JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, August 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8833
Hal 495-506
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Quantitative Analysis of Training Completion Using Multivariate Linear Regression

Yudo Devianto[1,*], Saruni Dwiasnati[1], Wawan Gunawan[1], Marco Alfan Sumarto[2], Dony Ramadhan Saputra[2]

[1] Faculty of Computer Science, Universitas Mercu Buana
[2] Ministry of Communication and Digital Republic of Indonesia
Email: [1,*]yudo.devianto@mercubuana.ac.id, [2]saruni.dwiasnati@mercubuana.ac.id, [3]wawan.gunawan@mercubuana.ac.id, [4]marc004@komdigi.go.id, [5]donyrsaputra22@gmail.com.
Correspondence Author Email: yudo.devianto@mercubuana.ac.id
Submitted **08-07-2025;** Accepted **01-08-2025;** Published **14-08-2025**

Abstract

The urgency of this research stems from the strategic need to monitor and evaluate the achievements of digital training implemented by various academies under government coordination, including VSGA, FGA, DEA, TA, and GTA. In the context of the national digital transformation program, the availability of an analytical model that can predict the success of participants in completing training is critically crucial to support the achievement of the Ministry's Key Performance Indicators (KPIs). The purpose of this study is to develop a predictive model based on multivariate linear regression that combines two main variables, the percentage of participants accepted and the percentage of participants who participate in onboarding, to project the level of training completion. This model is expected to provide a quantitative and objective assessment of the effectiveness of digital training implementation in each academy. The targeted outputs of this study include the development of a predictive model with performance validation through the calculation of R^2, which yielded a value of 0.9448, as well as the provision of technical reports and data-driven recommendations for enhancing digital training governance. The Technology Readiness Level (TKT) of this study is at TKT 3, and there is evidence of conceptual validation of the predictive model based on real data collected from the implementation of the training. This stage marks the readiness of the research to continue developing the system model and implementing it on the training evaluation platform in the next stage.

Keywords: Digital Training; Predictive Model; Multivariate Linear Regression; Training Evaluation; Technology Readiness Level

1. INTRODUCTION

In today's digital era, the ability to master technology has become essential for workers across various sectors. Therefore, various digital-based training programs are designed to support the development of individual competencies, enabling them to adapt to the ongoing dynamics and challenges in the industrial world. However, the effectiveness of these training programs can vary depending on the learning methods, curriculum, and other factors applied by each organizing academy [1].

Several earlier studies have examined similar approaches and relevant literature on evaluating and predicting the effectiveness of digital training. It is worth noting that Means et al. (2010), after reviewing a range of studies, observed that students participating in well-organized online classes often perform just as well as those attending conventional, face-to-face sessions. In fact, under certain circumstances, online instruction has been found to deliver even better results than traditional classroom teaching [2]. Other studies have employed descriptive statistics, comparative analysis, and predictive modeling, such as multivariate linear regression and logistic regression, to evaluate graduation rates and participant performance in various digital education settings [3][4][5][6][7][8]. These methods enable researchers to identify significant factors contributing to training completion and to develop targeted interventions that can improve outcomes.

Furthermore, studies have employed both parametric and non-parametric statistical tests, such as the t-test, ANOVA, and the Mann-Whitney U test, to investigate differences in training outcomes across multiple institutions or cohorts [9][10][11]. Additionally, the application of correlation analysis, particularly Spearman correlation, has been practical in identifying the relationships between onboarding rates, acceptance rates, and training completion [12][13]. Research focusing on Indonesian digital training initiatives such as the *Kartu Prakerja* program also underscores the positive impact of digital literacy and structured onboarding on participant success [5][14][15].

The primary gap and difference of this study compared to prior research is the integration and comparative validation of multiple statistical approaches, namely Spearman correlation, the Mann-Whitney U test, and multivariate linear regression, applied to real data from several government-coordinated digital academies in Indonesia. Unlike much of the earlier work, which often looks at individual academies in isolation, this study draws on a combined dataset from several programs. This wider lens helps reveal more profound insights into what drives participants to complete digital training. It supports the formulation of recommendations that are grounded in evidence and specifically relevant to the conditions in Indonesia. Moreover, while past research has highlighted the importance of digital skills and training program design, there remains a lack of studies quantitatively modeling the combined effect of participant acceptance and onboarding rates on graduation using multivariate regression, particularly within Indonesia's national digital transformation agenda. By developing and validating a predictive model with a high R^2 value, this research addresses that gap and offers practical insights for improving the digital training program outcomes [6][7][8].

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, August 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8833
Hal 495-506
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Digital training is adequate as, or comparable to, face-to-face training. A meta-analysis by Means et al. (2010) found that, on average, participants who learned online performed slightly better than those who learned face-to-face (effect size = 0.20, $p < 0.001$). Other review studies have also reported that blended or online learning can produce the same or higher outcomes as conventional methods. This suggests that digital training can improve learning outcomes on par with or even exceed traditional training when well-designed [2].

One indicator of the success of a digital training program is the participant completion rate, which measures the extent to which participants successfully complete the training and obtain certification [3]. Based on the data collected, several academies organize digital training programs with varying numbers of participants at each stage, from registration and onboarding to accreditation. However, it remains unclear whether there is a significant difference in completion rates between academies and what factors contribute to this difference [4].

Competency and graduation enhancement in a vocational context. In vocational education, the acquisition of digital skills is positively related to graduation rates. Mbambo and du Plessis' (2025) study in South African vocational-technical colleges found that students' lack of digital skills was significantly correlated with low graduation rates; therefore, increasing digital literacy is recommended to improve throughput and graduation rates [5]. The example of Indonesia's national program, Kartu Prakerja, also shows a positive impact: according to a BPS survey, 87% of participants felt that their work competence had increased after taking the digital training (Program Kartu Prakerja | Department of Economic and Social Affairs) [14]. This suggests that digital-based training can enhance competency and increase the likelihood of passing certification for vocational training participants.

An analysis of the variation in completion rates across different academies can provide valuable insights for training organizers and policymakers in their efforts to optimize program effectiveness [15]. The insights gained from this study can inform the development of more effective strategies to support participant success and ensure that training initiatives provide meaningful benefits for all involved [16].

