Check for updates

Identifying Factors Influencing Student Academic Performance Using Feature Selection Method by Weight in Private Higher Institution

Emi Amielda Binti Ahmad Mokhtar[1*] ,Nik Nur Syazreen Binti Nik Rohaimi[2], Suriana Binti Ishak[3], Noor Farah Wahida , Binti Abdul Rashid[4] Hazrina Binti Tajudin[5], Syahirah Binti Abdul Kadir[6]

[1*,2,3,4,5.6] Universiti Poly-Tech Malaysia, Malaysia
amielda@uptm.edu.my*, nursyazreen@uptm.edu.my, suriana@uptm.edu.my, noor_farah@uptm.edu.my, hazrina@uptm.edu.my syahirah_ak@uptm.edu.my

Abstract. Low academic achievement is a significant issue for higher educational institutions, hindering their ability to equip students with the necessary skills to succeed in a competitive and rapidly evolving society. At private higher education institutions in Malaysia, students are required to achieve a minimum Cumulative Grade Point Average (CGPA) of 2.50 to remain eligible for financial aid or scholarships provided by Majlis Amanah Rakyat (MARA). Failure to meet this requirement may result in financial hardship, student dropouts and delayed graduations. This issue not only affects the students personally and financially but also impacts institutional reputation. Therefore, identifying and understanding the factors that influence student achievement is crucial for both educational institutions and policymakers in order to implement targeted interventions. Previous studies have highlighted several determinants, such as socio-economic background, parental involvement, school environment, student motivation, peer influence and the quality of teaching. The objective of this study is to identify key variables using four methods which are Information Gain, Information Gain Ratio, Relief and Chi Square Statistics which assesses the statistical relevance of each variable. The significant variables include Take of Value (TOV), program, semester, gender, total household income and college status. Therefore, identifying factors that influence student achievement is crucial for both educational institutions and policymakers.

Keywords: academic performance; data mining; filter method; private higher institution.

1 Introduction

In the dynamic landscape of global education today, student success plays a more vital role than ever. For institutions striving to highlight academic excellence, drive innovation, and remain relevant in a competitive arena, student performance is a central focus. This is especially the case for private universities and colleges, which typically operate in market-oriented settings where student outcomes significantly impact

N. A. Ishak et al. (eds.), *Proceedings of the International Conference on Cross- Disciplinary Academic Research 2025 - Track 1 Advances in Computing, Electronics, Engineering, and Mathematics (ICAR-T1 2025)*, Advances in Engineering Research 296,
https://doi.org/10.2991/978-94-6239-636-4_16

enrolment, financial support, and institutional prestige. Identifying the core factors behind student success is not just an academic pursuit but a strategic priority for private higher education institutions.

Private higher education institutions differ significantly from public ones in aspects such as funding sources, governance, admissions policies, and resource allocation. These distinctions can greatly shape the overall student experience. Students at private universities often encounter unique challenges, particularly in balancing academics, financial responsibilities, and social life. For instance, the burden of higher tuition fees may compel them to work part-time, potentially affecting their academic focus. However, private institutions also present notable benefits, including smaller class sizes, more favourable student-to-teacher ratios, and personalized academic support, all of which can contribute to student success.

Student achievement is shaped by a wide range of interrelated internal and external factors (Mappadang et al., 2022). Internally, elements like teaching quality, curriculum relevance, faculty involvement, support services, classroom facilities, and access to academic advising and counselling are key contributors (Sanfo & Malgoubri, 2023). Externally, characteristics such as socio-economic background, parental education, past academic performance, personal motivation, and peer influence significantly affect outcomes (Rodríguez-Hernández et al., 2020). In addition, modern challenges such as adapting to digital learning, managing mental health concerns, and integrating technology into education add more layers of complexity to understanding the factors behind student success. Therefore, this study aims to identify the key internal and external factors influencing student academic performance in a private higher education institution, building upon prior empirical findings in the literature.

