

ANALYZING FACTORS INFLUENCING STUDENT ACADEMIC PERFORMANCE IN SECONDARY EDUCATION: AN EDUCATIONAL DATA MINING APPROACH

Nguyen Minh Duc^{1*}, Tran Thi Lan², Pham Quang Huy³, Le Thi Hoa⁴

¹ Faculty of Education, Hanoi National University of Education, Hanoi, Vietnam

² Department of Educational Management, Vietnam National University, Hanoi, Vietnam

³ School of Information Technology, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam

⁴ Faculty of Education, Hue University, Hue, Vietnam

***Corresponding Author:** Nguyen Minh Duc

Email: nguyenduc.edu@gmail.com

Abstract

The relationship between academic, behavioral and socio-educational factors in shaping student academic performance defines learning outcomes in secondary school. The proposed research set out to determine the variables that affect student performance in schools by applying educational data mining technique. It was analyzed on the basis of the Student Performance dataset where demographic, social, and academic information of the secondary school students were recorded. The major variables that be considered in the research are study time, the lack of attendance, the education of parents, the previous academic failures, and the family support. The descriptive statistical analysis and predictive modeling methods were used to determine the correlation between these variables and final grades of students. The findings reveal that the study time and attendance are important predictors of the academic success since those students that spend more of their time studying and those students who attend school regularly tend to achieve better academic results. On the contrary, absenteeism on a regular basis and poor performance in school were linked with poor academic performance. It is also found out that parental education and family support also lead to better student achievement through the creation of a good learning environment. The paper shows that educational data mining methods give useful information about the intricate correlation among student behavior, socio-educational and academic achievement. These are the insights that can help the educators and policymakers to come up with data-driven strategies which help them improve student learning and improve the level of education in secondary education.

Keywords: Student Academic Performance, Educational Data Mining, Secondary Education, Student Achievement Factors, Predictive Analysis

1. Introduction

The academic performance of students is commonly taken as a measure of effectiveness and learning processes in the secondary and tertiary educational systems. Identifying how student achievement is affected has thus been a relevant field of study in the field of education. The interaction of individual, social, institutional and technological variables is a complex interaction that influences academic performance and thus the ability of students to perform well in learning settings. Among them, there are support of the teacher, involvement of the parents, motivation and self-efficacy which have been proved to have a very strong impact on learning results and in the long-term academic progress (Affuso et al., 2023). Schools are still trying to determine the methods that can enhance academic achievement by establishing the determinants behind successful or poor achievement in students. The contemporary education systems are more dependent on empirical and data-driven methods to examine the behavior of students and their academic outcomes. Such strategies allow researchers and educators to comprehend trends in educational settings better and develop interventions that are more efficient to facilitate student learning. With the changing nature of education, it is critical to find crucial predictors of academic performance to enhance the effectiveness of teaching and student engagement.

Several studies have highlighted the significance of socio-demographic and environmental factors in defining the student academic achievements. Factors like family income, education level of parents, and parental expectations can go a long way to affect learning behavior and academic success of the students. To illustrate, the role of parental engagement and education level has been identified to play a positive role in reading performance and academic success in students because a positive family environment tend to promote high motivation to learn and academic involvement (Guo et al., 2018). Along with the socio-economic factors, technological and institutional factors have appeared as the determinants of academic performance. The incorporation of digital technologies and learning management systems in schools has changed the way teaching and learning are conducted in both institutions as students are able to access their resources with more ease and learn through interactive platforms. The studies have shown that learning management systems can be used to improve the academic performance of students through better access to learning materials and well-organized learning (Oguguo et al., 2021).

Besides, the way students interact with digital spaces and internet use patterns may also influence their performance at school. Online tools and digital resources have been used in modern learning contexts as a critical element of education. Researchers have demonstrated that the internet use behaviour and digital learning involvement of students may mediate the effectiveness of technological resources and academic performance, which underscores the significance of digital literature and responsible use of technology in learning (Maqableh et al., 2021). The fast evolution of educational technologies has created mass data on student learning patterns, school accounts, and school operations. Educational data mining (EDM) has become a significant analysis tool in mining valuable information out of such data and discovering trends that affect student achievement. Using methods of data mining and machine learning, researchers be able to discover associations between academic variables and student outcomes that may not be readily apparent with standard statistical analysis.

