



Predicting Student Project Performance using Machine Learning

Faisal Asad ur Rehman¹, Abdulkarim Kanaan Jebna^{1,*}, Ramesh Kumar Ayyasamy¹, Abdelhak Senadjki²

¹ Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, 31900, Kampar, Perak, Malaysia

² Teh Hong Piow Faculty of Business and Finance, Universiti Tunku Abdul Rahman, 31900, Kampar, Perak, Malaysia

ARTICLE INFO

ABSTRACT

Article history:

Received 2 March 2026

Received in revised form 7 March 2026

Accepted 15 March 2026

Available online 18 March 2026

Keywords:

Digital literacy; entrepreneurial traits;
machine learning; project performance;
students

The transition towards Industry 4.0 and digital transformation has increased the demand for graduates equipped with both behavioural and technical competencies to execute complex tasks successfully. While the current project-based curriculum highlights technical proficiency, emerging evidence suggests that entrepreneurial traits and digital literacy significantly enhance project outcomes. However, current students lack these combined skills, creating a gap between industry expectations and workforce capabilities. This challenge reflects evolving demands and limits graduates' employability in advancing today's industries. Motivated by this gap, this research aims to investigate the influence of entrepreneurial traits and digital literacy on students' project performance. A machine learning based data driven framework was proposed in this study to capture complex and hidden relationships. The data were collected from 691 undergraduate students. The results demonstrate that the XGBoost model achieved an average R^2 of 0.549, with an MSE of 0.444, an RMSE of 0.658, and an MAE of 0.457 under 5-fold cross-validation, showing good predictive capability and low estimation error with a limited dataset. Proactiveness, risk-taking, and digital literacy are the most significant factors in this context. The study recommends that universities redesign the competency-based curricula of their project-based courses to produce skilled graduates ready to work in today's industry.

1. Introduction

The adoption of modern Industry 4.0 technologies, including artificial intelligence (AI), automation, and digital platforms, has advanced industrial processes and workforce skill requirements [1,2]. Today's industry requires a workforce that is capable of integrating technical skills with digital and behavioural competencies to solve complex problems. The graduates who enter these ecosystems often face a substantial skills gap, with a lack of basic competencies to work effectively in real-world projects [3-5].

While the university curricula are based on technical skills, emerging research indicates that a broader set of competencies beyond technical proficiency alone influences successful performance

* Corresponding author.

E-mail address: abdulkarim@utar.edu.my

<https://doi.org/10.37934/araset.57.5.6781>

in such environments [6]. In particular, entrepreneurial traits (ETs), including proactiveness, risk-taking, innovativeness, autonomy, and self-efficacy, are essential skills for the modern workforce to perform well on complex problems and innovative tasks [7,8]. The trait theory of entrepreneurship (TTE) further describes how ETs' competencies influence the leveraging of opportunities, proactive behaviour, and adaptability in uncertain situations that closely mirror the settings of modern industrial-based projects [9,10].

Digital literacy (DL) is also an important competency in the Industry 4.0 workforce [11,12]. It enables individuals to learn and apply new technologies and tools in sustainable industrial systems [13]. Moreover, the digital competency framework (DCF) considers DL as a crucial factor that enables individuals to work effectively in complex technological systems [14,15]. Despite these competencies being frequently recognized in the framework of workforce development, their predictive potential for the students' project performance (PP) remains underexplored. However, the understanding of which competencies among ETs and DL most significantly enhance students' PP is important for both industry and educator stakeholders seeking to enhance learning interventions and align the skills of university graduates with modern industry demands.

Meanwhile, the availability of good amounts of data has stimulated the use of machine learning (ML) approaches to predict and analyse performance in the student context. The traditional methods are centred on linear assumptions, and sometimes fail to identify the exact associations between variables or hidden patterns in educational datasets [16-18]. Unlike traditional methods, ML has the potential to capture complex, nonlinear, and hidden relationships among variables, and to smoothly handle high-dimensional interactions [19-21]. Therefore, it provides more reliable and effective predictions and interpretive insights than traditional approaches [22,23]. The existing research has successfully employed ML techniques to predict academic achievement and student retention, but it relies on conventional factors such as demographic variables and grades [21,24,25]. However, the integration of competency-based factors that aligned with Industry 4.0 talent development has still received limited attention.

This integration of the sustainable industrial requirement, competency-focused students' projects, and advanced analytical capabilities reveals a significant research gap. Existing research has examined ETs, DL, and performance outcomes separately [7,8,11,12], but there is minimal exploration of their combined effects on predicting students' PP. Moreover, the potential of ML to capture the multidimensional relationships in predicting students' PP has also been minimally explored [21,24,25]. However, there is a lack of data-driven and comprehensive frameworks to help design the educational curricula that align with today's industry requirements to produce Industry 4.0-ready talent.

To address this significant gap, this study adopts an ML-driven analytical approach to examine the extent to which ETs and DL jointly affect students' PP. By integrating these competencies into a predictive modelling framework, this study seeks to provide a more realistic understanding of these factors affecting performance in the student context. Accordingly, the research aims of this study are to:

- Examine the influence of ETs in shaping students' PP.
- Demonstrate the applicability of the ML approach in predicting students' PP leveraging student competencies.
- Identify the most significant competencies to enhance students' PP.

Through these examinations, this research seeks to contribute to the literature in several ways. First, it aims to explore the influence of ETs' and DL competencies on students' PP, moving beyond traditional academic factors. Secondly, the study intends to propose a methodological contribution

by applying ML techniques to capture the true relationships between competencies and performance. Lastly, this study seeks to offer practical insights for universities to develop competency-based project curricula aligned with Industry 4.0 workforce demands. However, this research presents new insights through its competency-based predictive framework, which combines ETs and DL within an ML modelling to explore students' PP. This data-driven approach provides a new methodological framework for a deeper understanding of talent development in universities for Industry 4.0.

This study is organized into different sections. The upcoming sections present the methodology, which starts by proposing a workflow for the research framework, explains the data description and the selected ML models for predicting students' PP, and concludes with the evaluation metrics. After this, the results and discussion are presented, which provide a detailed analysis and overview of how the study's findings relate to the existing literature. The last section concludes the study's outcomes, presents the significance of this work, identifies its limitations, and suggests directions for future research.

2. Methodology

This study proposes a research framework that is based on ML to generate robust predictive and empirical insights into the influence of ETs and DL on students' PP (as shown in Figure 1). It also identifies the most significant factors for the same context.

