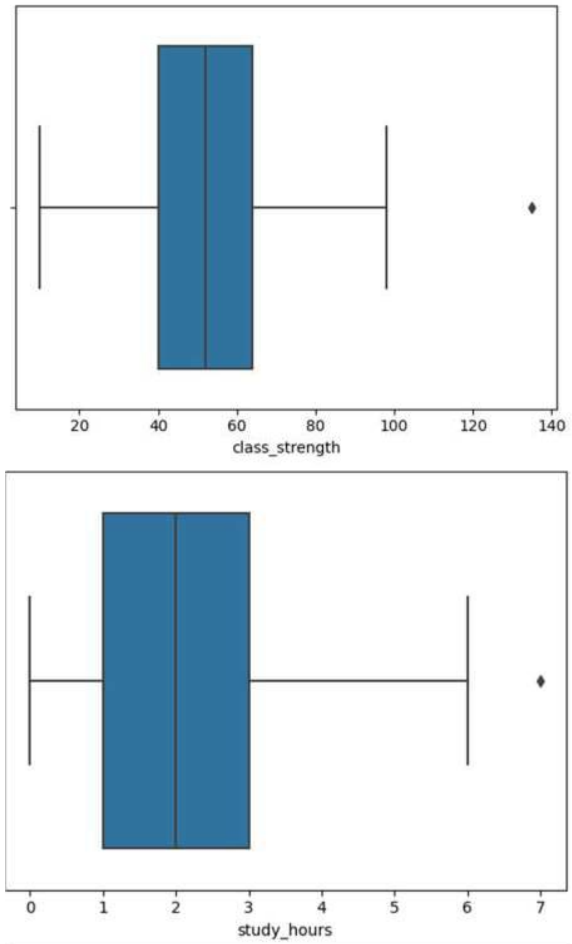

Fig. 2 : Outlier Detection

2) Data Transformation

Following the meticulous process of data cleaning, the dataset advanced into a critical phase of transformation, essential for rendering it amenable to comprehensive analysis. This transformation was a strategic endeavour to convert the dataset into a format suitable for sophisticated statistical analyses, particularly regression analysis. The primary objectives were to encode categorical variables, standardize or normalize numerical variables, and ensure homogeneity in the dataset, laying the groundwork for a robust examination of factors influencing student engagement.

Encoding Categorical Variables: Categorical variables, such as gender, teaching style, course applied, and technology used, inherently pose a challenge for many statistical models that expect numerical input. To overcome this challenge, the dataset underwent encoding processes, translating qualitative information into a numerical format that could be effectively utilized in subsequent analyses.

One-Hot Encoding: One-Hot Encoding was employed for categorical variables with no inherent order or ranking, such as teaching style and technology used. This technique creates binary columns for each category, indicating the presence or absence of that category for each observation. For instance, the "teaching style" variable, initially comprising categories like "project-based" or "lecture-based," was transformed into binary columns with clear numerical representations (1 or 0) for each teaching style category.

Label Encoding: Label Encoding, on the other hand, was applied to categorical variables where there existed a natural ordinal relationship, such as the course applied variable. This method assigns a unique numerical label to each category based on its order. For instance, if there are three courses labeled A, B, and C, Label Encoding might assign them values 1, 2, and 3, respectively.

Numerical variables, including age and total marks gained, underwent standardization or normalization to bring them onto a comparable scale. This step is crucial in regression analysis, where variables with different units or magnitudes can introduce biases and inconsistencies in the model.

Standardization: To create a mean of 0 and a standard deviation of 1, numerical variables must be transformed. This procedure guarantees that the variables have a standard scale and keeps the analysis from being dominated by those with bigger magnitudes. For example, after being standardized to a dimensionless number, the "age" variable—which was formerly measured in years—was directly comparable to other standardized variables.

Normalization: Normalization, akin to standardization, is applied to bring numerical variables within a specific range, often between 0 and 1. This is particularly useful when dealing with variables that have different scales. In the case of the "total marks gained" variable, normalization ensured that its values fell within a standardized range, eliminating potential biases arising from disparate magnitudes.