2. RESEARCH METHODOLOGY

2.1 Research Questions

Based on the background described, this study formulates several main questions that will be the focus of the analysis. First, what is the graduation rate of participants in each academy involved in the digital training program? Second, is there a significant difference in graduation rates between academies, and what factors influence these differences? Third, what is the relationship between the number of participants registered, accepted, onboarded, and completed training and the graduation rate, and what recommendations can be given to improve the graduation rate in the digital training program? These questions are expected to provide a comprehensive picture of the training program's effectiveness and serve as the basis for developing strategies to improve participant success in the future.

2.2 Methodological Approach

This problem formulation will be answered through statistical analysis and data evaluation to understand participant graduation patterns and identify steps that can be taken to improve training effectiveness.

2.3. Research Stages and Workflow

This research began by collecting data from multiple digital academies, followed by a cleaning process to ensure the data was accurate and consistent. Afterward, statistical methods were used to compare the pass rates across the academies. The results of this analysis helped inform recommendations aimed at enhancing the effectiveness of future digital training programs.

A research flowchart has been created to visually represent the entire research process, from initial data gathering to the final recommendations. This diagram illustrates the stages that have been completed and will be completed, as well as their relationship to previous research.

To clarify the flow of this research, the overall stages are illustrated in Figure 1 below. Each stage represents a key step, from the initial acquisition and preparation of data through to statistical analysis, modeling, and the formulation of recommendations.

Figure 1 below illustrates the overall research workflow, highlighting each key stage from data collection to the formulation of recommendations.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, August 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8833
Hal 495-506
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

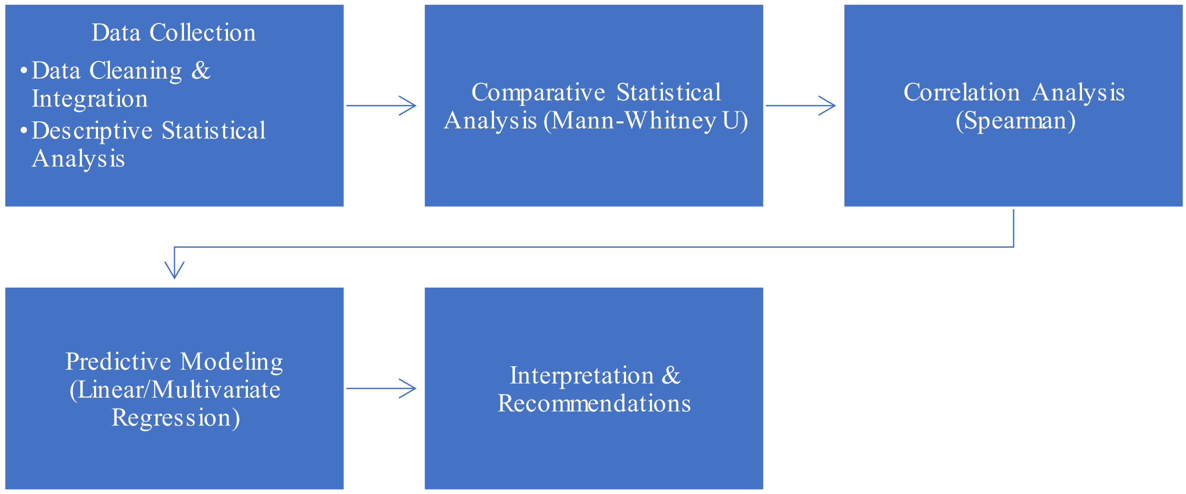

Figure 1. Research Workflow Diagram

Description of Each Research Stage:

a. Data Collection: Gathering raw data from various government digital academies (VSGA, FGA, DEA, TA, GTA), including variables such as registration, acceptance, onboarding, completion, and certification.
   1. Data Cleaning & Integration: Removing missing or inconsistent values, standardizing formats, and integrating all data into a single dataset for comprehensive analysis.
   2. Descriptive Statistical Analysis: Calculating and summarizing key metrics (e.g., participant numbers at each stage, completion rates) to provide an overview of each academy's performance.

b. Comparative Statistical Analysis: Applying non-parametric tests such as the Mann-Whitney U test to determine whether significant differences exist between academies in terms of graduation or completion rates.

c. Correlation Analysis: Using the Spearman correlation coefficient to explore the relationships between variables, such as between onboarding rates and final completion rates.

d. Predictive Modeling: Developing linear or multivariate regression models to predict completion rates based on variables such as acceptance and onboarding percentages.

e. Interpretation & Recommendations: Interpreting results, drawing practical conclusions, and proposing evidence-based recommendations for improving digital training program outcomes.

2.4 Descriptive Statistical Analysis

To address the issue of differences in graduation rates between academies in digital training programs, the following approach will be applied. In this descriptive analysis, the first step taken is to calculate the pass rate of digital training participants at each academy. The calculation is done using the formula: Pass Rate (%) = (Number of Participants Passing Certification / Number of Participants Taking Certification) × 100. In addition, this analysis presents summary statistics for each academy, including the number of participants at each stage of training, from registration to accepted participants, those who complete the onboarding process, and those who complete the training and take certification. The presentation of this data aims to provide an overview of the distribution of participants and their success at each academy, as well as to serve as an initial basis for identifying patterns or differences that may occur between academies. [17][18].

2.5 Comparative Statistical Analysis

To analyze the differences in graduation rates between academies, a statistical approach is employed that is suitable for the characteristics of the data. Given that this dataset contains only two academies as the objects of analysis, the comparison of graduation rates is conducted using a non-parametric statistical test, specifically the Mann-Whitney U test. This test was chosen because it does not assume a normal distribution and is more appropriate for small samples or data that is not normally distributed. This analysis aims to determine whether there is a statistically significant difference in graduation rates between the two academies involved in the digital training program. The results of this test are expected to provide empirical information that supports program evaluation and the formulation of strategies to improve the effectiveness of training in each academy [9], because the sample size is too small to use ANOVA or other parametric tests [10][11]. If there is more data on other academies (by combining datasets from other sheets), the statistical test can be extended to the Kruskal-Wallis test.