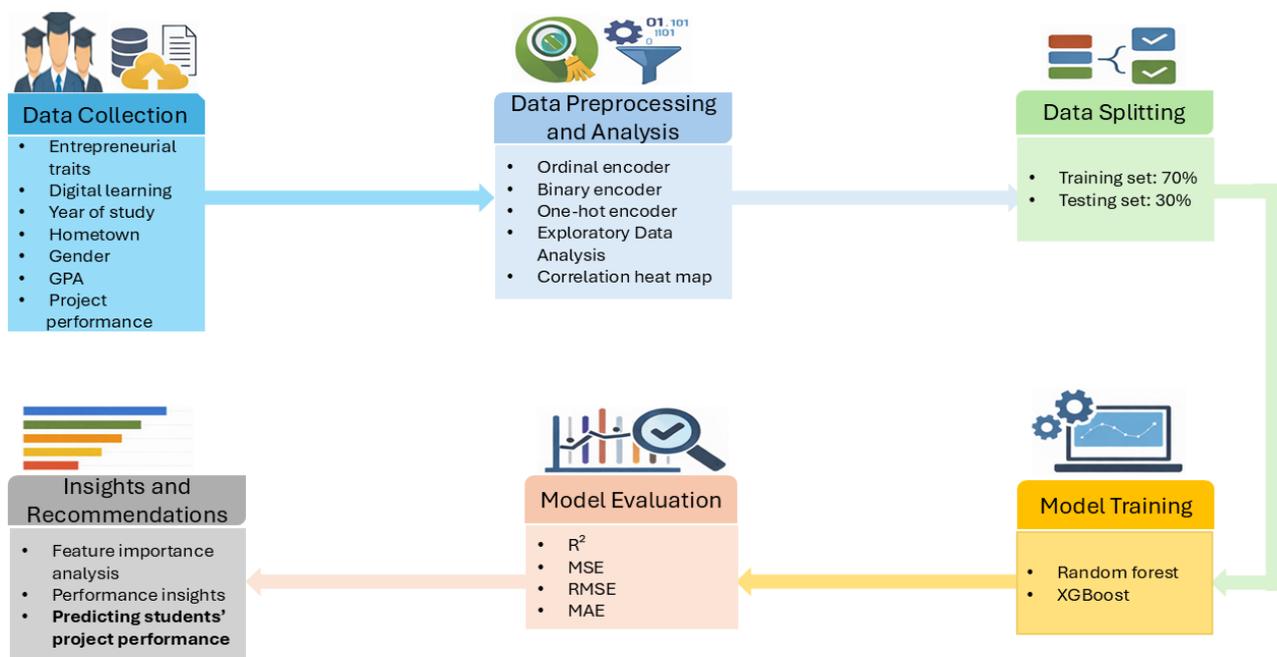


Fig. 1. The workflow of the research design

This approach captures the nonlinear and complex nature of relationships between the competencies and performance that traditional methods can not effectively capture. The process of data collection, data preprocessing, data splitting, model training, and model evaluation ensures the effectiveness and reliability of the proposed framework. The research design of this study supports accurate performance and empirical insights into the students' PP by integrating grounded competencies into data-driven ML models.

2.1 Dataset Description

The data for this study were collected through a questionnaire from undergraduate students at different public Malaysian universities. The target audience was those enrolled in project-based courses focused on ETs and DL skills. The students participated in this data collection process voluntarily, and assurances were given that their anonymity and data confidentiality would be maintained. A total of 330 responses were collected. The description of each study variable is presented in Table 1.

Table 1

The description of the variables

S. No.	Variable Full Name	Variable Abbreviation	Data Type	Scale	Description
1	Project performance	PP	Categorical	Five-point-likert	Students' PP
2	Digital literacy	DL	Categorical	Five-point-likert	Students' ability to adopt DL.
3	Proactiveness	PRA	Categorical	Five-point-likert	Students' ability to work proactively on their project tasks.
4	Innovation	INN	Categorical	Five-point-likert	Students' ability to look into new creativity in their projects.
5	Risk-taking	RT	Categorical	Five-point-likert	Students' willingness to take risks in their project tasks.
6	Autonomy	AUT	Categorical	Five-point-likert	Students' ability to take responsibility and independently manage their project tasks.
7	Appearance self-efficacy	ASE	Categorical	Five-point-likert	Students' subjective evaluation of their appearance values.
8	Social self-efficacy	SSE	Categorical	Five-point-likert	Students' subjective evaluation of their social values.
9	Gender	-	Categorical	Binary	Students' gender
10	Hometown	-	Categorical	Binary	Students' hometowns: are they from rural or urban areas?
11	Year of study	-	Categorical	Nominal	Students' study year
12	GPA	-	Numeric	Continuous	Students' GPA

2.2 Data Preprocessing and Exploratory Analysis

All the variables in this study were categorical except GPA, which was a continuous variable. As shown in Table 1, the subfactors of ETs, including proactiveness, innovation, risk-taking, autonomy, social self-efficacy, and appearance self-efficacy, and other variables such as DL and PP, were measured using a five-point Likert scale (strongly agree, agree, neutral, disagree, strongly disagree). As Likert-scale responses are ordinal, an ordinal encoder was used to convert the categories into the corresponding numerical values while maintaining their inherent order [26]. While gender and hometown were the binary variables, a binary encoder was applied to convert the categories into numerical values [27]. On the other side, the year of study was a nominal variable; a one-hot encoder was used to convert these categories into numerical forms [27]. As ML models operate exclusively on numeric data, this preprocessing step is essential. Therefore, all categorical variables were converted to numeric forms.

Moreover, to understand the dataset characteristics, an exploratory data analysis (EDA) was used in this study. It gives an overall idea of the dataset, data preprocessing, and selection of the ML models based on the data characteristics [28,29]. It presents the descriptive statistics to summarise each variable's characteristics and help to select the variables for final modelling [30,31]. It is the essential part of the ML workflow. This study also used a correlation analysis to assess associations among variables. It identifies linear interrelationships among the features and gives a clear picture of the selection of relevant features for the model [32,33].

2.3 Machine Learning Models

To capture the complex and nonlinear relationships among ETs, DL, educational factors, and PP, this study uses two ensemble ML models: RF and XGBoost. Both models have shown strong predictive power in academic research and learning analytics applications [34-36]. RF is an ensemble ML model based on the bagging concept, generating multiple decision trees from random subsets of training examples and features. In a final prediction of RF, the prediction of each tree is aggregated. It reduces the overfitting issue and enhances the model's reliability and stability. It also handles heterogeneous and noisy data, which is very common in an educational context. To provide the feature importance score, increase its power to enhance the interpretability. It identifies the set of best predictors for the outcome variable [37-40].

On the other hand, XGBoost is also one of the reliable models in ML. The architecture of XGBoost is based on sequential trees, in which each sequential tree reduces the error of the previous tree, resulting in robust predictions. It overcomes the chance of overfitting by incorporating a regularisation term and also outperforms in situations where features have different data types and have missing values [41,42].