The overarching goal of these transformations was to homogenize the dataset, creating a standardized and

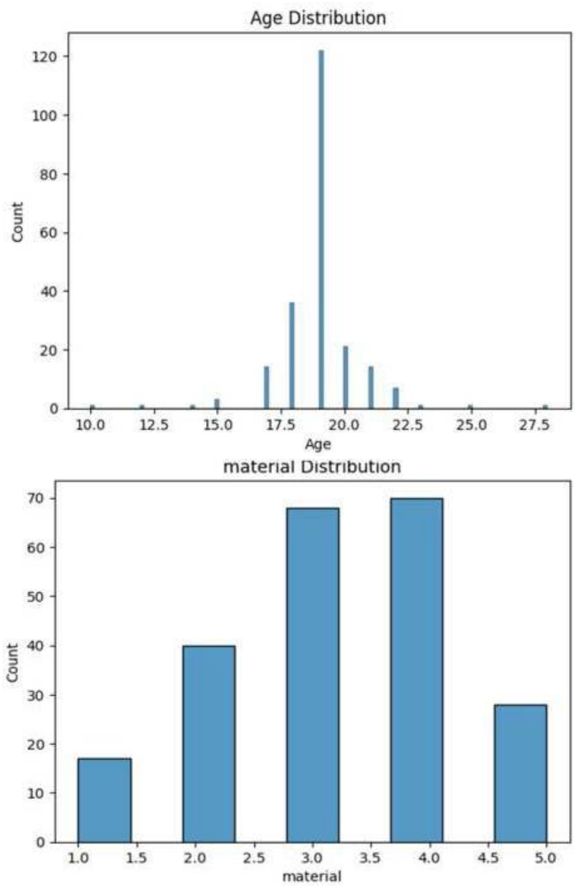

Fig. 3 : Distribution of parameters

consistent foundation for subsequent analyses. By encoding categorical variables and standardizing or normalizing numerical variables, the dataset was tailored to meet the assumptions and requirements of regression analysis. This strategic transformation ensures that each variable, irrespective of its nature, contributes meaningfully to the exploration of factors influencing student engagement.

3) Feature Selection

In the intricate process of preparing the dataset for analysis, feature selection emerged as a pivotal step, addressing the nuanced task of identifying and isolating the most influential features likely to impact student engagement—the focal point of investigation. This critical phase involved the strategic curation of relevant features, eliminating those that proved redundant or lacked a significant contribution to the analysis. The dual lens of domain knowledge and statistical techniques guided this feature selection endeavor, ensuring a focused exploration of the variables most informative in the context of student engagement.

VIF (Variance Inflation Factor) Analysis: VIF analysis stood out as a crucial component of the feature selection process, specifically geared toward assessing multicollinearity among predictor variables. Multicollinearity, the condition where predictor variables in a regression model are highly correlated, can impede the model's stability and interpretability. The VIF analysis, offering insights into the extent of correlation between each predictor and other variables, played a significant role in discerning potential issues.

The VIF factors were computed for key variables in the study, including "study_hours", "courses", "hours", "class_strength", "teaching", "technology", and "material". Conventionally, a VIF factor less than 5 is deemed acceptable, signifying that the variance of a predictor is not excessively inflated by correlations with other predictors. The analysis revealed that the VIF for "technology" and "material" exceeded the threshold, indicating a potential concern about multicollinearity.

Addressing multicollinearity is paramount for ensuring the robustness of the regression model. When predictors are highly correlated, it becomes challenging to isolate the individual effects of each variable on the dependent variable. Therefore, thoughtful consideration and potential adjustments in the modeling process may be necessary to mitigate the impact of correlated predictors on the stability and interpretability of the model.

Heatmap: The heatmap visualization served as a complementary tool in the feature selection toolkit, providing an intuitive representation of the correlation matrix among predictor variables. Darker shades in the heatmap signified higher correlation coefficients, allowing for quick visual identification of variables with strong interrelationships. In the context of this study, the heatmap was instrumental in corroborating and enhancing the insights gained from the VIF analysis.