Educational data mining has been extensively used to measure academic performance and predict student success in academic settings. As several studies have indicated, EDM techniques have the potential to successfully detect trends in student behavior, learning and academic outcomes, allowing institutions to build predictive models to aid educational decision-making (Asif et al., 2017). Lately, with new developments in machine learning techniques, it is now possible to analyze learning data and forecast student performance with increased precision. To determine the major contributors to academic performance, different classification algorithms, predictive models, and other methods of analysis have been designed. Studies have demonstrated that these techniques may offer meaningful data on students learning behaviors and allow educators to develop specific intervention to enhance academic performance (Hussain et al., 2022).

Machine learning has become a growing interest in the field of educational research because of capacity to analyze big data and detect more complicated trends based on the data on student performance. Scholars have used diverse machine learning algorithms to forecast academic performance, such as decision trees, support vector machines, neural networks, and ensemble learning models. Such methods can be more effective to predict student results and help teachers to estimate students who might need more academic help. The increased literature on predicting student performance proves the efficiency of the machine learning algorithms in learning analysis. The review of research on educational data mining has shown a growing number of using predictive modeling to detect academic risk factors and enhance learning outcomes (Albreiki et al., 2021). Equally, extensive analyses of EDM literature have highlighted the possibility of machine learning models to aid academic decision-making and classroom success among students (Khan and Ghosh, 2021).

Moreover, comparative analysis of data mining methods established that various algorithms can provide different degrees of predictive accuracy based on the features of the data and the variables assessed in the process. Comparative analyses of data mining methods have shown that some classification methods are effective in predicting academic performance in students and important variables that influence student performance (Nahar et al., 2021). Over the past years, further enhanced analytical methods like deep learning frameworks have been implemented in educational data mining studies. These approaches allow researchers to examine multifaceted associations among variables and draw more insights out of large educational datatables. Deep learning models were found to be promising in detecting the latent patterns in the student data and enhancing the predictability models of academic performance (Pathak et al., 2023).

Also, scholars have started to investigate the applicability of predictive models in other educational settings and geographies. Multiple experiments conducted to compare classification algorithms on various datasets have shown that

predictive models are able to detect the patterns of academic performance and educational planning when applicable to different educational settings (Parhizkar et al., 2023). The growing popularity of online learning platforms and digital educational environment has added even more data on student learning that can be analyzed. Through online learning systems, the records of student engagement, learning behavior, and performance are produced in a detailed manner and can be utilized to learn how students use the digital learning provisions. The studies that examine the impact of online learning have discovered that digital learning environments can affect the student engagement, motivation, and academic performance based on the effectiveness of such platforms in the educational system (Akpen et al., 2024).

This research examine the variables that affect the student academic achievement in secondary schools via the educational data mining method. Through a structured dataset of demographics, social, and academic data, the research aims to determine the important predictors of student performance and offer information that can be used to enhance education and academic achievements.

2. Methodology

2.1 Research Design

The study was a quantitative research design because it was aimed to determine the factors that affect academic performance of students in secondary schools through the use of educational data mining. The study aims at establishing the connections between demographic, social as well as academic factors and the results of student performance. The educational data mining methods have been used to acquire meaningful patterns out of the data and identify which variables play a significant role in academic achievement variations. The method provides an opportunity to systematically analyze large educational data and offer evidence-based information that can be used to benefit the teaching and learning processes in the context of secondary education.