The dataset for this study was split into 80:20 for training and testing. To assess the reliability and unbiasedness of both the RF and XGBoost models, they were trained on the training set, and their performance was evaluated on the test set. In addition to that, k-fold cross-validation (CV) was utilized to mitigate the overfitting issue and optimize the hyperparameters [43].

It splits the training data into k folds. The k-1 folds were used for model training, while the remaining fold was used for validation purposes of the model. This study used $K = 5$, which means that performance measures were measured across all 5-folds and averaged to get the stable performance of the model before external evaluation on the unseen test set. It enhances the model. It enhances model confidence and also addresses the heterogeneity and variability concerns of the data [44]. To determine the most important features, the study used Shapley additive explanations

(SHAP). SHAP quantifies each feature's marginal contribution by calculating Shapley values. It represents the average contribution of a feature across all feature coalitions [45].

2.4 Model Evaluation

A set of evaluation measures was used to assess the ML model's performance, including coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). They interpret the model's predictive quality and assess its fitness [46-48]. These are common evaluation matrices for regression tasks in educational research [34, 49, 50]. The purpose of the evaluation matrices is shown in Table 2.

Table 2
 Purpose of the evaluation measures

Measure	Purpose	Interpretation
R^2	It is used to measure the explained variance by independent variables (factors).	Better explanatory power explains if its value is higher.
MSE	It highlights a larger prediction error to monitor the overall model accuracy.	Lower value indicates less overall error.
RMSE	It offers easy-to-interpret model accuracy by computing the prediction error in the same units as the outcome variable.	For better model performance, the lower value should be considered.
MAE	Calculating the average magnitude of the prediction error yields a good, interpretable measure of model performance.	For better model performance, the lower value should be considered.

3. Results and Discussion

The section presents the findings of this study. First, it presents the overall insights from the dataset using different statistical analyses. Secondly, it compares the performance of the ML models and then identifies the success factors in predicting students' PP. Table 3 presents descriptive statistics of the numerical variables of the study. It shows the high mean values of risk-taking, proactiveness, innovation, and appearance self-efficacy, while social self-efficacy and autonomy demonstrated lower mean and higher standard deviation (SD) values among the ETs. It indicates that there are variations in perceived independence and social confidence in students.

Table 3
 The summary of the numeric variables of the study

Variable	Count	Mean	Median	SD	Min	25%	50%	75%	Max
GPA	691	3.31	3.34	0.44	1.58	3.00	3.34	3.65	4.0
INN	691	3.74	4.00	0.71	1.00	3.33	4.00	4.00	5.0
RT	691	3.93	4.00	0.73	1.00	3.50	4.00	4.50	5.0
PRA	691	3.85	4.00	0.65	1.00	3.33	4.00	4.00	5.0
AUT	691	3.51	3.50	0.77	1.00	3.00	3.50	4.00	5.0
ASE	691	3.67	3.67	0.76	1.00	3.00	3.67	4.00	5.0
SSE	691	3.36	3.33	0.80	1.00	2.83	3.33	4.00	5.0
DL	691	3.81	4.00	0.59	2.00	3.25	4.00	4.00	5.0
PP	691	3.84	3.86	0.63	1.00	3.43	3.86	4.14	5.0

The academic factor (i.e., GPA) shows good performance of the students in their academics. The DL and PP factors also show good mean values along with low SDs. The descriptive statistics indicate sufficient variation across all variables, supporting the ML applicability to work well in heterogeneous datasets.

The distribution of the categorical variables of the study is shown in Figure 2. It indicates that there are more female students than male students in the dataset. The year-wise distribution indicates that early-year students were engaged in the project-based courses. Regarding hometown background, the distribution shows that more students are from urban areas than from rural areas. Overall, these findings reveal that diverse student representation in the dataset supports generalizability across various academic and demographic backgrounds.

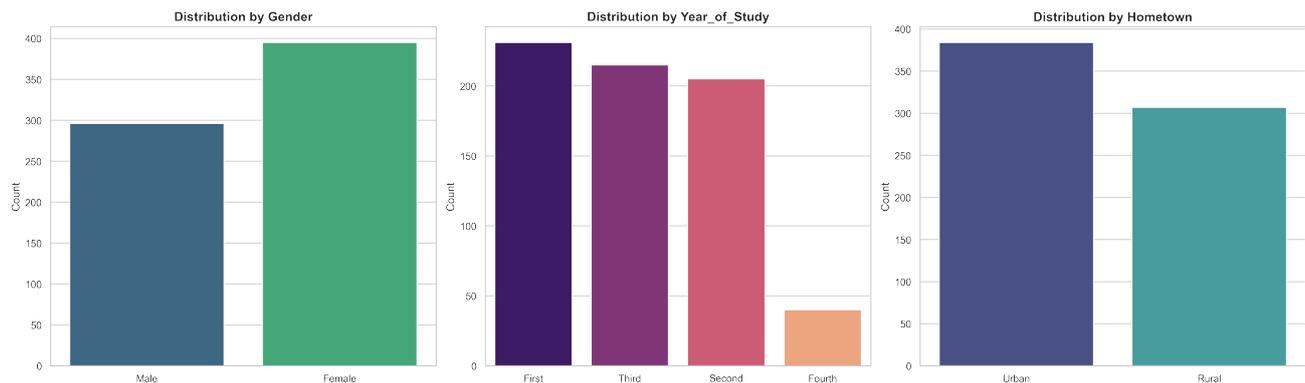


Fig. 2. Demographic distribution of respondents by gender, year of study, and hometown

The strength and direction of associations among quantitative study variables are illustrated in Figure 3. PP shows moderate positive correlations with proactiveness, risk-taking, innovation, DL, and appearance self-efficacy, suggesting that undergraduate students with higher entrepreneurial ability and digital competence achieve better project outcomes. Social self-efficacy and autonomy demonstrate a weaker, however positive, correlation with PP, indicating supportive but less important factors. GPA shows only a weak association with PP, underscoring that the traditional academic construct is a weak predictor of real-world, complex projects. Notably, strong intercorrelations between innovation, risk-taking, and proactiveness indicate conceptual relatedness without showing severe multicollinearity. Therefore, supporting their joint inclusion in subsequent ML models to capture interactive effects and nonlinearity on students' PP.