The heatmap, presented in Fig. 4 portrayed the correlations between key variables. Notably, class strength exhibited negative correlations with marks, study hours, courses, and teaching. This observation sheds light on the potential detrimental effects of larger class sizes on academic performance, study hours, course engagement, and teaching effectiveness. On the other hand, teaching demonstrated a strong positive correlation with

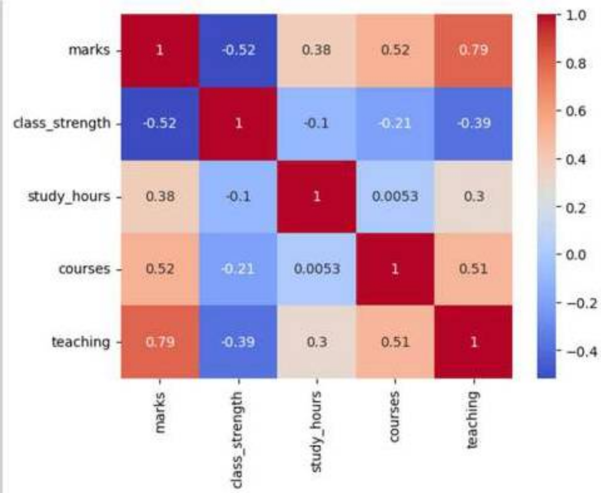

Fig. 4 : Heatmap

marks, suggesting that effective teaching methods could significantly enhance academic performance.

Together, the VIF analysis and heatmap visualization formed a comprehensive feature selection strategy. The VIF analysis provided quantitative insights into multicollinearity, while the heatmap offered a qualitative, visual representation of inter-variable correlations. The synergy between these techniques guided the refinement of the model, ensuring a judicious selection of features that not only adhered to statistical criteria but also resonated with the contextual understanding of the factors influencing student engagement.

5. Implementation

The implementation phase of this study represents a pivotal transition from meticulous preparation to the empirical exploration of the factors influencing student engagement. Guided by the chosen methodologies, including Multiple Linear Regression, Random Forest, Lasso, Ridge Regression, and Ordinary Least Squares (OLS) regression, this phase engages with the curated dataset to derive insights that hold the potential to reshape educational practices.

A. Dataset Integration and Preliminary Analysis

Before delving into the regression analysis, the first step in implementation involves the seamless integration of the cleaned and transformed dataset. This dataset, a comprehensive tapestry of student-related variables ranging from demographic information to study habits, now stands poised for examination. Preliminary analyses, including descriptive statistics and exploratory data visualizations, serve as the precursor to the regression analysis. These initial insights illuminate the characteristics of the dataset, offering a contextual backdrop for the ensuing analysis.

B. Regression Analysis: Unveiling Dynamics

The crux of the implementation lies in the application of various regression models to determine the most effective approach for understanding the dynamics between various factors and student engagement levels.

1) Models Employed

In the pursuit of understanding the intricate dynamics influencing student engagement, this study employed a suite of advanced regression models, each offering unique insights and analytical strengths. By leveraging multiple linear regression, bayesian ridge regression, lasso regression, random forest regression, and ordinary least squares (OLS) regression, the study aimed to comprehensively capture the relationships between critical educational variables and student engagement. These models were meticulously chosen to address different aspects of the data, from linear relationships and variable selection to handling multicollinearity and capturing non-linear interactions. The following sections detail the specification, estimation, and comparison of these models, ultimately leading to the identification of the most effective approach for predicting student engagement.

Multiple Linear Regression: Multiple Linear Regression is a foundational statistical technique that models the linear relationship between one dependent variable (student engagement) and multiple independent variables (class size, teaching style, technology use, and class hours). The aim is to understand how each independent variable influences the dependent variable when the other variables are held constant.

R-squared: 0.7388: This indicates that approximately 73.88% of the variance in student engagement can be explained by the independent variables included in the model. Although this is a significant proportion, it suggests that there are still other factors influencing student engagement that are not captured by this model.

Bayesian Ridge Regression: Bayesian Ridge Regression is an extension of linear regression that incorporates Bayesian principles. It applies a prior distribution to the regression coefficients, which helps in mitigating overfitting, especially in the presence of multicollinearity (when independent variables are highly correlated).

