South American Publishing

SAP Multidisciplinary Open. 2025; 3:51
doi: 10.62486/mo202551

ORIGINAL

Factors influencing academic performance: an empirical study using predictive analytics

Factores que influyen en el rendimiento académico: Un estudio empírico utilizando análisis predictivos

Farheen Islam[1] ✉, Aprajita Krishna[2] ✉, Devanshu Kumar[3] ✉, Sangeeta Kumari[4] ✉

[1]Department of Education, Patna Women's College. Patna, India.
[2]Babasaheb Bhimrao Ambedkar Bihar University. Muzaffarpur, Bihar, India.
[3]Department of Computer Science, V. K. S. University. Ara, India.
[4]School of Education, Nalanda Open University. Nalanda, India.

Cite as
Islam F, Krishna A, Kumar D, Kumari S. Factors Influencing Academic Performance: An Empirical Study Using Predictive Analytics. SAP Multidisciplinary Open. 2025; 3:51. https://doi.org/10.62486/mo202551

Publication History
Submitted: 10-03-2024
Revised: 13-06-2024
Accepted: 26-10-2024
Published: 01-01-2025

Corresponding Author
Farheen Islam ✉

ABSTRACT

Introduction: student academic performance is commonly measured using indicators like credit scores, class rankings, or passing thresholds, each offering insight into a student's comprehension and progress. These measures give a numeric evaluation of a student's understanding of the course content, enabling educators and institutions to monitor growth over time. Moreover, the capacity to forecast academic performance proves highly beneficial for both students and teachers, as it enables early interventions and course corrections. This predictive ability can lead to tailored learning strategies, ultimately enhancing educational outcomes and supporting overall student success.
Method: to carry out a statistical analysis on the dataset (kaggel), the first step will involve cleaning and preprocessing the data to address any missing or non-numeric entries. Once the dataset is refined, hypothesis testing and ANOVA will be employed to pinpoint the key factors that influence student performance.
Result: initial findings suggest that study hours and attendance are significant predictors of exam scores, with higher values generally leading to better performance. Additionally, motivational factors, parental involvement, and access to academic resources showed varying degrees of influence, highlighting their potential role in shaping educational outcomes.
Conclusion: through descriptive statistics, correlation analysis, and regression models, the dataset offers insights that can be used by educators and policymakers to enhance teaching strategies and student support systems. By identifying key factors that affect student outcomes, the dataset provides a foundation for future research and practical interventions aimed at improving academic performance in educational settings.

Keywords: Education; Performance; Statistics; Strategies.

RESUMEN

Introducción: el rendimiento académico de los estudiantes suele medirse mediante indicadores como la puntuación de los créditos, la clasificación de la clase o los umbrales de aprobación, cada uno de los cuales ofrece una visión de la comprensión y el progreso de un estudiante. Estas medidas proporcionan una evaluación numérica de la comprensión de los contenidos del curso por parte del alumno, lo que permite a los educadores y a las instituciones supervisar su crecimiento a lo largo del tiempo. Además, la capacidad de predecir el rendimiento académico resulta muy beneficiosa tanto para los alumnos como para los profesores, ya que permite realizar intervenciones tempranas y correcciones del curso. Esta capacidad de predicción puede dar lugar a estrategias de aprendizaje personalizadas que, en última instancia, mejoren los resultados educativos y contribuyan al éxito general de los estudiantes.
Método: para llevar a cabo un análisis estadístico del conjunto de datos (kaggel), el primer paso consistirá en limpiar y preprocesar los datos para corregir las entradas que falten o no sean numéricas. Una vez depurado el conjunto de datos, se realizarán pruebas de hipótesis y ANOVA para determinar los factores clave que influyen en el rendimiento de los estudiantes.
Resultados: los resultados iniciales sugieren que las horas de estudio y la asistencia son predictores significativos de las puntuaciones de los exámenes, y que los valores más altos suelen conducir a un mejor rendimiento. Además, los factores motivacionales, la participación de los padres y el acceso a los recursos académicos mostraron diversos grados de influencia, lo que pone de relieve su papel potencial en la configuración de los resultados educativos.