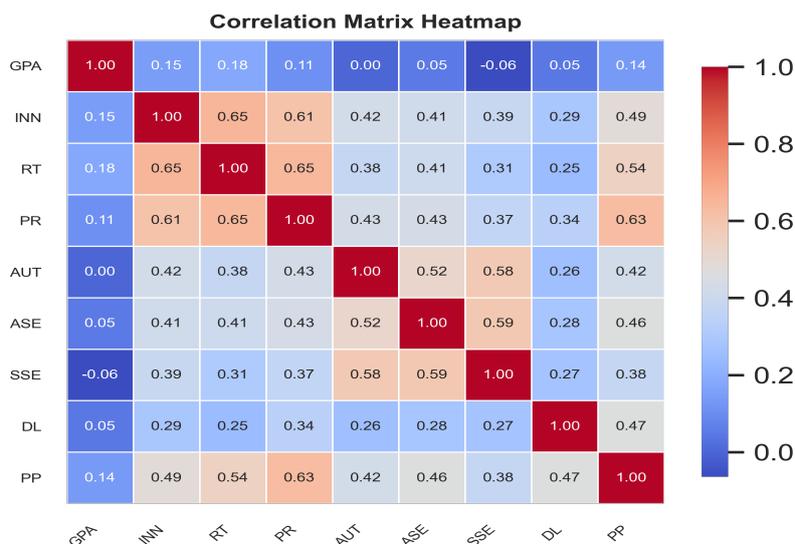


Fig. 3. Correlation heatmap emphasizing the association among quantitative study variables

Table 4 compares the prediction performance of the RF and XGBoost models across the training, CV, and test phases. Both models demonstrate good predictive power for PP among university students. The CV R^2 values are 0.524 and 0.549 for RF and XGBoost, respectively, showing that predictors explain a significant proportion of PP variance.

Table 4
 Performance analysis of the ML models

Measure	RF Performance			XGBoost Performance		
	Training	CV	Testing	Training	CV	Testing
R^2	0.780	0.524	0.589	0.731	0.549	0.598
MSE	0.216	0.474	0.434	0.265	0.444	0.424
RMSE	0.465	0.676	0.658	0.514	0.658	0.651
MAE	0.326	0.459	0.450	0.368	0.457	0.448

XGBoost exhibits slightly better generalization, as evidenced by higher CV and test R^2 and lower test errors (MSE, RMSE, and MAE) than RF. At the same time, the RF model achieves a marginally higher training R^2 value. Still, the gap between RF's training and test performance suggests a greater tendency toward overfitting than XGBoost. It suggests that the XGBoost model delivers better and more reliable results while maintaining accuracy and generalizability in predicting students' PP.

The feature importance scores are shown in Figure 4. It shows that proactiveness is the most important individual factor in predicting students' PP, with almost 25% of the total importance. However, ETs collectively account for 63% of the significance. In which proactiveness, risk-taking, appearance self-efficacy, and innovation are the higher contributors, while autonomy and social self-efficacy are the lower. Meanwhile, the DL contributes almost 15% of the total importance. So, it indicates that students who are proactive, take risks to add new things, have the capability to use digital tools, generate unique ideas, and have significant confidence in their abilities, achieve good performance in their projects.

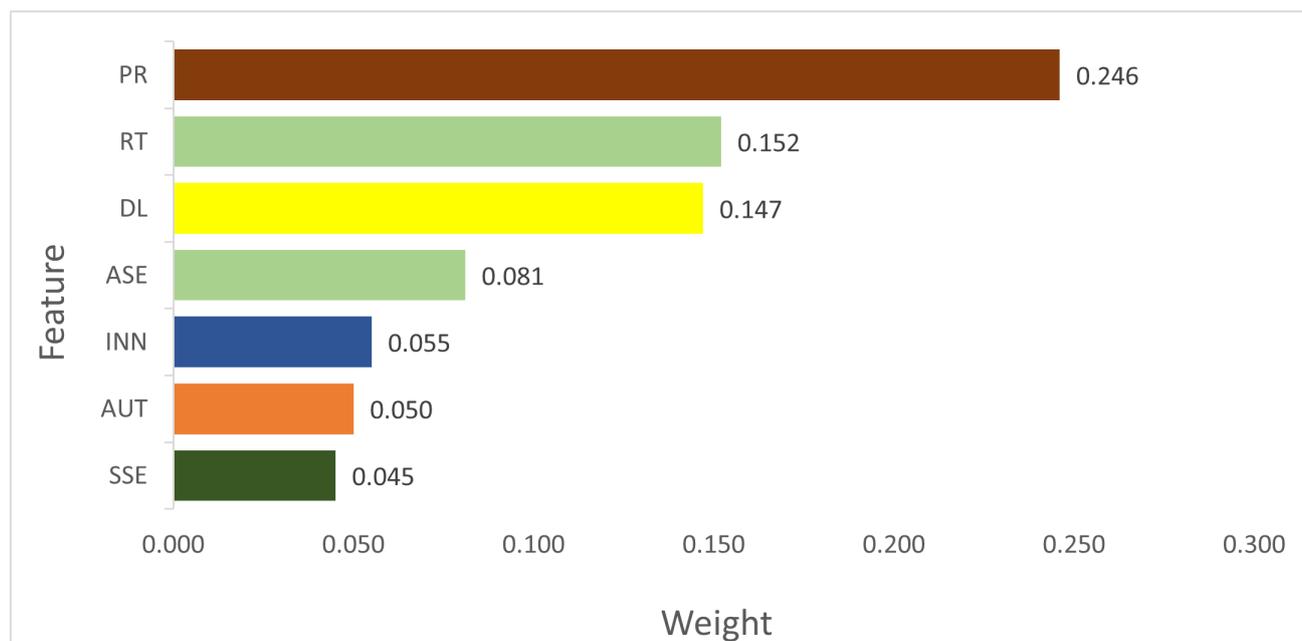


Fig. 4. Feature importance for this study using RF

However, this research examines the impact of ETs and DL on students' PP, develops a suitable ML model for prediction, and identifies the most significant factors, thereby proposing strategies for the universities to produce skilled graduates who meet the requirements of today's modern industry. The findings provide empirical evidence that supports the proposed framework of this study and extend existing research by integrating ETs and DL competencies within a data-driven analytical approach.

Consistent with prior studies, the findings show that XGBoost and RF models achieved satisfactory predictive performance across the training, CV, and test datasets [51, 52, 53, 54]. The R^2 values reveal moderate explanatory power. At the same time, error metrics indicate low values that confirm the robustness of the models in predicting students' PP. These findings align with earlier research that found ensemble models superior at capturing the nonlinear and complex relationships among the educational variables [25, 55, 56, 57]. Thus, the findings support one of the objectives of this study, which confirms the feasibility of using ML techniques to model competency-based students' PP.