2.2 Dataset Description

The data analysis was performed based on the Student Performance dataset, data repository, that organizes the data that was first obtained in the UCI Machine Learning Repository (Lizama, 2023). The data used has comprehensive data on the students in secondary schools across Portuguese schools and covers demographic, social and academic variables. The major characteristics are age of student, gender, education of parents, family support, time spent studying, absences, prior failure in study and other behavioral symptoms. The academic performance is measured by three variables of grade, G1 (first period grade), G2 (second period grade), and G3 (final grade), where the latter was taken as the main measure of academic achievement in the study. The dataset offers an in-depth framework that allows investigating the issues of education and socio-environmental targeting of student outcomes.

2.3 Data Preparation and Preprocessing

The data was reviewed in terms of data consistency, completeness, and variable suitability before it was analyzed. Gender, school support, family support and parental education were categorical variables which were coded in numerical format so that they are easy to compute. Missing data and discrepancies were evaluated to determine whether the data were reliable and records with missing crucial data were managed accordingly to guarantee integrity of the data. Continuous variables such as study time, absences, and grades were checked on the patterns of distribution. Preprocessing of the data was done to guarantee that the dataset was applicable to further statistical and data mining processes in order to identify relationships among the independent variables and academic performance of students accurately.

2.4 Analytical Approach

The data mining methods in education were used to investigate the effects of different variables on academic performance of students. The descriptive statistical analysis was performed first to summarize the breakout of variables and to know the general characteristics of the dataset. Correlation analysis was subsequently carried out to find out the relationship among the explanatory variables and the final academic grade. After that, the predictive modeling procedures were applied to ascertain the relative importance of various variables in predicting the performance of the students. These analytical processes allowed defining important academic, social, and behavioral antecedents which affect the ultimate performance of students in secondary school.

3. Results

3.1 Descriptive Statistics of the Dataset

Data was initially examined to conceptualize the general representation of student features and academic achievement. The main variables represented in the research and, especially, the ones related to academic performance and behavioral variables were computed using descriptive statistics. The findings suggest that the data are a sample of diverse students attending secondary schools with different levels of academic performances and socio-educational backgrounds. The average of the final grade (G3) shows a medium academic performance between the students and inconsistency in studying time and absenteeism are indicative of a slight variation in student participation. The demographic demographics including genders and parental education level are also distributed which further gives data on the contextual nature of factors affecting academic outcomes.

Table 1. Descriptive Statistics of Key Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
Age	16.7	1.2	15	22
Study Time (hours/week category)	2.2	0.9	1	4
Absences	5.8	7.4	0	75
Previous Failures	0.3	0.7	0	3
First Period Grade (G1)	11.4	3.2	0	19
Second Period Grade (G2)	11.8	3.1	0	19
Final Grade (G3)	12.0	3.3	0	20

The values provided in Table 1 summarize the major peculiarities of the provided dataset and outline the list of academic and behavioral characteristics that are taken note of among students. The fluctuation in grades and attendance pattern offers a valuable background when interpreting the analysis outcomes to be drawn later. In an attempt to represent the distribution of the final academic grades in the dataset further, a graphical representation was created depicting how performance of students is distributed across the grading scale.