© 2025; The Authors. This Open access article is distributed under the Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conclusiones: mediante estadísticas descriptivas, análisis de correlación y modelos de regresión, el conjunto de datos ofrece información que puede ser utilizada por educadores y responsables políticos para mejorar las estrategias de enseñanza y los sistemas de apoyo a los estudiantes. Al identificar los factores clave que influyen en los resultados de los alumnos, el conjunto de datos proporciona una base para futuras investigaciones e intervenciones prácticas encaminadas a mejorar el rendimiento académico en entornos educativos.

Palabras clave: Educación; Rendimiento; Estadísticas; Estrategias.

INTRODUCTION

It is essential to understand that the entirety of a person's existence is often shaped by the knowledge they acquire and how effectively they use this knowledge to improve themselves, their community, and the broader world. This underscores the fundamental importance of education in human life.[1] Education is not merely about acquiring facts but is a transformative process that provides individuals with the tools to shape their future and contribute to society. It serves as the foundation upon which personal and societal growth is built. At its core, the most fundamental benefit of education is knowledge.[2] Through education, individuals are exposed to a wide array of subjects and disciplines that expand their understanding of the world.[3] They learn about diverse topics, from the logical structures of Mathematics to the lessons of History, the creativity of Literature to the complexities of Political Science. Each subject plays a unique role in shaping a well-rounded individual, helping them develop critical thinking, problem-solving abilities, and a deeper understanding of the world around them. By gaining knowledge in these areas, individuals are better equipped to tackle the challenges of life, make informed decisions, and contribute meaningfully to society.

The knowledge we gain through education also significantly influences our future. It provides us with the skills necessary to navigate the complexities of modern life, from understanding financial systems to grasping political structures and societal dynamics. More importantly, education empowers individuals to interpret and make sense of current events and the changes happening in the world. A well-educated person can approach global issues—whether environmental, economic, or social—with a critical eye and a deeper understanding, which allows them to participate actively in finding solutions. Furthermore, education is not just about acquiring knowledge but also about applying that knowledge in meaningful ways. When people utilize their education to innovate, solve problems, and create new ideas, they contribute to the development of their communities, nations, and the world at large.[4] For instance, someone with a background in science might work on solutions for global health issues, while someone educated in political science could contribute to improving governance systems. Success in education is commonly measured by academic performance, which reflects how well students absorb and apply the knowledge they gain. While academic success is often associated with grades and test scores, it is also about how well students understand concepts and their ability to use that knowledge in practical, real-world situations. A student's ability to perform academically is a key indicator of their preparedness to enter the workforce and contribute to society effectively.[5]

Student academic performance is often evaluated through various metrics such as credit scores, class rankings, or passing criteria, all of which serve to indicate a student's level of learning and overall progress.[6] These metrics provide a quantitative assessment of how well students grasp the subject matter, helping educators and institutions track student development. Beyond this, the ability to predict academic performance is an invaluable tool for both students and educators as it allows for timely interventions and adjustments, which can significantly improve learning outcomes.[7] For students, predicting academic performance plays a crucial role in self-assessment and supervision. By understanding potential outcomes in advance, students can receive an early warning if their performance is not on track. This foresight empowers them to make necessary changes to their study habits, time management, and overall approach to learning.[8] For instance, if a student is predicted to struggle in a particular subject, they can seek additional resources, tutoring, or guidance before it is too late. As a result, this proactive approach helps minimize the risk of failure, allowing students to better control their academic journey and increase their chances of success.

Teachers also benefit significantly from academic performance prediction, as it allows them to tailor their teaching strategies to better suit the needs of their students.[9] If a teacher sees that many students are predicted to perform poorly in a subject, they can modify their teaching techniques, focusing more on areas where students struggle or offering more personalized support. For example, the teacher might incorporate more interactive lessons, provide additional practice materials, or use different instructional methods to cater to various learning styles. This adaptability in teaching methods ensures that students receive the support they need, ultimately improving their chances of passing and mastering the subject matter.[4] Similarly, academic counselors can leverage performance predictions to provide better support for students, particularly those at risk of failing.[10] By identifying struggling students early on, counselors can intervene with targeted advice, whether it's guiding students toward specific study resources, helping them manage their workload, or addressing personal issues that may be affecting their academic performance.[11] This early intervention can make a significant difference, helping students overcome obstacles before they lead to failure. Furthermore, counselors can monitor the overall performance trends of the student body and offer guidance to ensure academic success on a broader scale.[12]

At an institutional level, performance predictions provide valuable insights that can influence curriculum design and student training programs. Higher education institutions can analyze performance data to identify trends, such as which courses or teaching methods result in higher failure rates. This information enables institutions to make data-driven decisions about how to modify their

teaching methods, curriculums, and student support systems to improve student outcomes. For instance, if a course consistently results in low performance, the institution might offer additional workshops or revise the curriculum to make the material more accessible.