R-squared: 0.7660: The R-squared value here indicates that 76.60% of the variance in student engagement is explained by the model. The slight improvement over multiple linear regression suggests that Bayesian Ridge Regression effectively handles some of the complexities and correlations within the data.

Lasso Regression: Lasso Regression (Least Absolute Shrinkage and Selection Operator) is a type of linear regression that employs L1 regularization. This technique not only fits the model but also shrinks the coefficients of less important features to zero, effectively performing variable selection and thus simplifying the model.

R-squared: 0.7659: The R-squared value shows that 76.59% of the variance in student engagement is explained by the independent variables. Similar to Bayesian Ridge, Lasso Regression handles the dataset's complexity by reducing the influence of less important variables, leading to a slightly more interpretable model.

Random Forest Regression: Random Forest Regression is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction of the individual trees. This approach helps improve the model's accuracy and stability by reducing overfitting and variance.

R-squared: 0.7000: The R-squared value indicates that 70.00% of the variance in student engagement is explained by the model. Although this is lower than the other models, Random Forests are beneficial for capturing non-linear relationships and interactions between variables, providing a different perspective on the data.

Ordinary Least Squares (OLS) Regression: Ordinary Least Squares (OLS) Regression is a classical linear regression method that estimates the relationships between variables by minimizing the sum of the squared differences between the observed and predicted values. It is known for its simplicity and interpretability.

R-squared: 0.9520: This high R-squared value indicates that 95.20% of the variance in student engagement is explained by the independent variables in the model. This suggests that OLS regression provides the best fit for the data among the models considered, making it the most effective method for this study.

2) Model Specification

The heart of the implementation lies in the meticulous specification of the OLS regression model. The dependent variable, student engagement, is intricately linked to a set of carefully selected independent variables: class size, teaching style, technology use, and class hours. The model is expressed as:

$$\text{Student\_Engagement} = \beta 0 + \beta 1 \times \text{Class\_Size} + \beta 2 \times \text{Teaching\_Style} + \beta 3 \times \text{Technology\_Use} + \beta 4 \times \text{Class\_Hours} + \epsilon$$

Where,

Student_Engagement is the dependent variable.

$\beta 0$ is the intercept.

$\beta 1, \beta 2, \beta 3, \beta 4$ are the coefficients for class size, teaching style, technology use, and class hours, respectively.

ϵ is the error term.

3) Model Estimation and Comparison

Each model was employed to estimate the coefficients that best explain the variation in student engagement. The R-squared values of these models were compared to determine the effectiveness of each.

Table 1:
Results of various models employed
Model Interpretation

Models employed	R-squared values
Multiple Linear Regression	0.7388
Bayesian Ridge Regression	0.7660
Lasso Regression	0.7659
Random Forest Regression	0.7000
OLS Regression	0.9520

The OLS model demonstrated the highest R-squared value, indicating that it explained the most variance in student engagement compared to the other models.

Interpreting the OLS regression output is crucial for extracting meaningful insights. The coefficients (β) quantify the impact of each independent variable on student engagement. A positive coefficient indicates a positive relationship, while a negative coefficient suggests a negative impact. The magnitude of the coefficients gauges the strength of the influence.

C. Validation and Refinement

The validity of the model is assessed through various diagnostics, including checking for multicollinearity, heteroscedasticity, and normality of residuals. These diagnostic steps ensure the robustness of the model and enhance its reliability for drawing meaningful conclusions.

D. Practical Implications and Recommendations

As the regression analysis unfolds, the findings unearthed bear profound implications for educational practitioners and policymakers. A nuanced understanding of how class size, teaching style, technology use, and class hours intertwine with student engagement becomes a compass for informed decision-making.

1. Educational Practices

Insights derived from the regression analysis empower educators with evidence-based guidance. The impact of teaching style on engagement, the optimal class size for fostering student interaction, and the judicious integration of technology all become tangible considerations in refining pedagogical approaches.

2. Policy Decisions

Policymakers, grappling with decisions on class size limits, teacher training initiatives, and technology investments, find solace in the empirical evidence presented by the regression results. The study's implications extend beyond individual classrooms, informing systemic changes that resonate across educational levels.