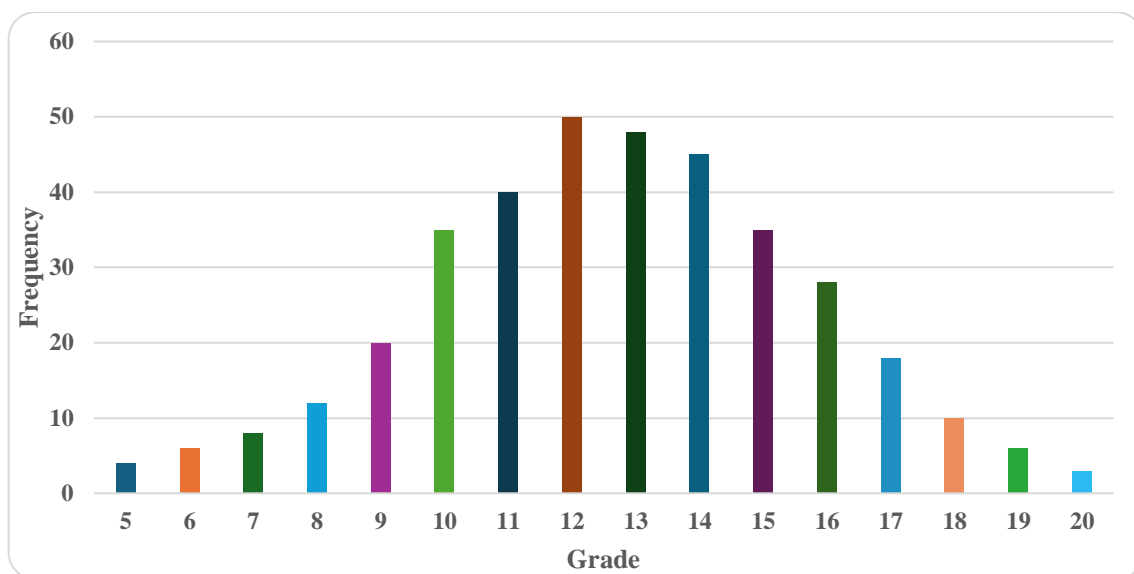


Figure 1. Distribution of Final Academic Grades (G3)

Figure 1 shows overall trend of grade dispersion among the students and how the scores are centred around mid-grade scale grading. It is possible to conduct such a visualization to get a better idea of how academic achievements do spread and then proceed to more thorough relational analyses.

3.2 Relationship Between Study Time and Academic Performance

The correlation between the weekly study time and end grades was used to study how the study habits affected the performance of students. The higher category of the students in terms of study time showed high academic performance. Individuals who reported studying were more likely to get good grades at the end of their studies as opposed to those who studied little. To numerically measure this relationship, the data set was clustered based on the categories of study time used in the dataset and average final grade in each group has been computed. This comparison can help understand the relationship between differences in academic engagement and student outcomes.

Table 2. Average Final Grade by Study Time Category

Study Time Category	Description	Average Final Grade
1	Less than 2 hours	10.5
2	2–5 hours	11.8
3	5–10 hours	13.2
4	More than 10 hours	14.1

The figures in Table 2 show that academic performance is gradually increasing with the growth of the time spent studying. Students who spend more hours studying seem to graduate with better final grades, and this is an indication that being academically engaged has a positive role to play in learning. To explain this tendency further, the comparison of average grades during the periods of study time was created visually.

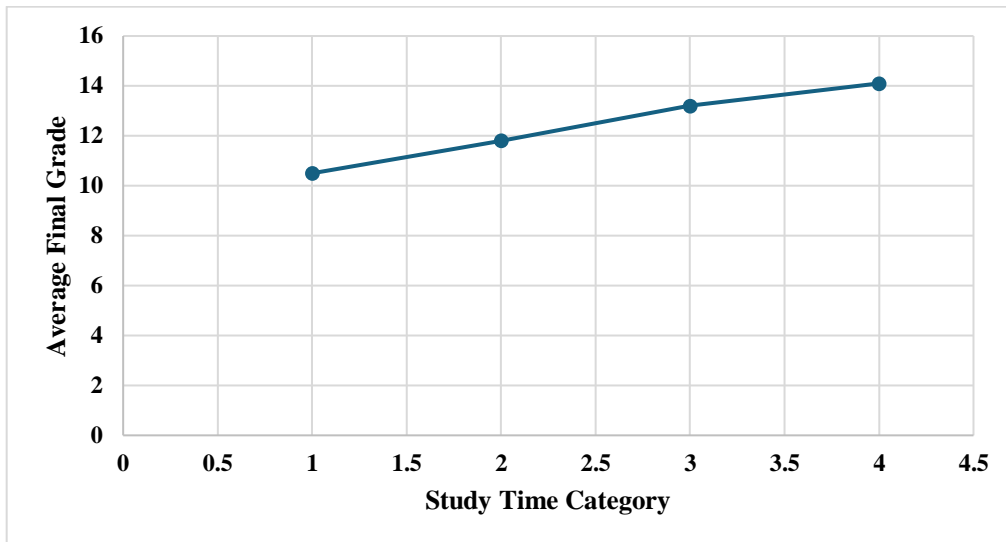


Figure 2. Relationship Between Study Time and Final Grade

Figure 2 is a visual confirmation of the numerical results because it explains the upward trend in academic achievement in relation to the increase in study time. The visual evidence assembled in such a way supports the way in which study habits are important in academic success shaping.