The rest of the manuscript is organized as follows: in "Methodology" section, related works are discussed, followed by an analysis of the influencing factors and the presentation of the "Student Performance Factors" section. "Experimental Analysis" section discusses the experimental analysis and presents a comparative assessment, and the manuscript is concluded in "Conclusion" section.

METHOD

Data Sources

Publicly available datasets like the one provided by Kaggle[13] (https://www.kaggle.com/datasets/lainguyn123/student-performance-factors/data)

This dataset from Kaggle offers a detailed analysis of the diverse factors that contribute to students' performance in examinations. It captures data on key elements such as study routines, class attendance, parental participation, and other variables that significantly affect academic achievement. By providing these insights, the dataset allows for a broader understanding of how these influences interplay and impact student success. One of the most important elements included in the dataset is study habits. Consistent study routines are often seen as critical to academic success, and this dataset helps highlight the different approaches students take toward studying. It tracks various patterns, such as time management, revision practices, and preferred learning methods, providing valuable insights into how these habits may contribute to better exam results.

This dataset is particularly valuable for researchers, educators, and policymakers who are focused on improving educational outcomes. By analyzing the data, patterns can be identified, interventions can be designed, and targeted support systems can be developed to address specific needs. For instance, if a strong correlation between parental involvement and exam success is observed, schools could promote initiatives that encourage more parental participation in their children's education.

The dataset contains 6,607 entries, with each entry representing a student and their corresponding academic performance measured by exam scores. The dataset includes 20 columns covering various attributes such as:

- Hours_Studied: Number of hours a student spends studying.
- Attendance: Percentage of attendance.
- Parental_Involvement: Levels of involvement (Low, Medium, High).
- Access_to_Resources: Whether the student has access to academic resources (High, Medium, Low).
- Extracurricular_Activities: Participation in extracurricular activities (Yes, No).
- Motivation_Level: Student's self-reported motivation level (Low, Medium, High).
- Family_Income: Income level of the student's family (Low, Medium, High).
- Teacher_Quality: Perception of the quality of teaching (Low, Medium, High).
- Learning_Disabilities: Presence of any learning disabilities (Yes, No).
- Gender: Gender of the student.
- Exam_Score: Final exam score.

The dataset uses a combination of categorical and numerical variables, which allows for both qualitative and quantitative analysis of student performance.

Data Analysis

To perform statistical analysis on the dataset, I will first clean and preprocess the data to handle any missing or non-numeric values, then conduct hypothesis tests and ANOVA to identify significant factors affecting student performance, and finally visualize the results to interpret the findings. Let's proceed with this plan.

Based on our statistical analysis, we've found:

1. No missing values in the dataset.
2. T-test for Internet Access:
   T-statistic: 4.08756319841647
   P-value: 4.4128283860873265e-05

The low p-value (< 0.05) suggests a significant difference in exam scores between students with and without internet access.

1. ANOVA for School Type:
   F-statistic: 0.7531963873423186
   P-value: 0.3854987810260675

The high p-value (> 0.05) indicates no significant difference in exam scores between public and private schools.

Distribution of Exam Scores

The histogram of exam scores shows a roughly normal distribution with a slight left skew, indicating that most students scored around the average, with fewer students achieving very high or very low scores.