6. Discussion On Result

The Ordinary Least Squares (OLS) regression model serves as the analytical linchpin in unraveling the complex relationship between various factors and student engagement, represented by academic 'marks'. The results obtained from the rigorous analysis offer a wealth of insights, shedding light on the critical determinants that significantly impact student engagement levels.

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                  marks   R-squared (uncentered):                   0.952
Model:                            OLS   Adj. R-squared (uncentered):              0.951
Method:                 Least Squares   F-statistic:                              1066.
Date:                Sun, 30 Apr 2023   Prob (F-statistic):                   2.18e-141
Time:                        11:12:29   Log-Likelihood:                         -1388.4
No. Observations:                 221   AIC:                                      2785.
Df Residuals:                     217   BIC:                                      2798.
Df Model:                           4
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
study_hours       36.1974      6.024      6.009      0.000      24.325      48.070
courses           62.6510     22.794      2.749      0.006      17.724     107.578
class_strength     7.1757      0.281     25.539      0.000       6.622       7.729
teaching         164.5034     21.992      7.480      0.000     121.158     207.849
==============================================================================
Omnibus:                        0.719   Durbin-Watson:                   1.984
Prob(Omnibus):                  0.698   Jarque-Bera (JB):                0.433
Skew:                          -0.054   Prob(JB):                        0.805
Kurtosis:                       3.188   Cond. No.                         175.
==============================================================================
```

Fig. 6 : OLS regression results

A. Model Fit and Robustness:

The statistical metrics portraying the model's fit and robustness showcase a promising foundation. The R-squared value of 0.952 indicates that approximately 92% of the variability in student engagement can be explained by the chosen predictors. This signifies a robust model that captures a substantial portion of the intricate dynamics influencing academic performance.

$$R^2 = 1 - \frac{SST}{SSE} \tag{7}$$

The Adjusted R-squared value of 0.951 reinforces the model's strength, adjusting for the number of predictors and providing a more accurate representation of the true relationship. These high values instill confidence in the model's ability to generalize well to new data.

$$R^2{}_{adj} = 1 - \frac{(1 - R^2)(n-1)}{n-k-1} \tag{8}$$

The coefficients for each predictor offer insights into their individual impact on student engagement:

Study Hours: The coefficient associated with study hours is 36.1974, implying that each additional hour of study corresponds to an increase of approximately 36.2 marks. This result establishes a positive correlation between study hours and student engagement, underscoring the pivotal role of dedicated study time in academic performance. The implication is that students who invest more time in focused study are likely to achieve higher levels of engagement and success in their academic endeavors.

Courses: For the variable 'courses,' the coefficient is 62.6510, indicating that participation in additional courses contributes around 62.7 marks. This interpretation emphasizes the positive impact of enrolling in supplemental courses, highlighting their value in enhancing both student engagement and overall academic performance. The implication is that students who diversify their learning experiences through additional courses tend to achieve higher academic success.

Class Strength: The coefficient for class strength is 7.1757, suggesting that a unit decrease in class strength is associated with a rise of 7.2 marks. This finding implies a positive correlation between smaller class sizes and increased student engagement, aligning with existing literature on class size and educational outcomes. The implication is that students in smaller classes may experience a more engaged and supportive learning environment.

Teaching Style: In the context of teaching, the coefficient is 164.5034, indicating a significant influence on student engagement. Each unit change in teaching style contributes 164.5 marks, underscoring the pivotal role of effective teaching methods in fostering student engagement and success. This implies that educators employing impactful teaching styles positively influence student engagement, leading to improved academic outcomes.