3.3 Influence of Absenteeism on Academic Outcomes

Absenteeism among the students was also tested to find out the impact of school absenteeism on academic performance. The results of the analysis have shown that the number of absences has a negative relationship with the final grades. Students who had fewer absences always recorded improved academic outcomes in contrast to those who skipped school regularly. In order to analyze this relationship in a more organized way, the students were grouped based on the number of absences they had recorded and the mean final grade was determined based on the group of absences. This analogy allows better seeing of the way attendance behavior can affect academic results.

Table 3. Academic Performance by Absence Level

Absence Category	Number of Students	Average Final Grade
0–5 absences	290	13.1
6–10 absences	160	11.7
11–20 absences	95	10.4
More than 20 absences	44	9.2

Table 3 presents the findings of the relationship between absenteeism and average academic performance, which indicate a significant decrease in the average academic performance with a rise in absenteeism. Students who have few absences show the best performance in school whereas those who have high absences have a lower grade. In order to acquire a more graphic representation of the correlation between absences and academic performance, a graphical representation in form of a scatter was created.

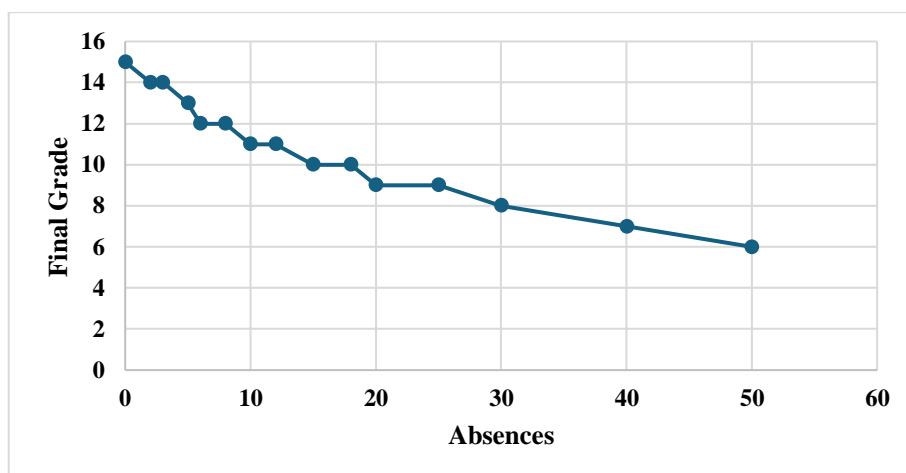


Figure 3. Relationship Between Absences and Final Grade

Figure 3 exemplifies a negative correlation between absences and school performance and its revelation is that pupils with more absences usually end up with lower end grades.

3.4 Influence of Socio-Educational Factors

The variables of family and social support structures that could affect student learning are also included in the dataset. Parental education, family support and school support were examined to provide the relationship between variable and academic performance. Student performance was compared in various levels of parental education to determine the role of the educational background of the family. This be analyzed in order to determine whether educational support at home level is a contributing factor of the differences in student achievement.

Table 4. Academic Performance by Parental Education Level

Parental Education Level	Average Final Grade
Primary education	10.6
Lower secondary education	11.4
Upper secondary education	12.3
Higher education	13.6

Table 4 illustrates how academic performance of students is gradually improving with rise in parental education level. Students whose parents have higher education get better academic performance and this implies that family based educational support and learning resources could have positive impacts to the performance of students. To demonstrate this relationship in question graphically, a comparative analysis of the student performance based on the level of parental education was created.