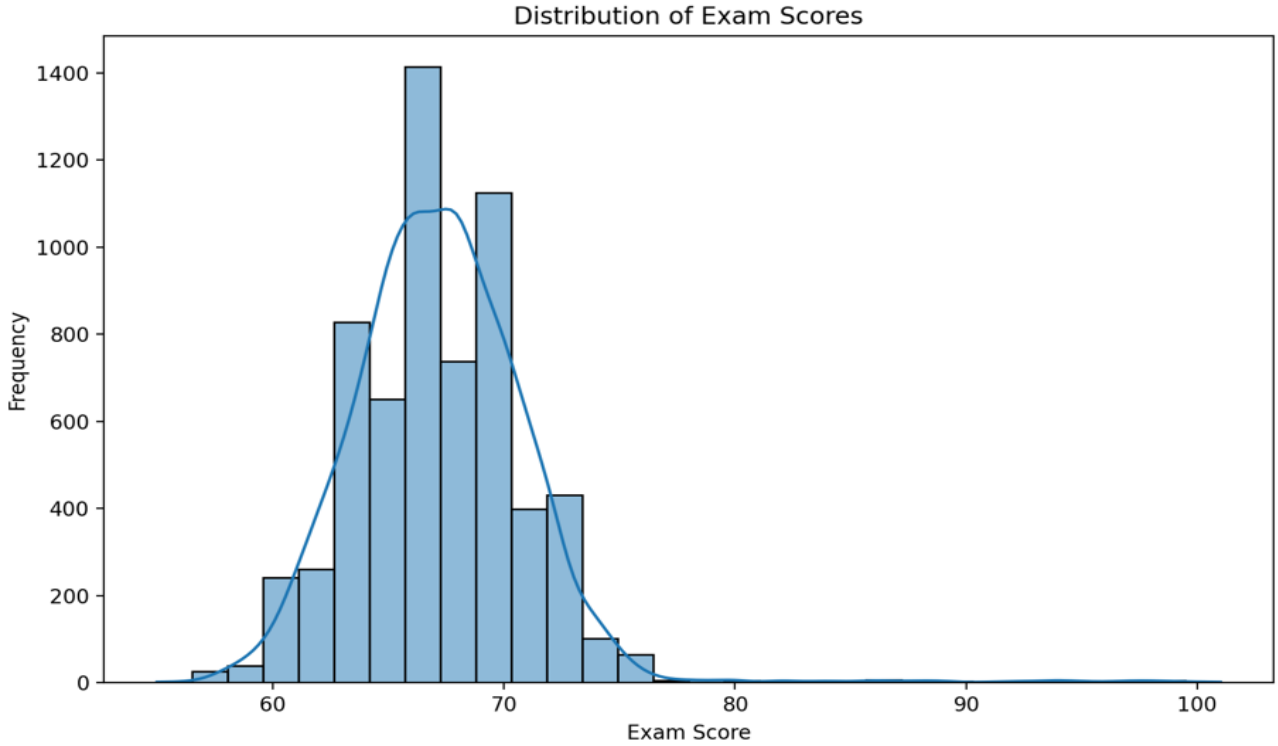


Figure 1. Distribution of Exam Scores[13]

Hours Studied vs. Exam Score

This scatter plot illustrates the relationship between the number of hours studied and the exam scores, with points colored by gender. It helps to visualize if more study hours correlate with higher exam scores.