B. Statistical Significance

The overall F-statistic for the model is 1066, with a p-value of 2.18e-141. This statistical significance indicates that the chosen predictors collectively have a substantial impact on student engagement. The implication is that the model is a valid and reliable representation of the relationships between the selected variables and student engagement.

$$F = \frac{\frac{SSR}{K}}{\frac{SSE}{n-k-1}} \qquad (9)$$

The Durbin-Watson statistic, with a value of 1.984, suggests no significant autocorrelation in the model's residuals. This result enhances the reliability of the model by indicating that the independence assumption of OLS is not violated.

Additionally, the Omnibus and Jarque-Bera statistics, with values of 0.719 and 0.433, respectively, suggest that residuals follow a normal distribution. These relatively small statistics reinforce the assumption of normality in the residuals, further validating the robustness of the model.

The OLS regression results suggest that study hours, participation in additional courses, class strength, and teaching style are significant factors influencing student engagement and subsequent academic performance. Educators and policymakers can leverage these findings to tailor interventions and strategies that enhance these key determinants, ultimately improving overall student outcomes.

The pie chart in Fig.6 presents a visual representation of the influential elements affecting students' academic outcomes. The chart delineates four primary factors: "study hours," "courses," "class strength," and "teaching style," each contributing to the overall academic performance.

The data reveals that "teaching style" has the most substantial influence, emphasizing the critical role of effective teaching methodologies in fostering student engagement and success. This finding underscores the importance of pedagogical approaches in shaping educational outcomes and suggests a need for continuous innovation in teaching practices to optimize student learning experiences.

The chart highlights the contributions of "study hours" and "courses" to academic achievement, indicating the significance of dedicated study time and diversified learning experiences in enhancing student outcomes. Conversely, the diminished representation of "class strength" implies that an increase in class size may correspond to a decrease in academic marks, emphasizing the intricate relationship between class dynamics and student performance, warranting further examination.

Conclusion

Based on the analysis conducted in this study, it is evident that several factors significantly impact student engagement. The Ordinary Least Squares (OLS) regression analysis identified that smaller class sizes are associated with higher levels of engagement, as they enable more personalized interactions and support. Effective teaching styles that are innovative and interactive also play a crucial role in enhancing student engagement. The use of technology in the classroom further boosts student interest and participation.

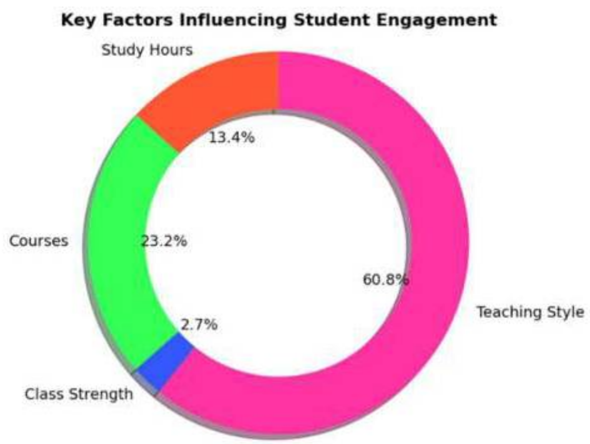

Fig . 7 : Factors influencing Student Engagement

Interestingly, the number of class hours did not show a significant influence on engagement, suggesting that the quality of instructional methods is more critical than the quantity of time spent in class.

These findings highlight the need for educators to focus on optimizing class sizes, adopting dynamic teaching methodologies, and integrating technology to foster a more engaging learning environment. Future research could explore these relationships in different educational contexts and among diverse student populations to develop more tailored strategies for improving engagement. By prioritizing these factors, educational institutions can enhance student learning experiences, leading to better academic outcomes and higher retention rates. This study provides a valuable framework for understanding and improving student engagement through strategic educational practices.

References

Rashid, T., & Asghar, H. M. (2016). Technology use, self- directed learning, student engagement and academic performance: Examining the interrelations. Computers in human behavior, 63, 604-612.

Delfino, A. P. (2019). Student engagement and academic performance of students of Partido State University. Asian Journal of University Education, 15(1), n1.