Regarding the ETs, the feature importance analysis indicates that proactiveness, risk-taking, and innovation contribute to students' PP. These findings are consistent with prior studies indicating that proactive behaviour improves initiative and task engagement in the projects. At the same time, the ability to take risks facilitates problem-solving and experimentation, and the innovative behaviour is associated with the development of creative solutions [58-60].

Meanwhile, autonomy and social self-efficacy showed relatively low importance to students' PP. This divergence supports the prior studies that highlight the role of autonomy and self-efficacy in learning performance [61-63]. A promising explanation is that the projects involve technological problems and digital work processes, in which proactive engagement and technical competencies may overcome an individual's perceived autonomy and confidence. These findings emphasize the contextual nature of the effectiveness of ETs in the modern project context.

Moreover, the DL emerged as a significant factor in students' PP, which has a relatively good feature weight. It supports the existing studies that have shown that students with high DL exhibit superior task management, improved problem-solving efficiency, more effective digital collaboration, and better learning performance [64-66]. These findings extend the research domain

by empirically showing the dominant role of DL within an ML-based performance model that highlights its centrality to achieving successful project outcomes.

The combined inclusion of ETs and DL in predictive ML models improved performance accuracy. These findings are supported by previous studies showing that Industry 4.0 demands these skills in the workforce for sustainable manufacturing [67-70]. The findings align with theoretical perspectives grounded in TTE and DCF, as well as the integrated competency framework. It indicates that behavioural and digital skills improve performance. This joint effect has been mainly overlooked in previous empirical studies, which have often examined these competencies separately.

However, the findings address all objectives of the study by showing the significant roles of ETs and DL competencies, validating the ML approach for competency-based predictive modelling, and finding the most significant factors. The dominance of ETs, including proactiveness and risk-taking with DL, suggests that project-based curricula should be grounded in tasks that require entrepreneurial mindsets and digital capabilities to support problem-solving and adaptive learning. Moreover, the limited influence of autonomy and self-efficacy suggests that not all ETs contribute equally to the students' PP. This insight helps refine competency development strategies that enable educators to prioritize competencies with performance relevance.

These findings suggest that universities should incorporate entrepreneurial and digital competencies into project-based curricula rather than focusing solely on technical skills. By adding proactive approaches, risk-taking tasks, creative solutions to problems, and the use of digital tools in the projects, students are prepared for tasks similar to those required by today's industry.

4. Conclusions

This study proposed a data-driven approach to predict the students' PP. It investigates the importance of entrepreneurial and digital competencies through an ML approach. The findings of this proposed work suggest that entrepreneurial and digital capabilities should align with educational pathways that equip students with essential skills to work in a modern industry.

This study reveals that ML ensemble models can predict the students' PP. It provides a meaningful interpretation of how ETs and DL interact to measure the students' PP. It also emphasizes how to assess students' readiness for today's industries. The findings suggest that academic curricula need to add entrepreneurial and digital capabilities to project-based courses to prepare students for the modern industry.

This study has some limitations. The dataset has only 691 examples that were collected from university students. It may affect the generalizability of the model. This study is based solely on the Malaysian context; in addition, the self-reporting of the measures may introduce subjective bias. These limitations highlight the need for further research in the same context. Future research may improve this work by adding more data to enhance the reliability of the ML models. To assess the external validity of the model and explore the contextual differences, the data were collected from different countries. The longitudinal research may also be applied for a deeper understanding of how entrepreneurial and digital competencies develop over time in university students, which may provide meaningful insights into how these capabilities affect performance at different levels.

For educators and policymakers, this work provides practical and strategic insights for bridging the gap between universities and the growing needs of Industry 4.0. This study concludes with offering an ML-based solution for redesigning a new curriculum for project-based courses that focus on entrepreneurial and digital competencies in individuals to produce graduates who fulfil the requirements for Industry 4.0.

Acknowledgement

The authors sincerely thank Universiti Tunku Abdul Rahman (UTAR) for providing financial support for this work under UTAR Research Fund, Project No. IPSR/RMC/UTARRE/2023-C1/A07. It plays a primary role in successfully conducting this research.

Conflict of Interest Statement

The authors declare that they have no conflicts of interest in this work.

Author Contributions Statement

Faisal Asad ur Rehman: Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing a original article & editing, and Visualization. Abdulkarim Kanaan Jebna: Supervise, Conceptualization, Validation, Resources & editing. Ramesh Kumar Ayyasamy: Supervise. Abdelhak Senadjkia: Supervise.

Data Availability Statement

The data of this study are available from the corresponding author upon reasonable request.

Ethics Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

References

- [1] Peres, Ricardo Silva, Xiaodong Jia, Jay Lee, Keyi Sun, Armando Walter Colombo, and Jose Barata. "Industrial artificial intelligence in industry 4.0-systematic review, challenges and outlook." *IEEE access* 8 (2020): 220121-220139. <https://doi.org/10.1109/ACCESS.2020.3042874>
- [2] Rocha-Jácome, Cristian, Ramón González Carvajal, Fernando Muñoz Chavero, Esteban Guevara-Cabezas, and Eduardo Hidalgo Fort. "Industry 4.0: a proposal of paradigm organization schemes from a systematic literature review." *Sensors* 22, no. 1 (2021): 66. <https://doi.org/10.3390/s22010066>
- [3] Bhattarai, Prakash C., Mahesh Nath Parajuli, Suresh Gautam, Prakash Kumar Paudel, Anup Bhurtel, and Amrita Sharma. "Education–work transition: skill gaps in the construction industry." *Frontiers in Built Environment* 11 (2025): 1623609. <https://doi.org/10.3389/fbuil.2025.1623609>
- [4] Hegde, S., and P. Kavade. "Bridging skill gap: A study on industry expectations vs. graduate competencies." *International Journal for Multidisciplinary Research* 7, no. 3 (2025): 1-6. <https://doi.org/10.36948/ijfmr.2025.v07i03.44225>
- [5] Mohammed, Farhad Sadik, and Fezile Ozdamli. "A systematic literature review of soft skills in information technology education." *Behavioral Sciences* 14, no. 10 (2024): 894. <https://doi.org/10.3390/bs14100894>
- [6] Akhilla, Kharismatul Zaenab, Anggraini Sukmawati, and Bagus Sartono. "Identifying the types of future skills needed in the manufacturing industry: a systematic literature review." *Journal of Application Business & Management/Jurnal Aplikasi Bisnis dan Manajemen* 11, no. 3 (2025). <https://doi.org/10.17358/jabm.11.3.1099>
- [7] Park, Jeong-Hyun, and Seon-Joo Kim. "Entrepreneurial Competencies in the Era of Digital Transformation: A Systematic Literature Review." *Digital* 5, no. 4 (2025): 46. <https://doi.org/10.3390/digital5040046>
- [8] Adeniyi, Adeshina Olushola, and Idris Olayiwola Ganiyu. "Reshaping education and entrepreneurial skills for Industry 4.0." In *Reshaping entrepreneurship education with strategy and innovation*, pp. 64-77. IGI Global Scientific Publishing, 2021. <https://doi.org/10.4018/978-1-7998-3171-6.ch004>
- [9] Thai, Quoc Hoang, and Khuong Ngoc Mai. "Entrepreneurial traits: a systematic literature review." *Management Review Quarterly* (2023): 1-69. <https://doi.org/10.1007/s11301-023-00370-4>
- [10] Xu, Miaoxin. "The effects of personality traits on entrepreneurial intention and creativity among Chinese and African college students in era of industry 4.0/5.0." *Frontiers in psychiatry* 14 (2023): 1098148. <https://doi.org/10.3389/fpsy.2023.1098148>
- [11] Yaacob, Teh Zaharah, Keerthi Poobalan, H. I. Che Hashim, M. Z. Hasan, Yogeewari Subramaniam, and Logaiswari Indiran. "The relationship between students' digital competency skills and adaptation to Industry 4.0 learning