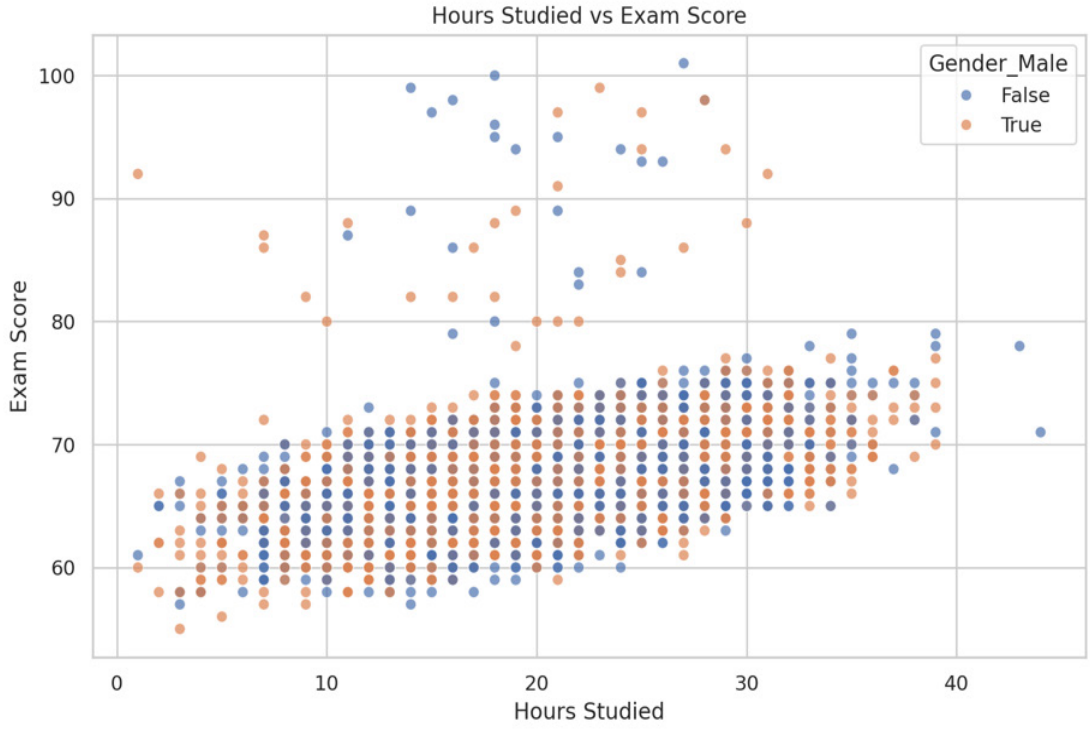


Figure 2. Hours Studied vs Exam Scores[13]

Exam Score by Attendance Level

This box plot displays the exam scores categorized by attendance levels, providing insights into how attendance impacts performance.

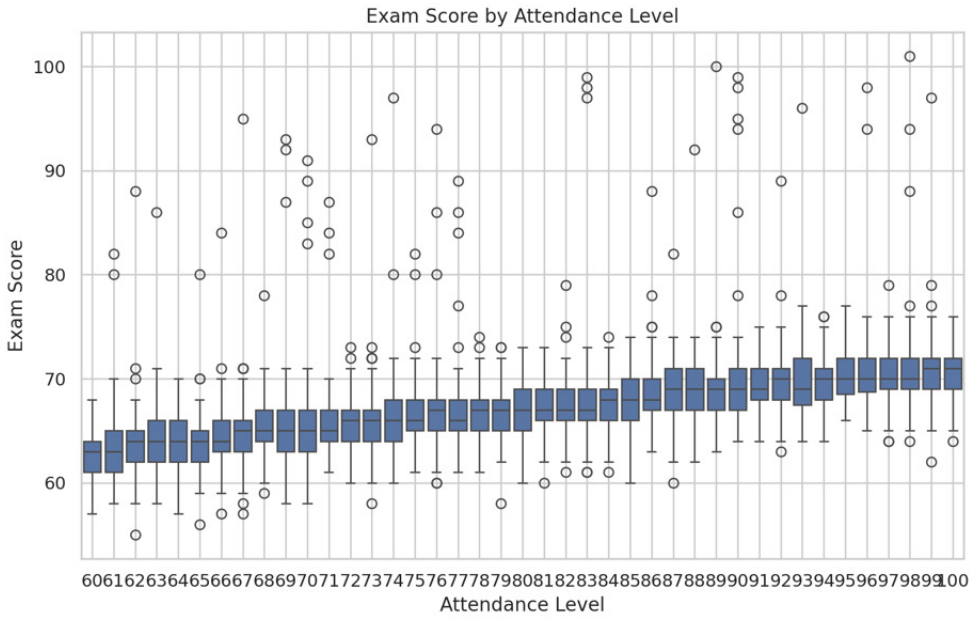


Figure 3. Exam Score by Attendance Level[13]

RESULTS AND DISCUSSION

Initial analysis indicates that certain factors, such as hours studied, attendance, and previous scores, show a significant correlation with exam performance. Students who studied more hours and had better attendance generally performed better. Interestingly, variables such as motivation level and access to resources also showed a strong relationship with exam scores. For instance, students with high motivation and better access to academic resources tended to have higher scores. However, the data also suggests that other factors like parental involvement and family income had a more mixed impact on performance. While high parental involvement was often associated with better outcomes, this was not universally true, implying that other environmental or personal factors may play a role.