2 Literature Review

Recent studies have thoroughly examined the various factors that affect student academic performance by utilizing both traditional methods and data-driven modeling approaches. Kocsis and Molnár (2024) explored institutional and demographic factors such as GPA, European Credit Transfer and Accumulation System (ECTS) credits, and gender, underlining their predictive significance for academic success and dropout likelihood. In a related study, Lu et al. (2024) used machine learning techniques to evaluate how psychological and behavioral factors including metacognitive awareness, learning motivation, and class engagement affect academic results, showcasing the increasing utility of artificial intelligence in assessing educational outcomes. Similarly, Al-Abyadh and Abdel Azeem (2022) highlighted the significance of self-management and perceived self-efficacy, suggesting that confidence and organizational abilities are correlated with improved performance. Mappadang et al. (2022) added to this research by demonstrating that intrinsic academic interest plays a crucial role in determining success, emphasizing the essential nature of motivation. Collectively, these studies indicate that student academic performance is influenced by a combination of academic background, demographic characteristics and psychological factors, highlighting the multifaceted nature of student success.

Orji and Vassileva (2022) extended these results by analyzing the relationship between motivational factors and study habits, determining that intrinsic motivation and autonomy are essential for predicting success. Similarly, Yağcı (2022) assessed machine learning models with academic data to demonstrate how predictive analytics can identify underperforming students early in the semester.

Simultaneously, the application of feature selection methods has become increasingly important in educational data mining (EDM) to enhance the accuracy of forecasting academic performance. Mustapha et al. (2023) utilized the Chi-Square method within RapidMiner to pinpoint the most significant predictors, demonstrating that proper preprocessing through feature selection can greatly improve prediction results. Bachri et al. (2017) integrated Chi-Square with neural networks to predict the anticipated study duration for students, affirming its effectiveness in real-world scenarios.

Several studies have validated the effectiveness of Information Gain Ratio (IGR). Enaro and Chakraborty (2020) used IGR to enhance classification performance and Madyatmadja et al. (2021) confirmed IGR's superiority over other techniques like Gini Index and classic Information Gain in terms of precision, recall, and overall accuracy.

Expanding on this idea, Deepika et al. (2022) introduced a hybrid method that integrates Relief-F with Budget Tree Random Forest (RFBTRF), leading to improved identification of academic predictors. Building on this, Roy and Farid (2024) advanced the work by creating an Adaptive Feature Selection Algorithm (AFSA) grounded in Relief-based ensemble ranking, effectively decreasing dimensionality while bolstering model resilience.

Alias et al. (2024) performed an extensive review of techniques for feature selection, covering methods like Information Gain and Chi-Square and the results highlight the significance of factors such as program, semester and gender in forecasting academic success. Similarly, Mubarak and Setiawan (2021) utilized Information Gain feature selection to determine key predictors of student performance. Their study showed that concentrating on important variables improves the precision of classification models, highlighting the value of Information Gain in the context of educational data mining.

In a related review, Albreiki et al. (2021) confirmed that filter-based methods like Chi-Square, IGR, and Relief are particularly effective when used alongside classifiers such as Naive Bayes and Random Forest, especially for categorical data in education. Overall, prior studies demonstrate that filter-based feature selection methods such as Chi-Square, Information Gain Ratio, and Relief consistently enhance model performance; however, their effectiveness varies depending on the dataset and institutional context. Collectively, these research findings illustrate the collaboration between educational theory and computational intelligence in identifying important academic predictors and developing dependable forecasting models. Despite these extensive findings, existing studies largely focus on general higher education contexts and specific academic or psychological variables, with limited attention given to institutional and demographic factors within private higher education institutions.

The objective of this study is to identify and examine key factors in the realm of private higher education institutions. The factors to be analysed include race, religion, gender, total household income, program, semester, OKU status, hostel status, college status and TOV. Gaining a comprehensive understanding of these factors will yield practical insights for stakeholders such as administrators, educators, policymakers, and students, aiming to improve educational strategies, enhance support systems and ultimately elevate academic performance across private institutions.

3 Methodology

This section discusses on Exploratory Data Analysis (EDA), Extract-Transform-Load (ETL) framework, and analysed through four filter-based feature selection methods: Information Gain, Information Gain Ratio, Chi-Square Statistic, and the Relief algorithm. These methods provide a robust approach to identifying the most influential factors in student performance, offering valuable insights for data-driven decision-making in higher education.