2.6 Analysis of Relationships Between Variables

Analysis of the Relationship of Factors to Graduation Using Spearman Correlation [12] to see the relationship between the number of applicants, the number accepted, the number of participants who completed the training, and the graduation

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, August 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8833
Hal 495-506
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

rate. Use logistic regression to examine the effect of onboarding factors and the number of participants who complete the certification on the graduation rate [13].

2.7 Review of Applied Methods and Algorithms

This study employs several statistical methods and algorithms to analyze and model the factors influencing digital training completion rates. A brief literature review of each method is provided below to justify their use and contextual relevance.

Descriptive statistics are fundamental for summarizing and visualizing data distributions, participant demographics, and key performance indicators in digital training [19][20]. This approach provides a clear overview and initial understanding before deeper statistical testing.

The Mann-Whitney U test is a non-parametric test used to assess whether there are significant differences between two independent groups. It does not require normality assumptions and is widely used for educational research involving small or non-normally distributed samples [9][17][18]. This method is suitable for comparing graduation rates between academies.

Spearman's rank correlation coefficient is a non-parametric measure of the strength and direction of association between two ranked variables. It is effective for analyzing monotonic relationships where variables may not follow a normal distribution. Spearman correlation is commonly applied in studies investigating educational factors and participant progression [10][11].

Linear regression and its multivariate extension are predictive modeling techniques that estimate the relationship between one dependent variable and one or more independent variables. These methods are widely used for projecting training outcomes and identifying key predictors in educational research [14][15][16]. In this study, regression models are used to predict the completion rate based on acceptance and onboarding percentages.

2.8 Growth and Impact of Digital-Based Training

Digital-based training has grown rapidly in recent years, with various studies showing a significant impact on learning effectiveness and participant success rates. However, the pass rate of participants in digital training programs between academies still shows considerable variation, which is influenced by several factors, including training design, the technology used, and the readiness of participants and the organizing institutions. Recent research indicates that digital-based training can make a significant contribution to enhancing participant performance. A study by Widihartono and Ahmadi (2024) found that training programs designed with a digital context in mind increased employee productivity by an average of 34%, compared to 21% with conventional methods. Additionally, the integration of technologies such as artificial intelligence (AI), virtual reality (VR), and augmented reality (AR) into training programs has been demonstrated to enhance knowledge retention by up to 65% compared to traditional methods [21].

2.9 Evaluation and Efficiency of Digital Training

In the context of learning evaluation, research by Aisy et al. (2024) demonstrates that digital-based training is practical in supporting the planning, implementation, evaluation, and follow-up of training. This training is also efficient in collecting data on participant and instructor achievements [19].

2.10 Factors Influencing Digital Training Effectiveness

Other relevant studies show that various factors influence the effectiveness of digital training. A survey by Setiawan and Casmiwati (2024) examined the efficacy of the Competency-Based Training Program at the Surabaya City Job Training Center. The results showed that the program had not been running effectively, as indicated by the lack of program targeting accuracy, suboptimal socialization, suboptimal goal achievement, and inadequate program monitoring. This study suggests that training should prioritize school graduates who are aligned with their majors and establish a structured socialization schedule [20].

2.11 Key Drivers of Graduation Rate Differences

Based on various studies reviewed, differences in graduation rates between academies in digital training programs are influenced by several key factors. Among them is the adoption of more sophisticated digital technologies, such as AI, VR, and AR, which have been proven to increase participant engagement and retention [22]. In addition, the use of digital-based evaluation and monitoring systems allows organizers to more quickly identify participants who are experiencing difficulties so that interventions can be carried out earlier [23]. Other factors include targeted training planning and implementation, such as adjusting training materials to meet participant needs and increasing socialization activities [24]. No less critical, the academy's readiness to support participants both in terms of technical aspects, infrastructure, and the quality of interaction with instructors also contributes to the success of the training.

3.5 Recommendations for Future Research

For further research, it is recommended to conduct comparative studies between academies that have significant differences in the application of technology, evaluation systems, and training approaches. Such studies will help identify the main factors that most influence variations in participant graduation rates in digital training programs. These findings

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, August 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8833
Hal 495-506
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

are expected to provide more specific and applicable recommendations for academies to increase the effectiveness of the digital training programs they organize.

3. RESULT AND DISCUSSION

Digital training programs from the Ministry of Communication and Informatics (Kominfo) are designed for various segments of society through the Digital Talent Scholarship initiative. The Vocational School Graduate Academy (VSGA) targets vocational high school graduates or their equivalents to equip them with digital work skills through BNSP certification, such as Junior Web Developer and Network Administrator. The Fresh Graduate Academy (FGA) is intended for D3/D4/S1 graduates to develop the latest competencies, such as Data Scientist and Cloud Engineer, through global collaboration (Cisco, AWS, Google). The Digital Entrepreneurship Academy (DEA) focuses on empowering MSMEs and entrepreneurs through digital entrepreneurship training. At the same time, the Thematic Academy (TA) offers thematic training tailored to local potentials, such as tourism and the creative economy. For ASN, the Government Transformation Academy (GTA) provides digital government transformation materials, including those related to data management and information security. All of these initiatives are integrated with the use of Information and Communication Technology (ICT), which serves as the primary foundation for building an inclusive and sustainable national digital ecosystem.

3.1. Research Data

The available data is processed in the form of an Excel table, which contains six digital training academies (VSGA, FGA, DEA, TA, GTA, ICT). Variables used: Academy, Applicants, Accepted, Onboarding, Completed, Entitled to Certification, Participated in Certification, Passed Certification. The six digital trainings have complete data that complete the training, and two (VSGA, FGA) trainings have a follow-up program, namely competency certification for participants who have completed the training activities according to the requirements determined by the activity committee. **Table 1** presents the dataset used in this study, detailing the number of participants at each stage across different digital training academies.