- technologies." *International Journal of Academic Research in Business and Social Sciences* 14, no. 7 (2024): 622-632. <http://dx.doi.org/10.6007/IJARBS/v14-i7/21938>
- [12] Keshar, A. "Advancing Industrial IoT and Industry 4.0 through Digital Twin Technologies: A comprehensive framework for intelligent manufacturing, real-time analytics and predictive maintenance." *World Journal of Advanced Engineering Technology and Sciences* 14, no. 1 (2025): 228-240. <https://doi.org/10.30574/wjaets.2025.14.1.0019>
- [13] Li, Ling. "Reskilling and upskilling the future-ready workforce for industry 4.0 and beyond." *Information systems frontiers* 26, no. 5 (2024): 1697-1712. <https://doi.org/10.1007/s10796-022-10308-y>
- [14] Martínez-Bravo, María Cristina, Charo Sádaba Chalezquer, and Javier Serrano-Puche. "Dimensions of digital literacy in the 21st century competency frameworks." *Sustainability* 14, no. 3 (2022): 1867. <https://doi.org/10.3390/su14031867>
- [15] Cetindamar Kozanoglu, Dilek, and Babak Abedin. "Understanding the role of employees in digital transformation: conceptualization of digital literacy of employees as a multidimensional organizational affordance." *Journal of Enterprise Information Management* 34, no. 6 (2021): 1649-1672. <https://doi.org/10.1108/JEIM-01-2020-0010>
- [16] Hilbert, Sven, Stefan Coors, Elisabeth Kraus, Bernd Bischl, Alfred Lindl, Mario Frei, Johannes Wild, Stefan Krauss, David Goretzko, and Clemens Stachl. "Machine learning for the educational sciences." *Review of Education* 9, no. 3 (2021): e3310. <https://doi.org/10.1002/rev3.3310>
- [17] Almalawi, Ahlam, Ben Soh, Alice Li, and Halima Samra. "Predictive models for educational purposes: A systematic review." *Big Data and Cognitive Computing* 8, no. 12 (2024): 187. <https://doi.org/10.3390/bdcc8120187>
- [18] Gomes, Cristiano Mauro Assis, Gina C. Lemos, and Enio G. Jelihovschi. "The reasons why the regression tree method is more suitable than general linear model to analyze complex educational datasets." *Revista Portuguesa de Educação* 34, no. 2 (2021): 42-63. <https://doi.org/10.21814/rpe.18044>
- [19] Bai, Haiyan, Xing Liu, Fangxing Bai, Yuting Chen, and Randyll Pandohie. "Machine Learning Method for High-Dimensional Education Data." *Journal of Methods and Measurement in the Social Sciences* 13, no. 1 (2022): 1-14. <https://doi.org/10.2458/jmmss.5396>
- [20] Da, Ting. "A three-stage machine learning and inference approach for educational data." *Scientific Reports* 15, no. 1 (2025): 11521. <https://doi.org/10.1038/s41598-025-89394-2>
- [21] Ahmed, Esmael. "Student performance prediction using machine learning algorithms." *Applied computational intelligence and soft computing* 2024, no. 1 (2024): 4067721. <https://doi.org/10.1155/2024/4067721>
- [22] Grebovic, Marko, Luka Filipovic, Ivana Katnic, Milica Vukotic, and Tomo Popovic. "Machine learning models for statistical analysis." *Int. Arab J. Inf. Technol.* 20, no. 3A (2023): 505-514. <https://doi.org/10.34028/iajit/20/3A/8>
- [23] Ismail, Leila, Huned Materwala, and Alain Hennebelle. "Comparative analysis of machine learning models for students' performance prediction." In *International Conference on Advances in Digital Science*, pp. 149-160. Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-71782-7_14
- [24] Saini, Balwinder Kaur. "Optimizing student academic performance prediction using heterogeneous ensemble learning." *European Journal of Artificial Intelligence and Machine Learning* 4, no. 4 (2025): 1-6. <https://doi.org/10.24018/ejai.2025.4.4.77>
- [25] Tin, Ting Tin, Lee Shi Hock, and Omolayo M. Ikumapayi. "Educational Big Data Mining: Comparison of Multiple Machine Learning Algorithms in Predictive Modelling of Student Academic Performance: Educational Big Data Mining." *International Journal of Advanced Computer Science & Applications* 15, no. 6 (2024). <https://doi.org/10.14569/ijacsa.2024.0150664>
- [26] Abdullahi, Kabir Bindawa. "Statistical mirroring-based ordinal analysis: A sensitive, robust, efficient, and ordinality-preserving descriptive method for analyzing ordinal assessment data." *MethodsX* 15 (2025): 103427. <https://doi.org/10.1016/j.mex.2025.103427>
- [27] Bolikulov, Furkat, Rashid Nasimov, Akbar Rashidov, Farkhod Akhmedov, and Young-Im Cho. "Effective methods of categorical data encoding for artificial intelligence algorithms." *Mathematics* 12, no. 16 (2024): 2553. <https://doi.org/10.3390/math12162553>
- [28] Dhummad, Sukhdevsinh. "The imperative of exploratory data analysis in machine learning." *Scholars Journal of Engineering and Technology* 13, no. 01 (2025): 30-44. <https://doi.org/10.36347/sjet.2025.v13i01.005>
- [29] Otero-Escobar, Alma Delia, and María Luisa Velasco-Ramírez. "Study on Exploratory Data Analysis Applied to Education." In *2023 IEEE International Conference on Engineering Veracruz (ICEV)*, pp. 1-5. IEEE, 2023. <https://doi.org/10.1109/ICEV59168.2023.10329702>
- [30] Cooksey, Ray W. "Descriptive statistics for summarising data." In *Illustrating statistical procedures: Finding meaning in quantitative data*, pp. 61-139. Singapore: Springer Singapore, 2020. https://doi.org/10.1007/978-981-15-2537-7_5
- [31] Páez, Antonio, and Geneviève Boisjoly. "Exploratory data analysis." In *Discrete choice analysis with R*, pp. 25-64. Cham: Springer International Publishing, 2023. https://doi.org/10.1007/978-3-031-20719-8_2