3.1 Exploratory Data Analysis (Data Description)

This study utilizes data from the University Poly-Tech Malaysia (UPTM) academic system, which includes 3,796 students and examines ten key variables that comprise both demographic and academic components. The demographic factors considered are gender, ethnicity, religion, total household income, college status, disability status (OKU) and hostel residency. These factors are analyzed for their potential impact on academic performance. The academic variables include the students' program of study, current semester, and TOV scores. The primary indicator of academic achievement used in this study is the Cumulative Grade Point Average (CGPA). The CGPA is categorized into two groups: those above 2.50 and those below 2.50. The dataset variables used in this study are categorised into three types:

1. Binomial - variables with two distinct categories.
2. Polynomial - variables with more than two categories.
3. Real - numerical variables represented with decimal values.

Table 1. Factors and Description

No.	Factors	Description
1.	CGPA (Target factor)	The students' Cumulative Grade Point Average: 1: 2.50 – 4.00 0: 0.00 – 2.49
2.	Gender	Gender of the students 1: Male (M) 2: Female (F)
3.	Race	- Bumiputera Sabah - Bumiputera Sarawak - India Muslim - Melayu - Orang Asli (Semenanjung) - Others
4.	OKU	Persons with disabilities: 1: No (N)

		2: Yes (Y)
5.	Hostel status	Hostel status: 1: Non-residence (NR) 2: Residence
6.	Religion	Religion of students: — Islam — Buddha — Kristian — Others
7.	Program	Program: 1. AA103 2. AA201 3. AB201/AB201YTP 4. AB202 5. AB202YTP 6. AB301 7. AB302 8. AC201 9. BE101 10. BE201 11. BE202 12. BE203 13. BK101 14. BK201 15. CC101 16. CM201 17. CT203 18. CT204 19. CT206 20. CT301
8.	Semester	Students' current semester.
9.	College status	Students' college status: - Full time (FT) - On leave (ONL) - Deferred (DFR)
10.	Total Household Income	Total household income of the students' family.
11.	TOV	Average scores for previous core subject.

Table 2. Type of Data

Factors	Type of Data
CGPA, Gender, OKU, Hostel status	Binomial
Semester, College status, Race, Program, Religion	Polynomial
TOV, Total Household Income	Real

3.2 Extract Transform Load (ETL)

The dataset was prepared using the Extract-Transform-Load (ETL) framework and implemented with RapidMiner. ETL is a widely adopted methodology for gathering, refining, and structuring data before analysis. During the extraction phase, data was sourced from a structured spreadsheet generated by the university's academic management system. In the transformation stage, the data underwent cleaning and standardization: missing values in numerical fields were filled using the mean, while missing categorical values were replaced with the mode. Categorical variables, such as gender, OKU status, and college status, were encoded numerically to facilitate analysis. In the final loading phase, the processed data was input into feature selection tools for further analysis. The selection of tools and procedures was guided by the nature and structure of the dataset to ensure reliability and precision. Figure 1 shows the RapidMiner design on the process of EDA and ETL.

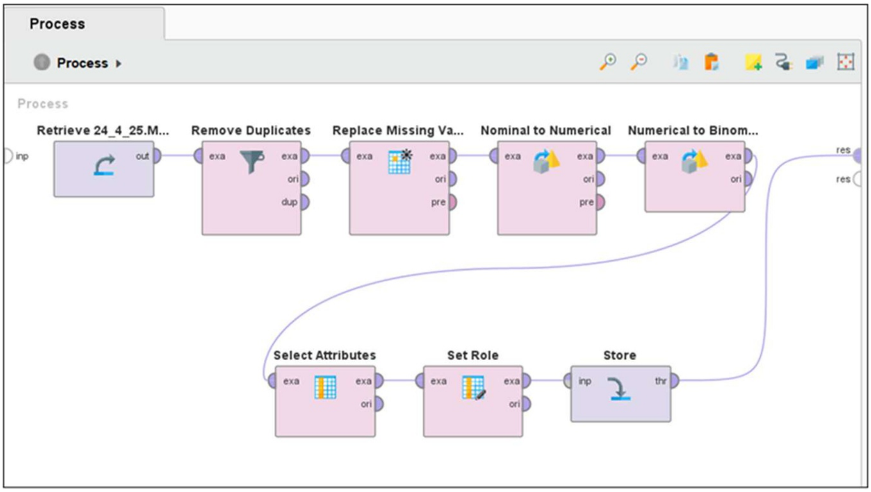

Fig. 1. RapidMiner Design for EDA and ETL

3.3 Feature Selection by Weights (Filter Method)

This study makes use of four filter-based feature selection methods: Information Gain, Information Gain Ratio, Chi-Square Statistics, and Relief. These techniques were selected because they are well-suited for handling large and complex datasets, especially in classification tasks commonly found in educational data mining. Together, they offer a solid and well-rounded approach to identifying the key factors that influence students' academic performance. While Information Gain and Chi-Square focus on measuring how relevant each feature is, Information Gain Ratio helps reduce bias in features that have many categories. The Relief algorithm adds further strength by identifying how features interact with one another, making the overall feature selection process more effective and insightful. Figure 2 show the process for feature selection method by using RapidMiner.