CONCLUSION

The analysis emphasizes the crucial role that study habits and class attendance play in shaping student performance. Although the time spent studying and regular attendance are key contributors to academic success, the variation in student scores suggests that these factors alone do not fully explain the differences in outcomes. Other aspects, such as a student's motivation, the level of parental involvement, and access to educational resources, likely also have a significant influence on performance. These elements may create conditions that either enhance or limit a student's ability to excel academically. To gain a more comprehensive understanding of what drives student success, future research should explore these additional factors, as doing so could provide deeper insights into the complexities of academic achievement.

AVAILABILITY OF DATA AND MATERIALS

The datasets used in this research are publicly available and properly cited in our dataset section for transparency and ease of replication. (https://www.kaggle.com/datasets/lainguyn123/student-performance-factors/data)

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTIONS

Conceptualization: Farheen Islam, Aprajita Krishna, Devanshu Kumar and Sangeeta Kumari.
Investigation: Farheen Islam, Aprajita Krishna, Devanshu Kumar and Sangeeta Kumari.
Methodology: Farheen Islam, Aprajita Krishna and Sangeeta Kumari.
Writing - original draft: Farheen Islam, Aprajita Krishna and Sangeeta Kumari.
Writing - review and editing: Farheen Islam, Aprajita Krishna, Devanshu Kumar and Sangeeta Kumari.

REFERENCES

1. Haque MA, Ahmad S, Hossain MA, Kumar K, Faizanuddin M, Islam F, et al. Internet of things enabled E-learning system for academic achievement among university students. E-Learning Digit Media. 2024; 20427530241280080.

2. Namoun A, Alshanqiti A. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Appl Sci. 2020;11(1):237.

3. Duch D, May M, George S. Enhancing Predictive Analytics for Students' Performance in Moodle: Insight from an Empirical Study. J Data Sci Intell Syst. 2024;

4. Kocsis Á, Molnár G. Factors influencing academic performance and dropout rates in higher education. Oxford Rev Educ. 2024;1–19.

5. Haque MA, Ahmad S, Haque S, Kumar K, Mishra K, Mishra BK. Analyzing University Students' Awareness of Cybersecurity. In: 2023 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC). IEEE; 2023. p. 250–7.

6. Haque S, Haque MA, Kumar D, Mishra K, Islam F, Ahmad S, et al. Assessing the Impact of IoT Enabled E-Learning System for Higher Education. SN Comput Sci [Internet]. 2023;4(5):459. Available from: https://doi.org/10.1007/s42979-023-01860-8

7. Haque MA, Haque S, Zeba S, Kumar K, Ahmad S, Rahman M, et al. Sustainable and efficient E-learning internet of things system through blockchain technology. E-Learning Digit Media [Internet]. 2023;0(0):1–20. Available from: https://journals.sagepub.com/doi/abs/10.1177/20427530231156711

8. D'Uggento AM, d'Ovidio FD, Toma E, Ceglie R. A framework for detecting factors influencing students' academic performance: A longitudinal analysis. Soc Indic Res. 2021;156(2):389–407.

9. Alimul Haque, Deepa Sonal, Shameemul Haque, Moidur Rahman KK. Learning Management System Empowered by Machine Learning. In AIP Publishing; 2021.

10. Prakash A, Haque A, Islam F, Sonal D. Exploring the Potential of Metaverse for Higher Education: Opportunities, Challenges, and Implications. Metaverse Basic Appl Res [Internet]. 2023 Apr 26;2(SE-Reviews):40. Available from: https://mr.saludcyt.ar/index.php/mr/article/view/40

11. Khanna L, Singh SN, Alam M. Educational data mining and its role in determining factors affecting students academic performance: A systematic review. In: 2016 1st India international conference on information processing (IICIP). IEEE; 2016. p. 1–7.

12. Alimul Haque * , Devanshu Kumar , Khushboo Mishra , Farheen Islam BKM. The Factor Affecting Student's Performance of E-Learning Environment Using Machine Learning Algorithm. In: MOL2NET'22, Conference on Molecular, Biomedical & Computational Sciences and Engineering, 8th ed congress USEDAT-08: USA-Europe Data Analysis Training Congress, Cambridge, UK-Bilbao, Basque Country-Miami, USA, 2022 [Internet]. sciforum,MDPI; 2022. Available from: https://sciforum.net/paper/view/12682

13. Student Performance Factors [Internet]. Available from: https://www.kaggle.com/datasets/lainguyn123/student-performance-factors/data