Table 1. Digital Training Dataset

Academy	Registrant	Accepted	Onboarding			Finish	Entitled to Certification	Participate in Certification	Passed Certification
			M	F	Total				
VSGA	2562	1349	875	457	1332	1276	1276	1180	1044
FGA	876	222	127	91	218	149	140	0	0
DEA	4739	2196	551	1643	2194	2146	0	0	0
TA	1778	1383	519	836	1355	1354	0	0	0
GTA	1619	1094	632	435	1067	1032	0	0	0

3.2. Descriptive Analysis

The data cleaning process is carried out to ensure the quality and consistency of the information to be analyzed. The initial stage includes removing missing values so as not to interfere with the analysis results or cause bias. Furthermore, data is combined from various stages of training at each academy, such as VSGA, FGA, DEA, TA, and GTA, to produce a single, integrated dataset that comprehensively reflects the entire training process. This step is essential to get a comprehensive picture of the effectiveness of the program and the achievements of participants at each academy; the results of the descriptive analysis [25]. **Table 2** provides a summary of the descriptive analysis results, showing the percentage of participants who completed each stage of the digital training program.

Table 2. Descriptive Analysis Results

Academy	Registrant	Accepted	Onboarding			Finish		Certification			% Accepted from Registration	% Onboarding of Accepted	% Completion from Onboarding	% Eligible for Certification from Completion	% Passed from Eligible for Certification
			M	F	Total	Yes	No	Entitled	Follow	Passed					
VSGA	2.562	1.349	875	457	1.332	1.276	56	1.276	1.180	1.044	52,65	98,74	95,80	100,00	81,82
FGA	876	222	127	91	218	149	69	140	-	-	25,34	98,20	68,35	93,96	-
DEA	4.739	2.196	551	1.643	2.194	2.146	48	-	-	-	46,34	99,91	97,81	-	-
TA	1.778	1.383	519	836	1.355	1.354	1	-	-	-	77,78	97,98	99,93	-	-
GTA	1.619	1.094	632	435	1.067	1.032	35	-	-	-	67,57	97,53	96,72	-	-

As shown in Figure 2, the percentage of participant progress across different academies is visualized through a bar graph, which allows for easy comparison of their completion rates.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, August 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8833
Hal 495-506
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

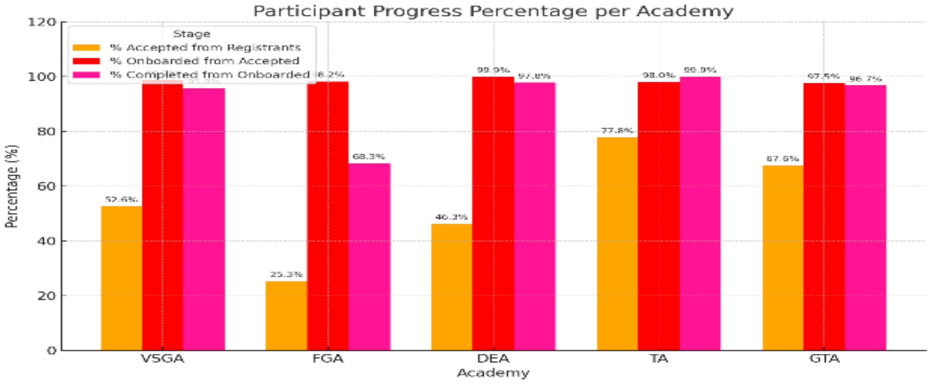

Figure 2. Percentage of Participant Progress Per Academy

Figure 2 displays a bar graph illustrating the percentage of participant progress across the five major academies, namely VSGA, FGA, DEA, TA, and GTA, based on three key stages in the training process: % Accepted from Registrants (in orange), % Onboarded from Accepted (in red), and % Completed from Onboarded (in deep pink).
Each academy has different performance characteristics in retaining participants at each stage, as explained below:

a. VSGA (Vocational School Graduate Academy) 52.65% of applicants were accepted to join the program. Of the participants who received it, 98.74% completed the onboarding process. The training completion rate was relatively high, with 95.80% of onboarding participants completing the program to the end.
b. FGA (Fresh Graduate Academy) Only 25.34% of applicants were accepted, reflecting a strict selection process. However, onboarding proceeded smoothly, with 98.20% of participants receiving the invitation to proceed to the next stage. The completion rate was relatively low, with only 68.35% of onboarding participants completing the training. This indicates challenges in participant retention during the training process.
c. DEA (Digital Entrepreneurship Academy): The acceptance rate is 46.34% of the total applicants. Almost all accepted participants take part in onboarding (99.91%). The training completion rate is exceptionally high, reaching 97.81%, which indicates the academy's effectiveness in retaining participants.
d. TA (Thematic Academy) Has a relatively high acceptance rate of 77.78%. As many as 97.98% of accepted participants proceed to the onboarding stage. TA recorded an almost perfect completion rate of 99.93%, making it one of the most stable and consistent training academies.
e. GTA (Government Transformation Academy) The participant acceptance rate is at 67.57%. Of the participants who are accepted, 97.53% successfully take part in onboarding. As many as 96.72% of onboarding participants complete the training well, indicating an efficient and structured training process.

Overall, all five academies demonstrated very high onboarding and completion rates, approaching 100%. However, it is essential to note that: FGA had significant challenges in retaining participants to the end of the training despite almost universally successful onboarding., TA and DEA were the top-performing academies, not only in terms of high acceptance rates but also in terms of retaining and ensuring participants complete the training.

This analysis can serve as a basis for evaluating the effectiveness of each academy and for improving future training processes, especially in terms of participant retention.

3.3. Correlation Analysis (Spearman)

Spearman correlation is calculated based on the ranking of each variable and then examined to determine the relationship between these rankings [26], Table 3 displays the input data used for the Spearman correlation analysis, comparing the percentage of accepted, onboarded, and completed participants across the academies.