- [32] Malik, Saleem, S. Gopal Krishna Patro, Chandrakanta Mahanty, Rashmi Hegde, Quadri Noorulhasan Naveed, Ayodele Lasisi, Abdulrajak Buradi, Addisu Frinjo Emma, and Naoufel Kraiem. "Advancing educational data mining for enhanced student performance prediction: a fusion of feature selection algorithms and classification techniques with dynamic feature ensemble evolution." *Scientific Reports* 15, no. 1 (2025): 8738. <https://doi.org/10.1038/s41598-025-92324-x>
- [33] Zaffar, Maryam, Manzoor Ahmad Hashmani, K. S. Savita, Syed Sajjad Hussain Rizvi, and Mubashar Rehman. "Role of FCBF feature selection in educational data mining." *Mehran University Research Journal Of Engineering & Technology* 39, no. 4 (2020): 772-778. <https://search.informit.org/doi/10.3316/informit.459135063399758>
- [34] Inyang, U. P., and E. A. Johnson. "Performance comparison of XG-Boost and Random Forest for the prediction of students' academic performance." *European Journal of Computer Science and Information Technology* 13, no. 2 (2025): 1-21. <https://doi.org/10.37745/ejcsit.2013/vol13n2121>
- [35] Bellaj, Mohamed, Ahmed Ben Dahmane, Said Boudra, and Mohammed Lamarti Sefian. "Educational Data Mining: Employing Machine Learning Techniques and Hyperparameter Optimization to Improve Students' Academic Performance." *International Journal of Online & Biomedical Engineering* 20, no. 3 (2024). <https://doi.org/10.3991/ijoe.v20i03.46287>
- [36] Ordoñez-Avila, Ricardo, Jaime Meza, and Sebastian Ventura. "Mining autonomous student patterns score on LMS within online higher education." *PeerJ Computer Science* 11 (2025): e2855. <https://doi.org/10.7717/peerj-cs.2855>
- [37] Ignatenko, Vera, Anton Surkov, and Sergei Koltcov. "Random forests with parametric entropy-based information gains for classification and regression problems." *PeerJ Computer Science* 10 (2024): e1775. <https://doi.org/10.7717/peerj-cs.1775>
- [38] Ngo, Giang, Rodney Beard, and Rohitash Chandra. "Evolutionary bagging for ensemble learning." *Neurocomputing* 510 (2022): 1-14. <https://doi.org/10.1016/j.neucom.2022.08.055>
- [39] YEAP, M., Stephanie Chua, and Arif Bramantoro. "A Comparative Analysis of Machine Learning Models for Prediction of Autism Spectrum Disorder Using Screening Data." *JOURNAL OF ADVANCED RESEARCH* (2024): 175-185. <https://doi.org/10.37934/araset.53.1.175185>
- [40] Subramanian, Kayalvizhi, and Gunasekar Thangarasu. "An effective air pollution prediction model using machine learning algorithms." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 47, no. 2 (2024): 68-75. <https://doi.org/10.37934/araset.47.2.6875>
- [41] Fatima, Sana, Ayan Hussain, Sohaib Bin Amir, Syed Haseeb Ahmed, and Syed Muhammad Huzaifa Aslam. "Xgboost and random forest algorithms: an in depth analysis." *Pakistan Journal of Scientific Research (PJO SR)* 3, no. 1 (2023): 26-31. <https://doi.org/10.57041/vol3iss1pp26-31>
- [42] Zuehlke, Monty-Maximilian, and Daniel Kudenko. "TCR: topologically consistent reweighting for XGBoost in regression tasks." *Machine Learning* 114, no. 4 (2025): 108. <https://doi.org/10.1007/s10994-024-06704-x>
- [43] Liu, Yuqing, Binbin Li, and Hejun Liang. "Building Electricity Prediction Using BILSTM-RF-XGBOOST Hybrid Model with Improved Hyperparameters Based on Bayesian Algorithm." *Electronics* 14, no. 11 (2025): 2287. <https://doi.org/10.3390/electronics14112287>
- [44] Leinonen, Tuija, David Wong, Antti Vasankari, Ali Wahab, Ramesh Nadarajah, Matti Kaisti, and Antti Airola. "Empirical investigation of multi-source cross-validation in clinical ECG classification." *Computers in Biology and Medicine* 183 (2024): 109271. <https://doi.org/10.1016/j.combiomed.2024.109271>
- [45] Gebreyesus, Yibrah, Damian Dalton, Sebastian Nixon, Davide De Chiara, and Marta Chinnici. "Machine learning for data center optimizations: feature selection using Shapley additive exPlanation (SHAP)." *Future Internet* 15, no. 3 (2023): 88. <https://doi.org/10.3390/fi15030088>
- [46] Chicco, Davide, Matthijs J. Warrens, and Giuseppe Jurman. "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation." *PeerJ computer science* 7 (2021): e623. <https://doi.org/10.7717/peerj-cs.623>
- [47] Hodson, Timothy O. "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not." *Geoscientific Model Development Discussions* 2022 (2022): 1-10. <https://doi.org/10.5194/gmd-15-5481-2022>
- [48] Orji, Ugochukwu, and Elochukwu Ukwandu. "Machine learning for an explainable cost prediction of medical insurance." *Machine learning with applications* 15 (2024): 100516. <https://doi.org/10.1016/j.mlwa.2023.100516>
- [49] Agyemang, Edmund F., Joseph A. Mensah, Obu-Amoah Ampomah, Louis Agyekum, Justice Akuoko-Frimpong, Amma Quansah, and Oluwaferanmi M. Akinlosotu. "Predicting Students' Academic Performance Via Machine Learning Algorithms: An Empirical Review and Practical Application." *Computer Engineering and Intelligent Systems* 15, no. 1 (2024): 86. <http://doi.org/10.7176/CEIS/15-1-09>
- [50] Maulana, Aga, Ghazi Mauer Idroes, Pati Kemala, Nur Balqis Mauldydia, Novi Reandy Sasmita, Trina Ekawati Tallei, Hizir Sofyan, and Asep Rusyana. "Leveraging artificial intelligence to predict student performance: A comparative machine learning approach." *Journal of Educational Management and Learning* 1, no. 2 (2023): 64-70. <https://doi.org/10.60084/jeml.v1i2.132>