Dowling §, C., Godfrey §, J. M., & Gyles, N. (2003). Do hybrid flexible delivery teaching methods improve accounting students' learning outcomes?. Accounting Education, 12(4), 373-391.

Chen, P. S. D., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners: The impact of Web-based learning technology on college student engagement. Computers & Education, 54(4), 1222-1232.

Halif, M. M., Hassan, N., Sumardi, N. A., Omar, A. S., Ali, S., Aziz, R. A., ... & Salleh, N. F. (2020). Moderating effects of student motivation on the relationship between learning styles and student engagement. Asian Journal of University Education, 16(2), 93-103.

Alshanqiti, A., & Namoun, A. (2020). Predicting student performance and its influential factors using hybrid regression and multi-label classification. IEEE Access, 8, 203827-203844.

Douglas, K. A., Merzdorf, H. E., Hicks, N. M., Sarfraz, M. I., & Bermel, P. (2020). Challenges to assessing motivation in MOOC learners: An application of an argument-based approach. Computers & Education, 150, 103829.

Eltahir, M. E., Alsalhi, N. R., Al-Qatawneh, S., AlQudah, H. A., & Jaradat, M. (2021). The impact of game-based learning (GBL) on students' motivation, engagement and academic performance on an Arabic language grammar course in higher education. Education and Information Technologies, 26, 3251-3278.

Gray, J. A., & DiLoreto, M. (2016). The effects of student engagement, student satisfaction, and perceived learning in online learning environments. International Journal of Educational Leadership Preparation, 11(1), n1.

Bun, Maurice JG, and Teresa D. Harrison. "OLS and IV estimation of regression models including endogenous interaction terms." Econometric Reviews 38.7 (2019): 814-827.

Terziev, Venelin. "Factors influencing education system." 50th International Scientific Conference on Economic and Social Development–Chelyabinsk. 2020.

Shah, Mahsood, Chenicheri Sid Nair, and Lorraine Bennett. "Factors influencing student choice to study at private higher education institutions." Quality Assurance in Education 21.4 (2013): 402-416.

Jabeen, Shazi Shah, and Aqila Rafiuddin. "Factors influencing the education decision making process." Second 21st Century Academic Forum at Harvard. Vol. 5. No. 1. 2015.

Parakh, S., Nalwade, Y., & Naresh, E. (2023). An Empirical Investigation of Engineering Students' Attitute Towards Sports and Physical Education, Cultural Activities and Stress Management. Journal of Engineering Education Transformations, 37(2), 7–16. Scopus. https://doi.org/10.16920/jeet/2023/v37i2/23145

Suchitra, M., Vivek D'Souza, A., Vinay, B. K., & Panchami, S. V. (2022). Activity based teaching learning in Engineering Education, a course on: Embedded Systems. Journal of Engineering Education Transformations, 36(Special Issue 2), 277–282. Scopus. https://doi.org/10.16920/jeet/2023/v36is2/23040

Affia Thabassum, M. S., Mohammed Thaha, M., Rajendran, N., & Abudhahir, A. (2022). Enhancing Attainment of Learning Outcomes Through Active Learning. Journal of Engineering Education Transformations, 36(Special Issue 1), 94–103. Scopus.

Alva, H. H., Uma, B., Shruthi, D. V., & Saroja, C. (2018). Enhancing learning outcomes in software engineering course through problem based learning and peer assisted learning. Journal of Engineering Education Transformations, 2018(Special Issue). Scopus. https://doi.org/10.16920/jeet/2018/v0i0/120948

Beena, B. R., & Suresh, E. S. M. (2022a). Analysis of learning outcomes of Civil Engineering students of Kerala state using dimension reduction Techniques. Journal of Engineering Education Transformations, 35(Special Issue 1), 120–128. Scopus.

Bharadia, B., Pandya, K., & Majethia, K. (2018). Improvement of student learning outcomes by implementation of model making approach for student evaluation. Journal of Engineering Education Transformations, 31(3), 187–191. Scopus.

Chandrasekaran, J., Anitha, D., & Uma, K. V. (2023). An Effective Instructional Design to Enhance Learning Outcomes of Information Security Course in Online Mode. Journal of Engineering Education Transformations, 36(Special Issue 2), 319–325. Scopus. https://doi.org/10.16920/jeet/2023/v36is2/23047

Ganesh, K. E., & Pranesha, T. S. (2018). Enhancement of learning outcomes through implementation of best practices in teaching learning process: A case study. Journal of Engineering Education Transformations, 32(1), 12–14. Scopus. https://doi.org/10.16920/jeet/2018/v0i0/120888

Hegde, R., Prasanna Kumar, M. K., & Oppiliappan, G. (2020). Learning outcomes of a national level project contest. Journal of Engineering Education Transformations, 34(2), 54–60. Scopus. https://doi.org/10.16920/jeet/2020/v34i2/146590