Table 3. Spearman Correlation Input Data

Academy	% Accepted	% Onboarding	% Finish
VSGA	52.65	98.74	95.80
FGA	25.34	98.20	68.35
DEA	46.34	99.91	97.81
TA	77.78	97.98	99.93
GTA	67.57	97.53	96.72

Figure 3 displays the Spearman correlation matrix, which visualizes the relationships between the different stages of participant progression.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, August 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8833
Hal 495-506
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

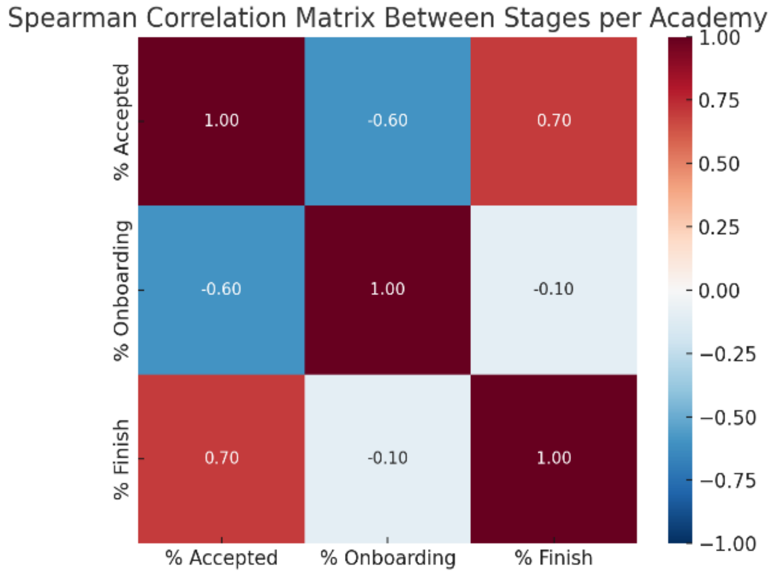

Figure 3. Spearman Correlation Matrix Between Stages Per Academy.

Table 4 below presents the percentage and ranking of each academy based on three main stages: % Accepted from Applications, % Onboarding from Accepted, and % Completed from Onboarding:

Table 4. Ranking of each column (large = rank 1).

Academy	% Accepted	A	% Onboarding	O	% Finish	F
VSGA	52.65	3	98.74	2	95.80	4
FGA	25.34	5	98.20	3	68.35	5
DEA	46.34	4	99.91	1	97.81	2
TA	77.78	1	97.98	4	99.93	1
GTA	67.57	2	97.53	5	96.72	3

The analysis of Table 4 reveals notable performance distinctions among academies in terms of participant progression. Thematic Academy (TA) stands out by ranking first in both acceptance (77.78%) and completion (99.93%), showcasing high efficiency from registration to graduation, although its onboarding ranks fourth. The Government Transformation Academy (GTA) ranks second in acceptance and third in completion but drops to fifth in onboarding, indicating possible participant drop-off at the early stage. The Digital Entrepreneurship Academy (DEA) excels in onboarding (99.91%), and completion yet ranks fourth in acceptance, suggesting a selective intake process with strong participant retention. The Vocational School Graduate Academy (VSGA) demonstrates balanced performance across all stages, ranking third in acceptance, second in onboarding, and fourth in completion. In contrast, the Fresh Graduate Academy (FGA) ranks lowest in acceptance and completion despite a moderate onboarding rank, which signals challenges in sustaining participant engagement through to the end of the program.

Calculate the difference in ranking for each pair of variables (e.g., Accepted and Completed). Table 5 shows the differences in ranking between the percentage of participants accepted and those who completed the program.

Table 5. Accepted vs Completed.

Academy	A	F	d=A-F	d^2
VSGA	3	4	-1	1
FGA	5	5	0	0
DEA	4	2	2	4
TA	1	1	0	0
GTA	2	3	-1	1
Total				6

Spearman Correlation Formula:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \quad (1)$$

The value $\boldsymbol{n} = \mathbf{5}$ is used because there are 5 academies, and each is compared between variables (% accepted, % onboarding, % completed) based on ranking. If the number of academies increases, the value of n must also be adjusted accordingly.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, August 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8833
Hal 495-506
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$$\rho = 1 - \frac{6 \times 6}{5(25-1)} = 1 - \frac{36}{120} = 1 - 0.3 = \mathbf{0.7}$$

It is the same as the result in the heatmap: $\boldsymbol{\rho}$ = **0.70** (Accepted vs Finish). Calculate the difference in ranking for each pair of variables (e.g., Accepted vs Onboarding).

Table 6 presents the differences in ranking between the percentage of participants accepted and those who completed the onboarding stage.

Table 6. Accepted vs. Onboarding

Academy	A	O	d=A-O	d^2
VSGA	3	2	1	1
FGA	5	3	2	4
DEA	4	1	3	9
TA	1	4	-3	9
GTA	2	5	-3	9
Total				32

$$\rho = 1 - \frac{6 \times 32}{5(25-1)} = 1 - \frac{192}{120} = 1 - 1.60 = \mathbf{-0.60}$$

Same as a heatmap: $\boldsymbol{\rho}$ = **-0.60** (Accepted vs. Onboarding). Calculate the difference in ranking for each pair of variables (e.g., Onboarding vs. Finish).

Table 7 illustrates the ranking differences between the percentage of participants who completed onboarding and those who finished the training.

Table 7. Onboarding vs. Finish

Academy	O	F	d=O-F	d^2
VSGA	2	4	-2	4
FGA	3	5	-2	4
DEA	1	2	-1	1
TA	4	1	3	9
GTA	5	3	2	4
Total				22

$$\rho = 1 - \frac{6 \times 22}{5(25-1)} = 1 - \frac{132}{120} = 1 - 1.10 = \mathbf{-0.10}$$

Same as a heatmap: ρ = -0.10 ((Onboarding vs. Finish). The correlation analysis between variable pairs reveals distinct relationships across stages of participant progression. A strong positive correlation (ρ = 0.70) exists between the percentage accepted and the percentage finished, indicating that academies with higher acceptance rates tend to also achieve higher completion rates; for instance, TA and GTA accepted more participants and recorded near-perfect completion levels. Conversely, a moderately strong negative correlation (ρ = -0.60) between the percentage accepted and the percentage onboarding suggests that as more participants are accepted, the proportion who proceed to onboarding tends to decline slightly, potentially due to limited onboarding capacity or mismatched readiness levels. Finally, the correlation between % Onboarding and % Finish is very weak and negative (ρ = -0.10), indicating no significant relationship; high onboarding rates do not necessarily lead to high completion rates, as demonstrated by FGA's high onboarding but low completion outcome.