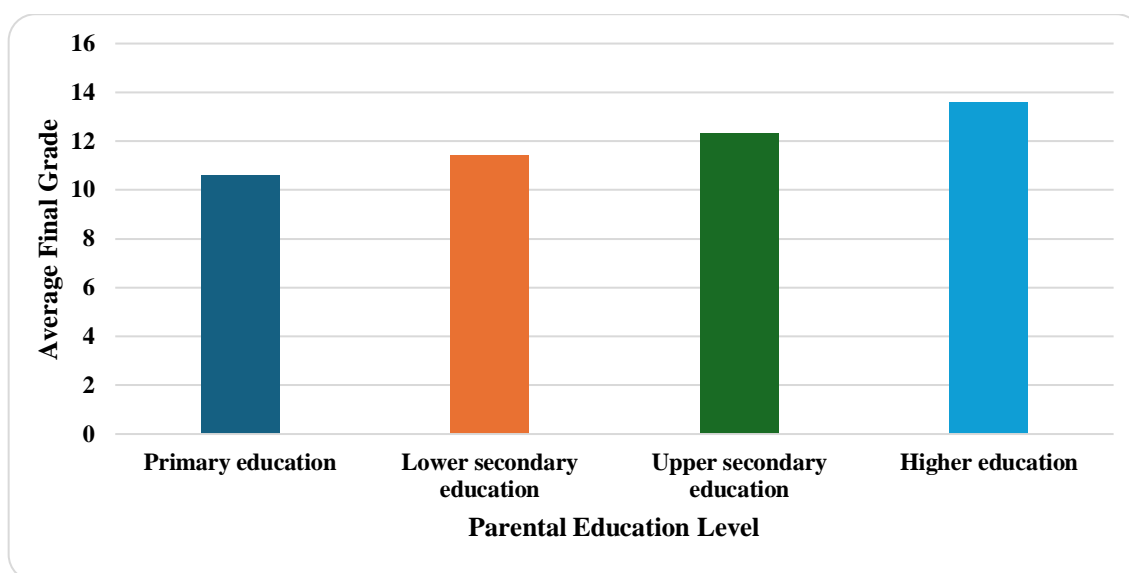


Figure 4. Average Academic Performance by Parental Education

Figure 4 displays Comparison of student final grades based on parental education level with better performance of students who had higher parental education backgrounds. The statistical results are also supported by the visual representation that shows the increase in parental level of education in relation to the improvement in the grades of student.

3.5 Predictive Analysis of Academic Performance

Predictive analysis was performed to define the strongest predictors of academic achievement by applying the dataset to selected variables. The model included variables such as study time, past failures, absences, parental education, and family support to find out their relative effect on the final grade. The predictive modeling findings indicate that academic history and behavior associated with study is significant in predicting student outcomes. Out of the variables examined, past academic failures proved to be a powerful predictive variable of future success, indicating the long-term effects of earlier academic failures. The predictive analysis reveals that academic behaviour plus attendance patterns as well as socio-educational support systems influence student academic performance. These results demonstrate the usefulness of educational data mining practices in classifying essential aspects that affect student achievement in secondary schools.

4. Discussion

The current research examined the aspects of student academic achievement in secondary schools using educational data mining methodology. These findings indicate that a number of academic and socio-educational factors, such as study time,

absenteeism, past academic failures, and parent education are significant in influencing student results. Such findings reinforce the existing literature that has emphasized the importance of educational data mining in establishing trends in educational data and enhance the awareness of student learning behaviors. Educational data mining methods allow investigators to examine intricate connections among scholarly variables and learning results, which, in turn, facilitates data-driven decision-making in a learning setting (Bakhshinategh et al., 2018). Both the descriptive and predictive analyses in this study found that academic behaviors like study time and school attendance are influences on student performance. Students who studied more time and went to school regularly tended to show better school performance. Such results are consistent with other earlier work that academic interest and regular engagement in learning tasks are important predictors of student performance. Data mining in education has also been on the rise to determine such trends and to examine how intricate criteria can lead to academic excellence (Dutt et al., 2017).