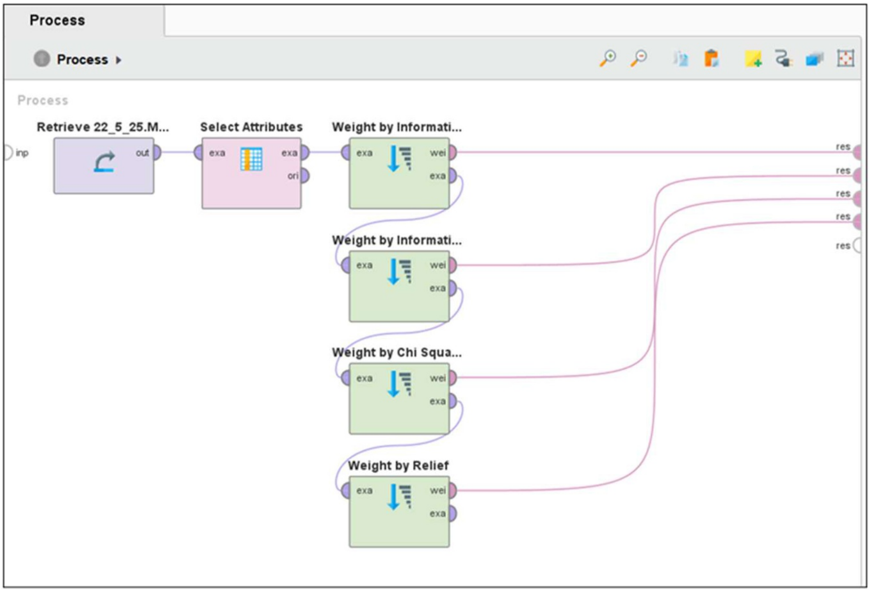

Fig. 2. RapidMiner Design for Feature Selection Method

3.3.1 Information Gain (IG)

Information Gain is a widely adopted feature selection method in data mining, particularly effective for classification tasks. It measures the extent to which a particular feature contributes to reducing uncertainty about the target variable which in this case, student CGPA. The method is grounded in Shannon's Information Theory, where entropy $H(Y)$ quantifies the level of randomness or unpredictability associated with a target variable Y and conditional entropy $H(X)$ represents the amount of uncertainty remaining about Y when the value of feature X is known.

IG is computed as the difference between the initial entropy and the conditional entropy after the dataset is split based on the values of the feature X. It reflects the reduction in entropy or gain in information achieved by incorporating the feature into the model. The formula is expressed as:

$$IG(Y,X) = H(Y) - H(Y \mid X) \quad (1)$$

Here, X denotes the independent variable (feature) and Y is the dependent variable (target). A higher IG value indicates that the feature provides more useful information in predicting the target, making it a strong candidate for model inclusion.

In the field of educational data mining, this approach is especially useful for identifying key characteristics like demographic or academic factors that play a significant role in shaping student performance. Its straightforward calculations and easy-to-understand results make it an essential tool in the process of selecting important features.

3.3.2 Information Gain Ratio (IGR)

The Information Gain Ratio (IGR) is an enhancement of the traditional Information Gain (IG) method which introduced to address the inherent bias of IG towards features with a large number of distinct values. While Information Gain evaluates the reduction in entropy achieved by partitioning the data based on a particular attribute, it tends to favour high-cardinality features which can lead to overfitting and reduced generalisability in predictive models.

To mitigate this issue, the IGR normalises the IG by the intrinsic entropy of the attribute itself, effectively penalising attributes with excessive categorical variation. The mathematical expression for IGR is given as:

$$IGR(Y,X) = \frac{IG(Y,X)}{H(X)} \quad (2)$$

where $H(X)$ represents the entropy of the feature X, and $IG(Y,X)$ denotes the Information Gain between the target variable Y and the feature X.