The Spearman correlation analysis indicates that the percentage accepted serves as a reliable indicator of the percentage completed, suggesting that academies with higher acceptance rates tend to have a higher percentage of participants completing the program. In contrast, the percentage of onboarding shows no significant correlation with completion, indicating that merely onboarding participants does not ensure their success. Therefore, the focus should shift from increasing onboarding numbers to enhancing the quality of onboarding processes and providing adequate training support and mentoring to improve overall participant outcomes.

3.4. Linear Regression

Linear regression analysis was conducted using data from three key stages of training participant progress per academy, aiming to predict the % Completed. Three models were developed: Model 1, using % Accepted as the predictor; Model 2, using % Onboarding; and a Multivariate Model that combines both % Accepted and % Onboarding as independent variables. All models are based on the same dataset used in the previous Spearman correlation analysis, as presented in Table 3**.** This approach enables a comparative assessment of how each stage contributes to predicting training completion rates across academies [6][7].

Manual calculation steps for simple linear regression [8] for:

$$Y = a + bX \tag{2}$$

Model 1 (%Accepted vs %Finish).

$Y = a + bX \; with \; X = \%Accepted, Y = \%Finish$

Step 1. Calculate the average of $\overline{X}$ and $\overline{Y}$.

$\overline{X} = \frac{52.65+25.34+46.34+77.78+67.57}{5} = \frac{269.68}{5} = 53.936$

$\overline{Y} = \frac{95.80+68.35+97.81+99.93+96.72}{5} = \frac{458.61}{5} = 91.722$

Step 2. Calculate $\sum(X - \overline{X})(Y - \overline{Y})$ and $\sum(X - \overline{X})^2$

Table 8 outlines the components involved in calculating the simple linear regression between the percentage accepted and the percentage finished.

Table 8. Linear Regression Calculation Components: % Accepted to % Finish.

Academy	X	Y	$\mathbf{X - \overline{X}}$	$\mathbf{Y - \overline{Y}}$	$(\mathbf{X - \overline{X}})(\mathbf{Y - \overline{Y}})$	$(\mathbf{X - \overline{X}})^2$
VSGA	52.65	95.80	-1.286	4.078	-5.247	1.654
FGA	25.34	68.35	-28.596	-23.372	668.843	818.034
DEA	46.34	97.81	-7.596	6.088	-46.231	57.713
TA	77.78	99.93	23.844	8.208	195.792	568.664
GTA	67.57	96.72	13.634	5.000	68.170	185.878
Total					881.33	1631.94

Table 8 shows the results of manual calculations of mean differences, product of differences, and squared differences used to calculate simple linear regression coefficients between the variables % Accepted and % Completed at five training academies.

Step 3. Calculate the coefficient b and intercept a.

$b = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\Sigma(X-\overline{X})^2} = \frac{881.33}{1631.94} \approx 0.540$

$a = \overline{Y} - b.\overline{X} = 91.722 - 0.540\ x\ 53.936 \approx 91.722 - 29.115 \approx 62.61$

Regression Equation Result $Y = 0.54X + 62.61$

- Coefficient (b) ≈ 0.540
- Intercept (a) ≈ 62.61

Model 2 (% Onboarding vs %Finish).

$Y = a + bX\ with\ X = \%Onboarding, Y = \%Finish$

Step 1. Calculate the average of $\overline{X}$ and $\overline{Y}$.

$\overline{X} = \frac{98.74+98.20+99.91+97.98+97.53}{5} = \frac{492.36}{5} = 98.472$

$\overline{Y} = \frac{95.80+68.35+97.81+99.93+96.72}{5} = \frac{458.61}{5} = 91.722$

Step 2. Calculate $\sum(X - \overline{X})(Y - \overline{Y})\ and\ \sum(X - \overline{X})^2$

Table 9 presents the calculation components for linear regression between the percentage of participants who completed onboarding and the percentage who finished the training.

Table 9. Linear Regression Calculation Components: % Onboarding to % Finish.

Academy	X	Y	$X - \overline{X}$	$Y - \overline{Y}$	$(X - \overline{X})(Y - \overline{Y})$	$(X - \overline{X})^2$
VSGA	98.74	95.80	0.268	4.078	1.093	0.072
FGA	98.20	68.35	-0.272	-23.372	6.367	0.074
DEA	99.91	97.81	1.438	6.088	8.756	2.068
TA	97.98	99.93	-0.492	8.208	-4.038	0.242
GTA	97.53	96.72	-0.942	5.000	-4.710	0.888
Total					7.468	3.344

Table 9 shows the results of manual calculations of mean differences, product of differences, and squared differences used to calculate simple linear regression coefficients between the variables % Onboarding and % Finish at five training academies.

Step 3. Calculate the coefficient b and intercept a.

$b = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\Sigma(X-\overline{X})^2} = \frac{7.468}{3.344} \approx 2.231$

$a = \overline{Y} - b.\overline{X} = 91.722 - 2.231 \cdot 98.472 \approx 91.722 - 219.684 \approx -127.962$

Regression Equation Results$Y = 2.231X - 127.962$

- Coefficient (b) ≈ 2.231
- Intercept (a) ≈ -127.962

Multivariate model (Combined % Accepted and % Onboarding).

$\hat{Y} = a + b_1X_1 + b_2X_2, \hat{Y} = \%Finish, X_1 = \%Accepted, X_2 = \%Onboarding$

$\hat{Y}$ prediction

VSGA

$\hat{Y} = -723.030 + (0.676 \times 52.65) + (7.904 \times 98.74)$

$\hat{Y} = -723.030 + 35.5794 + 779.45296 = 93.00236$

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, August 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8833
Hal 495-506
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

FGA
$\hat{Y} = -723.030 + (0.676 \times 25.34) + (7.904 \times 98.20)$
$\hat{Y} = -723.030 + 17.12784 + 776.17488 = 70.27264$
DEA
$\hat{Y} = -723.030 + (0.676 \times 46.34) + (7.904 \times 99.91)$
$\hat{Y} = -723.030 + 31.33984 + 789.67464 = 97.98448$
TA
$\hat{Y} = -723.030 + (0.676 \times 77.78) + (7.904 \times 97.98)$
$\hat{Y} = -723.030 + 52.56528 + 774.44792 = 103.98320$
GTA
$\hat{Y} = -723.030 + (0.676 \times 67.57) + (7.904 \times 97.53)$
$\hat{Y} = -723.030 + 45.66652 + 771.88792 = 93.52444$

Table 10 displays the predicted completion rates for each academy based on the multivariate linear regression model incorporating both the percentage accepted and the percentage onboarded.