- [51] Ariyanta, Nadindra Dwi, Anik Nur Handayani, Jevri Tri Ardiansah, and Kohei Arai. "Ensemble learning approaches for predicting heart failure outcomes: A comparative analysis of feedforward neural networks, random forest, and XGBoost." *Applied Engineering and Technology* 3, no. 3 (2024): 173-184. <https://doi.org/10.31763/aet.v3i3.1750>
- [52] Natras, Randa, Benedikt Soja, and Michael Schmidt. "Ensemble machine learning of random forest, AdaBoost and XGBoost for vertical total electron content forecasting." *Remote Sensing* 14, no. 15 (2022): 3547. <https://doi.org/10.3390/rs14153547>
- [53] Teodorescu, Vlad, and Laura Obreja Braşoveanu. "Assessing the validity of k-fold cross-validation for model selection: Evidence from bankruptcy prediction using random forest and XGBoost." *Computation* 13, no. 5 (2025): 127. <https://doi.org/10.3390/computation13050127>
- [54] Ren, Hancheng, Bo Pang, Ping Bai, Gang Zhao, Shu Liu, Yuanyuan Liu, and Min Li. "Flood susceptibility assessment with random sampling strategy in ensemble learning (RF and XGBoost)." *Remote Sensing* 16, no. 2 (2024): 320. <https://doi.org/10.3390/rs16020320>
- [55] Irfan, Muhamad, Abdul Sattar, Ahmad Sher, and Muhamad Ijaz. "Comparing Random Forest and Logistic Regression for Predicting Student Completion in Online University Courses Using Behavioral Data." *Artificial Intelligence in Learning* 1, no. 1 (2025): 1-19. <https://doi.org/10.63913/ail.v1i1.19>
- [56] Saragih, Rijoi IE. "Predicting student academic performance using random forest regression: a case study on lms behavioral data." *International Journal of Information System and Innovative Technology* 4, no. 1 (2025): 11-14. <https://doi.org/10.63322/rm0wcg63>
- [57] Winarsih, Winarsih, Heri Sutanto, and Aris Puji Widodo. "Implementation of the Ensemble Machine Learning Algorithm for Student Dropout Prediction Analysis." *Jurnal Sistem Informasi Bisnis* 15, no. 2: 159-166. <https://doi.org/10.14710/vol15iss2pp159-166>
- [58] García-Almeida, Desiderio Juan, and María Teresa Cabrera-Nuez. "The influence of knowledge recipients' proactivity on knowledge construction in cooperative learning experiences." *Active Learning in Higher Education* 21, no. 1 (2020): 79-92. <https://doi.org/10.1177/1469787418754569>
- [59] Graciano, Paola, Fernando Henrique Lermen, Fernanda Maciel Reichert, and Antonio Domingos Padula. "The impact of risk-taking and creativity stimuli in education towards innovation: A systematic review and research agenda." *Thinking Skills and Creativity* 47 (2023): 101220. <https://doi.org/10.1016/j.tsc.2022.101220>
- [60] Meekaewkunchorn, Nusanee, Katarzyna Szczepańska-Woszczyna, Chaiyawit Muangmee, Nuttapon Kassakorn, and Bilal Khalid. "Entrepreneurial orientation and SME performance: The mediating role of learning orientation." *Economics & Sociology* 14, no. 2 (2021): 294-312. <https://doi.org/10.14254/2071-789X.2021/14-2/16>
- [61] Puerta-Sierra, Lizbeth, and Rogelio Puente-Díaz. "Co-creation in entrepreneurship education: How autonomy support enhances the intention to develop entrepreneurial ideas." *Journal of Education for Business* 98, no. 8 (2023): 443-451. <https://doi.org/10.1080/08832323.2023.2224544>
- [62] Keshmiri, Fatemeh. "The effect of gamification in entrepreneurship and business education on pharmacy students' self-efficacy and learning outcomes." *BMC Medical Education* 25, no. 1 (2025): 491. <https://doi.org/10.1186/s12909-025-07052-7>
- [63] Ndlovu, Elona Nobukhosi, Patrick Ebong Ebewo, Semukele Hellen Mlotshwa, and Knowledge Shumba. "Entrepreneurial coaching and self-efficacy: A systematic review of its pedagogical integration into entrepreneurship education." *Education Sciences* 15, no. 2 (2025): 237. <https://doi.org/10.3390/educsci15020237>
- [64] Luengo-Aravena, Daniela, Patricio Cabello, and Beatriz Rodriguez-Milhomens Bachino. "Online collaborative problem-solving as a tangible outcome of digital skills in technical and vocational higher education." *Computers & Education* 218 (2024): 105079. <https://doi.org/10.1016/j.compedu.2024.105079>
- [65] Haryaka, Usfandi, Nur Khadijah Razak, Fazli Rachman, Khoe Yao Tung, and Loso Judijanto. "Integrating digital literacy, critical thinking, and collaborative learning: Addressing contemporary challenges in 21st century education." *Journal of Hunan University Natural Sciences* 52, no. 3 (2025). <https://doi.org/10.55463/issn.1674-2974.52.3.9>
- [66] Özeren, Ender. "Predicting Secondary School Students' 21st-Century Skills through Their Digital Literacy and Problem-Solving Skills." *International Education Studies* 16, no. 2 (2023): 61-75. <https://doi.org/10.5539/ies.v16n2p61>
- [67] Dvorakova, Zuzana, and Ilona Polents. "Entrepreneurship education and digital literacy as element of innovative learning." In *Digital transformation and new challenges*, pp. 253-263. Springer, Cham, 2021. https://doi.org/10.1007/978-3-030-71397-3_19
- [68] Iskandar, Joni, Gusli Chidir, and Yosua Novembrianto Simorangkir. "Digital literacy and entrepreneurial attitudes: A study of Indonesian university students." *Indonesian Journal of Management and Economic Research* 1, no. 02 (2024): 1-14. <https://doi.org/10.70508/87nr1468>

- [69] Heubeck, Tim. "Managerial capabilities as facilitators of digital transformation? Dynamic managerial capabilities as antecedents to digital business model transformation and firm performance." *Digital Business* 3, no. 1 (2023): 100053. <https://doi.org/10.1016/j.digbus.2023.100053>
- [70] Alkhodair, Majid, and Hanadi Alkhudhayr. "Harnessing industry 4.0 for SMEs: Advancing smart manufacturing and logistics for sustainable supply chains." *Sustainability* 17, no. 3 (2025): 813. <https://doi.org/10.3390/su17030813>