Normalisation through IGR helps ensure a fairer evaluation of feature importance by reducing the risk of overvaluing attributes that simply split the data into many small groups without truly improving prediction accuracy. This is especially useful in educational data mining, where categorical variables like program codes, course enrolments, or college affiliations often have many unique values. IGR offers a more balanced way to assess these features, leading to models that are not only more accurate but also more generalisable when applied to complex educational datasets.

3.3.3 Chi-Square Statistic (X^2)

The Chi-Square statistic is a well-established non-parametric test commonly used to assess the strength of association between categorical variables. In the context of feature selection, it measures the degree of dependence between an input feature and the target class by comparing observed frequencies with those expected under the assumption of independence. The test statistic is computed using the following formula:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (3)$$

where O_i represents the observed frequency and E_i denotes the expected frequency under the null hypothesis of no association. A higher Chi-Square value indicates a greater level of association between the feature and the class label, suggesting that the feature holds predictive relevance.

This approach works particularly well in educational research, where many of the input variables like gender, college status, or disability status are often categorical. The Chi-Square test is not only easy to use but also quick to compute, making it a handy option for rapidly pinpointing significant features in large datasets related to education. However, it is crucial to keep in mind that the reliability of the Chi-Square test hinges on having a sufficiently large sample size. If some categories have too few observations, the results might not be as reliable.

3.3.4 Relief Algorithm

The Relief algorithm takes a more intuitive, example-based approach to selecting important features. Instead of relying solely on traditional statistics, it looks at individual data points and compares each one to its closest match from the same group (called the nearest hit) and from a different group (the nearest miss). By doing this repeatedly, it figures out which features are most helpful in telling the groups apart, even when the differences are subtle. The relevance score for each feature is updated based on the differences observed between these pairs using the following formula:

$$W[A] = W[A] - diff(A, X, H)^2 + diff(A, X, M)^2 \quad (4)$$

where,
$W[A]$ represents the weight or importance score of feature A
$diff(A, X, H)$ denotes the difference in feature A between the instance X and its nearest hit H
$diff(A, X, M)$ denotes the difference between X and its nearest miss M

This scoring method increases the importance of features that clearly help distinguish between different classes, while lowering the importance of those that contribute little to that distinction. By comparing individual cases rather than looking at the data as a whole, the Relief algorithm can uncover more complex and context-specific relationships between features patterns those broader methods might miss. It also handles noisy data well and works effectively in situations involving multiple classes, making it especially useful in educational data mining, where subtle trends and varied student backgrounds often play a crucial role.

This research highlights that important factor like a student's program of study, the semester, and their gender have a more significant impact on academic performance than more fixed demographic aspects such as race, religion, or socio-economic status. While SPM results can influence a student's early success, it is really their ability to adapt, their motivation, and the support they receive that play crucial roles in their long-term academic journeys. Understanding these key elements is vital for schools and policymakers in their efforts to help all students succeed.

4 Results and Discussion

This study employed four weighted selection methods, namely Information Gain, Information Gain Ratio, Chi Squared Statistics and Relief, to examine the relative importance of variables affecting student academic performance. Table 3 displays the comparison of feature weights obtained from each method.

Table 3. Comparison of Feature Weight

Factors	Information Gain	Chi-Square Statistics	Information Gain Ratio	Relief
Religion	0.0003	0.8114	0.0065	-0.0411
Race	**0.0004**	**0.9536**	**0.0031**	**-0.0340**
Gender	**0.0030**	**16.6903**	**0.0034**	**0.7597**
Total Household Income	**0.0021**	**12.4618**	**0.0139**	**0.1135**
Program	**0.0316**	**191.1426**	**0.0085**	**1.0698**
Semester	**0.0058**	**57.8940**	**0.3409**	**0.6046**
OKU Status	0.0001	0.2696	0.0059	-0.0153
Hostel Status	0.0000	0.0000	0.0000	0.0000
College Status	**0.0012**	**14.1369**	**0.0271**	**0.0801**
TOV	**0.0090**	**0.1347**	**0.0178**	**0.0019**

The academic is the most significant variables indicates that the choice of academic program has the biggest impact on academic performance. This may be attributed to differences in curriculum complexity, available academic support, and student engagement or interest levels. According to the Information Gain Ratio, semester showed the highest weight factor, likely due to the role of academic maturity and progression in shaping student performance. Meanwhile, Take of Value (TOV) ranked second in importance based on Information Gain and was also consistently supported by the other three selection methods.