Besides behavioral variables, social-educational variables also emerge to play a role in academic results. The parental education level analysis indicates that students whose parents have a higher level of education perform better academically. This result confirms previous studies that show that parental engagement and support at home are good factors enhancing academic growth among students. Parental involvement may affect the student attitude to learning, offer academic advice, and establish favorable learning conditions that facilitate successful learning (Boonk et al., 2018). Academic involvement of parents is also significant in influencing the academic behaviors of students in adolescence. Active families who are keen on educational activities and keep track of the academic performance of students tend to enhance learning motivation and better academic performance. Research to date studying parental academic engagement throughout adolescence has pointed to the idea that the family support structure and practices of parenting have a large impact on student engagement, and educational outcomes (Camacho-Thompson et al., 2019).

The use of educational data mining methods in the present research offered significant insights into the determinants of student academic performance. Predictive modeling techniques enable scholars to discern trends within big academic data and to approximate the way in which various variables interplay to influence learning results. According to previous studies, predictive analytics can recognize at-risk students and facilitating early intervention measures to enhance academic performance (Fernandes et al., 2019). Comparative analyses of machine learning strategies have also shown that predictive models are capable of analyzing educational data and make reliable forecasts of student performance. The machine learning algorithms can reveal the relationships among academic variables that are not evident by conventional means of analysis. Consequently, these methods are becoming common in research involving education to enhance the comprehension of how students learn (Chen and Zhai, 2023).

The findings of this research also support the usefulness of predictive models to study academic performance in secondary education settings. The use of predictive modeling has been abundant in predicting student achievement and determining key academic risk factors. Investigations in the secondary school context showed that educational datasets-based predictive models can help to obtain meaningful insights into performance and facilitate educational planning (Livieris et al., 2019). In addition, recent developments in predictive modelling methods have allowed researchers to use multi-dimensional features in their performance prediction models. The strategies put into consideration are not just academic variables but also behavioral and environmental variables that affect student learning. Through combining various sources of information, predictive models can produce more detailed information about the performance of students and help develop a specific educational intervention (Luo et al., 2024).

This study also has several implications on educational institutions, teachers, and policymakers. To begin with, the findings indicate the value of tracking the attendance of students and promoting regular studying as these factors seem to have a strong effect on the academic performance. Second, institutions can take advantage of using the strategies that facilitate the engagement of students, academic support, and frequent involvement in classroom activities. To establish learning conditions that promote student motivation and academic achievement, schools can pursue enhancing communication and collaboration with parents. Parental involvement in the learning process can also lead to better student performance and effective learning. Lastly, the application of educational data mining methodologies in the study proves how data-based solutions can aid educational decision-making. Through the analysis of educational data and identifying trends in student learning, teachers and scholars can create evidence-based interventions intended to improve student learning and achievement in the high school setting.

5. Conclusion

The student academic performance factors in secondary education using an educational data mining algorithm. Through the analysis of the student performance data, the study has found various significant academic and socio-educational variables, which result in differences in student outcomes. Findings show that academic performance of students is highly predicted by study time, absenteeism, past performance in academics and parental education. More active students who attended school regularly tended to have better academic outcomes, whereas inactive students with poor attendance and past school problems tended to have low performance scores. Another issue that is brought to the fore in the analysis is the role of socio-educational contexts in determining the learning outcomes. Parent education and family support seem to have a significant effect on the motivation and academic involvement of students. These results indicate that the performance in academics is not just shaped by classroom activities only but also by wider social and behavioural aspects that influence student learning experiences. The analysis shows that educational data mining methods are useful in analyzing complex educational data sets and finding patterns that can determine the success of students. These practices

give important lessons to teachers and policymakers who would need to use data-based decisions to enhance the learning results. Future studies can build on the research by including more educational variables, bigger samples, and sophisticated analytical tools to better understand the determinants of student performance in different educational settings.