Table 10. Predicted % Finish Based on Multivariate Linear Regression Model (% Accepted and % Onboarding).

Academy	X_1	X_2	Y	a	b_1	b_2	$\hat{Y}$
VSGA	52.65	98.74	95.80	-723.030	0.676	7.904	93.00236
FGA	25.34	98.20	68.35	-723.030	0.676	7.904	70.27264
DEA	46.34	99.91	97.81	-723.030	0.676	7.904	97.98448
TA	77.78	97.98	99.93	-723.030	0.676	7.904	103.98320
GTA	67.57	97.53	96.72	-723.030	0.676	7.904	93.52444

Table 10 presents the predicted completion rates for each academy, showing that the Thematic Academy (TA) has the highest predicted completion rate at 103.98%, which is slightly above 100% due to the linear nature of the model, which lacks a natural upper bound. On the other hand, the Fresh Graduate Academy (FGA) records the lowest predicted completion rate at 70.27%, aligning with its actual performance, which is also relatively low compared to other academies. Predictions for other academies, such as VSGA, DEA, and GTA, demonstrate reasonable accuracy when compared to their actual data, indicating that the multivariate linear model performs effectively in estimating completion outcomes across different training programs.

Table 11 shows the difference between the actual and predicted values for each academy, along with the calculated sum of squared errors (SSE).

Table 11. Difference between Actual and Predicted Values and SSE (Sum of Squares Error) in Each Academy.

Academy	Y	$\hat{Y}$	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
VSGA	95.80	93.00236	2.79764	7.82679
FGA	68.35	70.27264	-1.92264	3.69654
DEA	97.81	97.98448	-0.17448	0.03044
TA	99.93	103.98320	-4.05320	16.42843
GTA	96.72	93.52444	3.19556	10.21160
SSE				38.19381

Table 12 presents the differences between actual and predicted values, along with the total sum of squares (SST) for each academy.

Table 12. Difference between Actual and Predicted Values and SST (Total Sum of Squares) in Each Academy.

Academy	Y	$Y - \bar{Y}$	$(Y - \bar{Y})^2$
VSGA	95.80	4.078	16.630
FGA	68.35	-23.372	546.250
DEA	97.81	6.088	37.064
TA	99.93	8.208	67.371
GTA	96.72	4.998	24.980
SST			692.295

$\bar{Y} = \frac{95.80+68.35+97.81+99.93+96.72}{5} = \frac{458.61}{5} = 91.722$
$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{38.19381}{692.295} = 1 - 0.05517 = 0.94483$

The comparison between the Y (Actual) and Y Predicted columns shows that the model exhibits good and consistent predictive performance, as indicated by an R^2 value of 0.9448. This means that approximately 94.48% of the variation in the % Complete data can be explained by the combination of % Accepted and % Onboarding.

This model can be used as a tool to project future training effectiveness, with the note that linear models have limitations in dealing with extreme values or nonlinearity in the data.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, August 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8833
Hal 495-506
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

4. CONCLUSION

Based on the implementation results of the research, it can be concluded that a predictive model for digital training success was successfully developed using a multivariate linear regression approach, with two key variables: the percentage of participants accepted from total applicants (% Accepted) and the percentage of accepted participants who proceeded to onboard (% Onboarding). The model demonstrates excellent performance with a coefficient of determination (R^2) of 0.9448, indicating that 94.48% of the variation in training completion rate (% Completed) can be explained by these two variables. The model's predictions closely match actual data, particularly for academies like DEA and TA, which showed the highest completion rates, highlighting the effectiveness of their selection and onboarding processes. This model serves as a valuable tool for monitoring and evaluating training programs, informing strategies to enhance academy effectiveness in the future. Overall, the research makes a significant contribution to strengthening the governance of national digital training and supports the achievement of the Ministry's Key Performance Indicators (KPIs).

REFERENCES

[1] S. Abdullah, B. Baharuddin, and A. N. Tanal, "Effectiveness of Digital Training for Educational Staff in Management Information Systems," *Int. J. ASIAN Educ.*, vol. 6, no. 1, pp. 141–150, 2025.

[2] B. Means, Y. Toyama, R. Murphy, M. Bakia, and K. Jones, "Evaluation of Evidence-Based Practices in Online Learning," *Structure*, p. 66, 2010, [Online]. Available: www.ed.gov/about/offices/list/opepd/ppss/reports.html

[3] L. M. Romero-Rodríguez, M. S. Ramírez-Montoya, and I. Aguaded, "Determining factors in MOOCs completion rates: Application test in energy sustainability courses," *Sustain.*, vol. 12, no. 7, 2020, doi: 10.3390/su12072893.

[4] B. Celik and K. Cagiltay, "Uncovering MOOC Completion: A Comparative Study of Completion Rates from Different Perspectives," *Open Prax.*, vol. 16, no. 3, pp. 445–456, 2024, doi: 10.55982/openpraxis.16.3.606.

[5] G. P. Mbambo and E. C. du Plessis, "Evaluating technical vocational education and training college student's digital skills versus throughput rate," *Discov. Educ.*, vol. 4, no. 1, 2025, doi: 10.1007/s44217-025-00396-8.

[6] D. Jatikusumo, B. Sukowo, and Y. Devianto, "Pembelajaran Algoritma Regresi Linier Menggunakan Python untuk Sekolah (Studi Kasus : SMK Media Informatika)," *Kapas Kumpul. Artik. Pengabdi. Masy.*, vol. 3, no. 1, pp. 81–89, 2024.