Next, gender, total household income and college status showed moderate performance in predicting the academic performance. In contrast, religion, race, disability (OKU) status, and hostel residence were found to have minimal impact on academic performance. The hostel status scored zero across all methods, likely due to different approaches in the learning process, which are online and hybrid learning methods.

Students at UPTM come from diverse regions across Malaysia and bring varying levels of academic preparedness, financial background, and personal ambitions. Despite these differences, the data indicate that student performance is influenced not by basic characteristics like race, religion, or physical abilities, but mostly by their ability to adapt to the academic environment.

According to the findings from Gain Ratio, the semester is identified as a significant factor influencing students' performance. A higher semester indicates greater acceptance and comprehension of the learning framework and related courses, which directly impacts the outcomes. As indicated by the TOV, former SPM results lay the groundwork for students' preparedness to pursue higher studies, particularly in essential subjects like Mathematics, Science, and English. Students who achieve strong SPM results generally find it easier to adapt to the academic requirements of diploma and degree courses, especially during the initial semesters. Nevertheless, as students advance, various external factors may come into play.

Utilizing the Relief method reveals differences in learning experiences between male and female students. In certain cases, female students excel compared to their male counterparts due to higher engagement or superior organizational skills. Conversely, male students may face challenges with motivation or struggle to connect with the course content or instructors. Interestingly, socio-economic factors such as household income and college status showed only a moderate influence.

5 Conclusion

This research highlighted that variables like the chosen field of study, semester, and gender have a more significant impact on student performance compared to unchangeable demographic factors such as race, religion, and socioeconomic status. Among all the variables, the academic program exhibits as a key factor determining the academic performance. This may reflect differences in the program difficulty, availability of resources, teaching quality or how well students resonate with their field of study.

In addition, while SPM results may affect initial success, the students' capacity to adapt, their motivation, and the support they receive are essential for long-term academic success. In some cases, students may select programs based on external pressures such as family expectations or scholarship conditions rather than their interests. Advisors can offer clearer guidance and realistic overviews of program requirements during the enrolment phase.

Furthermore, first-year students are still adjusting to new learning methods, especially those coming from rural or rigid educational settings, who often experience cultural and academic challenges. University life demands independence, self-discipline, and time management skills, which take time to develop. In contrast, students in advanced semesters have generally developed coping strategies that enhance their study techniques and boost their confidence in meeting university expectations. This underscores the necessity of implementing mentorship programs, time management workshops, and academic boot camps during the initial years of study.

Although these trends are not universally applicable, they highlight the necessity for support systems sensitive to gender. For instance, mentorship programs could offer diverse learning styles to provide more inclusive academic assistance.

With respect to modelling considerations, The factors of race, religion, disability status (OKU), and hostel status should be excluded when predicting academic performance. The data primarily consists of Malay and Muslim students, which may limit the broader applicability of the findings to different demographic groups. Furthermore, the low representation of students with disabilities makes it difficult to determine the key factors affecting their academic outcomes.

Moreover, hostel status appears to have a low impact on academic performance, especially considering the increase in online and hybrid learning options following the pandemic. This evolution allows students to attend classes from various locations, indicating that physical presence may no longer significantly affect academic achievements.

In summary, the data reflects authentic student experiences, yielding insightful information through statistical methods. By grasping the difficulties students encounter, educators and administrators can create policies and initiatives that are empathetic and effectively promote student success.