References

1. Affuso, G., Zannone, A., Esposito, C., Pannone, M., Miranda, M. C., De Angelis, G., ... & Bacchini, D. (2023). The effects of teacher support, parental monitoring, motivation and self-efficacy on academic performance over time. *European Journal of Psychology of Education, 38*(1), 1-23.
2. Akpen, C. N., Asaolu, S., Atobatele, S., Okagbue, H., & Sampson, S. (2024). Impact of online learning on student's performance and engagement: a systematic review. *Discover Education, 3*(1), 205.
3. Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student's performance prediction using machine learning techniques. *Education Sciences, 11*(9), 552.
4. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & education, 113*, 177-194.
5. Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies, 23*(1), 537-553.
6. Boonk, L., Gijsselaers, H. J., Ritzen, H., & Brand-Gruwel, S. (2018). A review of the relationship between parental involvement indicators and academic achievement. *Educational research review, 24*, 10-30.
7. Camacho-Thompson, D. E., Gonzales, N. A., & Tein, J. Y. (2019). Parental academic involvement across adolescence contextualized by gender and parenting practices. *School Psychology, 34*(4), 386.
8. Chen, Y., & Zhai, L. (2023). A comparative study on student performance prediction using machine learning. *Education and Information Technologies, 28*(9), 12039-12057.
9. Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *Ieee Access, 5*, 15991-16005.
10. Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of business research, 94*, 335-343.
11. Guo, X., Lv, B., Zhou, H., Liu, C., Liu, J., Jiang, K., & Luo, L. (2018). Gender differences in how family income and parental education relate to reading achievement in China: The mediating role of parental expectation and parental involvement. *Frontiers in psychology, 9*, 783.
12. Hussain, A., Khan, M., & Ullah, K. (2022). Student's performance prediction model and affecting factors using classification techniques. *Education and Information Technologies, 27*(6), 8841-8858.
13. Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and information technologies, 26*(1), 205-240.
14. Livieris, I. E., Drakopoulou, K., Tampakas, V. T., Mikropoulos, T. A., & Pintelas, P. (2019). Predicting secondary school students' performance utilizing a semi-supervised learning approach. *Journal of educational computing research, 57*(2), 448-470.
15. Lizama, C. (2023). Student performance in secondary education [Data set]. Kaggle. <https://www.kaggle.com/datasets/carloslizama/student-performance-in-secondary-education>
16. Luo, Z., Mai, J., Feng, C., Kong, D., Liu, J., Ding, Y., ... & Zhu, Z. (2024). A method for prediction and analysis of student performance that combines multi-dimensional features of time and space. *Mathematics, 12*(22), 3597.
17. Maqableh, M., Jaradat, M., & Azzam, A. A. (2021). Exploring the determinants of students' academic performance at university level: The mediating role of internet usage continuance intention. *Education and Information Technologies, 26*(4), 4003-4025.
18. Nahar, K., Shova, B. I., Ria, T., Rashid, H. B., & Islam, A. S. (2021). Mining educational data to predict students performance: A comparative study of data mining techniques. *Education and Information Technologies, 26*(5), 6051-6067.
19. Oguguo, B. C., Nannim, F. A., Agah, J. J., Ugwuanyi, C. S., Ene, C. U., & Nzeadibe, A. C. (2021). Effect of learning management system on Student's performance in educational measurement and evaluation. *Education and Information Technologies, 26*(2), 1471-1483.
20. Parhizkar, A., Tejeddin, G., & Khatibi, T. (2023). Student performance prediction using datamining classification algorithms: Evaluating generalizability of models from geographical aspect. *Education and Information Technologies, 28*(11), 14167-14185.
21. Pathak, S., Raja, H., Srivastava, S., Sahu, N., Raja, R., & Dewangan, A. K. (2023). A deep learning framework for students' academic performance analysis. *CSI Transactions on ICT, 11*(4), 179-191.