[7] D. Jatikusumo and R. R. Hidayat, "Optimasi penentuan lokasi bencana alam dengan regresi linier sederhana dan berganda," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 12, no. 3, 2024.

[8] R. A. Cardoso, G. A. B. Oliveira, G. M. J. Almeida, and J. A. Araújo, "A simple linear regression strategy for fretting fatigue life estimates," *Tribol. Int.*, vol. 198, no. May, p. 109852, 2024, doi: 10.1016/j.triboint.2024.109852.

[9] M. F. Călin and E. Tuşa, "Using t-Student and U-Mann-Whitney tests to identify differences in the study of the impact of the Covid 19 pandemic in online education in schools," *Analele Stiint. ale Univ. Ovidius Constanta, Ser. Mat.*, vol. 31, no. 2, pp. 39–59, 2023, doi: 10.2478/auom-2023-0018.

[10] D. Van Den Bergh *et al.*, "A tutorial on conducting and interpreting a bayesian ANOVA in JASP," *Annee Psychol.*, vol. 120, no. 1, pp. 73–96, 2020, doi: 10.3917/anpsy1.201.0073.

[11] Z. Yu, M. Guindani, S. F. Grieco, L. Chen, T. C. Holmes, and X. Xu, "Beyond t test and ANOVA: applications of mixed-effects models for more rigorous statistical analysis in neuroscience research," *Neuron*, vol. 110, no. 1, pp. 21–35, 2022, doi: 10.1016/j.neuron.2021.10.030.

[12] A. I. Yasril and F. Fatma, "Penerapan Uji Korelasi Spearman Untuk Mengkaji Faktor Yang Berhubungan Dengan Kejadian Diabetes Melitus Di Puskesmas Sicincin Kabupaten Padang Pariaman," *Hum. Care J.*, vol. 6, no. 3, p. 527, 2021, doi: 10.32883/hcj.v6i3.1444.

[13] J. Saputra and D. Nurwidyaningrum, "Analisis faktor-faktor yang mempengaruhi kompetensi lulusan melalui tracer study prodi d4 teknik konstruksi gedung PNJ," *J. Taman Vokasi*, vol. 10, no. 1, pp. 1–9, 2022.

[14] "Program Kartu Prakerja." Accessed: Apr. 21, 2025. [Online]. Available: https://sdgs.un.org/partnerships/program-kartu-prakerja#:~:text=beneficiaries are satisfied with the,following the training courses

[15] B. N. Alarifi and S. Song, "Online vs in-person learning in higher education: effects on student achievement and recommendations for leadership," *Humanit. Soc. Sci. Commun.*, vol. 11, no. 1, pp. 1–8, 2024, doi: 10.1057/s41599-023-02590-1.

[16] R. Anand and N. Gupta, "Impact of Online Learning on Student Engagement and Academic Performance," *Prax. Int. J. Soc. Sci. Lit.*, vol. 6, no. 7, pp. 29–40, 2023, doi: 10.51879/pijssl/060703.

[17] A. Anggrawan, "Analisis Deskriptif Hasil Belajar Pembelajaran Tatap Muka dan Pembelajaran Online Menurut Gaya Belajar Mahasiswa," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 18, no. 2, pp. 339–346, 2019, doi: 10.30812/matrik.v18i2.411.

[18] A. Lianawati, "Analisis Deskriptif Burnout Mahasiswa Bimbingan dan Konseling Selama Pembelajaran Daring," *Edukatif J. Ilmu Pendidik.*, vol. 4, no. 2, pp. 1678–1685, 2022, doi: 10.31004/edukatif.v4i2.2191.

[19] R. Aisy, E. Prihatin, and D. Nurdin, "Efektivitas Implementasi Evaluasi Pembelajaran pada Diklat Berbasis Digital," *Didakt. J. Kependidikan*, vol. 13, no. 3, pp. 3705–3714, 2024, [Online]. Available: https://jurnaldidaktika.org/contents/article/view/786%0Ahttps://jurnaldidaktika.org/contents/article/download/786/588

[20] H. B. Setiawan and D. Casmiwati, "Efektivitas Program Pelatihan Berbasis Kompetensi di Balai Latihan Kerja, Kota Surabaya," *J. Polit. dan Pemerintah. Drh.*, vol. 6, no. 1, pp. 24–33, 2024.

[21] R. P. Widihartono and M. A. Ahmadi, "Pengaruh Pelatihan Terhadap Kinerja Karyawan Di Era Digital," *J. Ilm. Ekon. Manaj. DAN BISNIS*, vol. 6, no. October, pp. 23–32, 2024, [Online]. Available: http://repository.usbypkp.ac.id/3699/

[22] L. Kircik and M. Goldust, "The Role of Virtual and Augmented Reality in Advancing Drug Discovery in Dermatology," *J. Cosmet. Dermatol.*, vol. 24, no. 2, pp. 3915–3924, 2025, doi: 10.1111/jocd.70071.

[23] Z. Li *et al.*, "Design of a Dynamic Monitoring and Early Intervention System for Left-Behind Children's Learning Power," *ACM Int. Conf. Proceeding Ser.*, pp. 305–312, 2024, doi: 10.1145/3696952.3696993.

[24] A. Derder, R. Sudaria, and J. Paglinawan, "Digital Infrastructure on Teaching Effectiveness of Public-School Teachers," *Am. J.*

Educ. Pract., vol. 7, no. 6, pp. 1–13, 2023, doi: 10.47672/ajep.1719.
[25] B. S. Nickerson, M. R. Esco, G. Schaefer, E. J. Aguiar, and S. A. Czerwinski, "A descriptive analysis of sarcopenia markers in young adults with down syndrome," *Exp. Gerontol.*, vol. 199, no. December 2024, p. 112655, 2025, doi: 10.1016/j.exger.2024.112655.
[26] J. Jiang, X. Zhang, and Z. Yuan, "Feature selection for classification with Spearman's rank correlation coefficient-based self-information in divergence-based fuzzy rough sets," *Expert Syst. Appl.*, vol. 249, no. PB, p. 123633, 2024, doi: 10.1016/j.eswa.2024.123633.