Acknowledgments. We would like to thank UPTM for providing the fund for this research under grant number UPTM.DVCRI.RMC.15 (129).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Al-Abyadh, M. H. A., & Abdel Azeem, H. A. H. (2022). Academic Achievement: Influences of University Students' Self-Management and Perceived Self-Efficacy. *Journal of Intelligence*, *10*(3), 55. https://doi.org/10.3390/jintelligence10030055
2. Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student performance prediction using machine learning techniques. *Education Sciences, 11*(9), 552. https://doi.org/10.3390/educsci11090552
3. Bachri, O. S., Kusnadi, K., Hatta, M., & Nurhayati, O. D. (2017). Feature selection based on Chi Square in artificial neural network to predict the accuracy of student study period. *International Journal of Civil Engineering and Technology, 8*(8), 660–670. https://iaeme.com/MasterAdmin/Journal_uploads/IJCIET/VOLUME_8_ISSUE_8/IJCIET_08_08_074.pdf
4. Deepika, K., Reddy, M. S., Pandian, N. R., & Kumar, R. D. (2022). Analysis of performance of accuracy by adding new features individually using Relief-F and Budget Tree Random Forest (RFBTRF) method. *International Journal of Nonlinear Analysis and Applications, 13*(1), 1239–1252. https://doi.org/10.22075/ijnaa.2021.24724.2808
5. Enaro, A. O., & Chakraborty, S. (2020). Feature selection algorithms for predicting students academic performance using data mining techniques. *International Journal of Scientific & Technology Research, 9*(4), 3622–3626. https://www.ijstr.org/final-print/apr2020/Feature-Selection-Algorithms-For-Predicting-Students-Academic-Performance-Using-Data-Mining-Techniques.pdf
6. Hadi Alias, M. A., Hambali, N., Abdul Aziz, M. A., Taib, M. N., & Jailani, R. (2024). Feature selection techniques and classification algorithms for student performance classification: a review. *International Journal of Electrical & Computer Engineering, 14*(3), 3230-3243. https://doi.org/10.11591/ijece.v14i3.pp3230-3243
7. Kocsis, Á., & Molnár, G. (2024). Factors influencing academic performance and dropout rates in higher education. *British Journal of Educational Studies, 72*(3), 414–432. https://doi.org/10.1080/03054985.2024.2316616

8. Lu, J., Liu, Y., Liu, S., Yan, Z., Zhao, X., Zhang, Y., Yang, C., Zhang, H., Su, W., & Zhao, P. (2024). Machine learning analysis of factors affecting college students' academic performance. *Frontiers in Psychology*, 15, 1447825. https://doi.org/10.3389/fpsyg.2024.1447825
9. Madyatmadja, E. D., Sembiring, D. J. M., Perangin Angin, S. M. B., Ferdy, D., & Andry, J. F. (2021). Big data in educational institutions using RapidMiner to predict learning effectiveness. *Journal of Computer Science, 17*(4), 403–413. https://doi.org/10.3844/jcssp.2021.403.413
10. Mappadang, A., Khusaini, K., Sinaga, M., & Elizabeth, E. (2022). Academic interest determines the academic performance of undergraduate accounting students: Multinomial logit evidence. *Cogent Business & Management, 9*(1). https://doi.org/10.1080/23311975.2022.2101326
11. Mubaroq, I. M., & Setiawan, E. B. (2020). The Effect of Information Gain Feature Selection for Hoax Identification in Twitter Using Classification Method Support Vector Machine. *Indonesian Journal on Computing (Indo-JC), 5*(2), 107-118.https://doi.org/10.34818/INDOJC.2020.5.2.499
12. Mustapha, M. F., Zulkifli, A. N. I., Kairan, O., Mat Zizi, N. N. S., Yahya, N. N., & Mohamad, N. M. (2023). The prediction of student's academic performance using RapidMiner. *Indonesian Journal of Electrical Engineering and Computer Science, 32*(1), 363–371. https://ijeecs.iaescore.com/index.php/IJEECS/article/view/26178
13. Orji, F. A., & Vassileva, J. (2022). Machine Learning Approach for Predicting Students' Academic Performance and Study Strategies Based on Their Motivation. arXiv preprint arXiv:2210.08186.
14. Rodríguez-Hernández, C. F., Cascallar, E., & Kyndt, E. (2020). Socio-economic status and academic performance in higher education: A systematic review. *Educational Research Review, 29*, 100305. https://doi.org/10.1016/j.edurev.2019.100305
15. Roy, K., & Dewan Md. Farid. (2024). An Adaptive Feature Selection Algorithm for Student Performance Prediction. *IEEE Access*. https://doi.org/10.1109/access.2024.3406252
16. Sanfo, J.-B. M. B., & Malgoubri, I. (2023). Teaching quality and student learning achievements in Ethiopian primary education: How effective is instructional quality in closing socioeconomic learning achievement inequalities? *International Journal of Educational Development, 99*, 102759.
17. https://doi.org/10.1016/j.ijedudev.2023.102759
18. Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments, 9*(1), 11. *https:*//doi.org/*10.1186/s40561-022-00192-z*