Journal of Applied Data Sciences
Vol. 7, No. 3, September 2026, pp. 1999-2019

Data Driven Evaluation of the New Learning Paradigm Using Machine Learning for Optimizing Graduate Outcomes

Mesra Betty Yel[1,*], Relita Buaton[2], Yuma Akbar[3], Novriyenni[4]

[1,3]Information System, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Jakarta Timur,13440, Indonesia

[2,4]Information Technology, STMIK Kaputama, Binjai, 20714, Indonesia

(Received: January 28, 2026; Revised: March 20, 2026; Accepted: June 18, 2026; Available online: July 20, 2026)

Abstract

In preparing students to face digital literacy and critical thinking transformations rapid, universities are therefore required to design and implement learning processes that are innovative, adaptive, and differentiated, in accordance with the new learning paradigm. However, the unemployment rate in Indonesia remains high approximately 5.98% vocational high school graduates and 4.8% diploma and university graduates. To develop highly skilled human resources, higher education must strengthen the competencies of students as future agents of change entering the workforce. The persistent unemployment rate among diploma and university graduates presents a national challenge that may hinder the progress of human capital development. Therefore, this study aims to develop a classification and association model linking new learning paradigm programs to student learning outcomes, in order to generate new knowledge and identify correlations among grade point average, employment waiting period, occupational field, and graduate income. The research employs a machine learning approach using association rule mining and the K-Nearest Neighbors algorithm to analyze correlations and predict graduate outcomes. Based on data processing of 450 graduate data who participated in the new paradigm learning program, the findings indicate that graduates under the new learning paradigm with grade point average ≥ 3.50 are significantly more likely to secure employment within ≤ 2 months, support = 20%, confidence = 100% based on processing a data set of 450 data. Participants in the teaching assistance program tend to experience longer waiting periods ≥ 6 months and lower initial earnings compared to those from other new learning paradigm pathways. Conversely, graduates involved in certified internships or independent study programs demonstrate higher earnings potential and stronger academic performance. The results confirm that the new learning paradigm exerts a positive influence on graduate employability, income level, and academic achievement, especially through experiential and industry-oriented learning mechanisms.

Keywords: New Learning Paradigm, Machine Learning Based Learning Model, Association Rule Mining, K-Nearest Neighbors, Graduate Employability, Learning Outcomes.

1. Introduction

In preparing students to face rapid social, cultural, industrial, and technological transformations, higher education institutions must ensure that student competencies are relevant and aligned with contemporary needs. The principle of link and match must extend beyond merely connecting education to industry and employment; it must also anticipate the demands of an increasingly dynamic and uncertain future. Universities are therefore required to design and implement innovative learning processes that enable students to achieve comprehensive learning outcomes encompassing attitudes, knowledge, and skills. One of the national initiatives launched to enhance student competencies in accordance with evolving global challenges is the new learning paradigm through the Independent Learning Campus Program framework. Independent learning is an educational policy introduced by the Indonesian government that grants students the right to take courses outside their study program for one semester and participate in off campus learning activities for up to two semesters.

The positive impacts of this policy have been evident across several dimensions. An impact study on graduate competencies indicates that participants of independent learning programs have an average job waiting period three months shorter and earn approximately 2.2 times higher salaries than the national average [1]. However, despite the

***Corresponding author: Mesra Betty Yel (mesrabettyyel@stikomcki.ac.id)**
DOI: https://doi.org/10.47738/jads.v7i3.1152

promising framework, several challenges persist in the Indonesian higher education landscape. According to the 2022 PISA report, national literacy and numeracy levels remain relatively low, and unemployment among diploma and university graduates continues to be a pressing issue [2].

Although the new learning paradigm aims to address these issues, its implementation has not yet produced a substantial or consistent impact. Many challenges remain at the levels of students, academic programs, and institutions, particularly in selecting and managing activities. The optimization of new learning paradigm implementation can be achieved through the integration of artificial intelligence (AI) and machine learning (ML) approaches [3], such technologies can uncover hidden patterns and generate new insights into the classification and correlation between learning programs and student outcomes—including employment waiting time, job field, and graduate income. Developing a data driven model based on machine learning is therefore essential for identifying and analyzing the relationships among these variables, ultimately serving as a reference for universities and students when choosing programs. Despite the existence of multiple new learning paradigm initiatives, no comprehensive analytical framework currently provides systematic knowledge on the correlation and classification of each program's impact on learning outcomes, employment waiting period, job field, and graduate income. To address this gap, this study proposes a model that applies machine learning techniques to analyze the data implementation. The model aims to identify the optimal classification and correlation between learning program participation and graduate employability indicators specifically waiting period, average income, and job field based on students' experiential learning performance during their participation, in addition, the students' experience and competence in information technology have a significant impact on the enhancement of graduate quality [4], [5].

2. Literature Review

The increasing integration of AI and ML across nearly all aspects of modern life has intensified the need for widespread AI literacy. AI and ML technologies have become embedded within everyday technological infrastructures that are often difficult for non experts to comprehend. Understanding how AI and ML operate is essential not only for individuals pursuing careers in STEM (Science, Technology, Engineering, and Mathematics) but also for all students, as these technologies are increasingly shaping every dimension of human experience, including education, recreation, and work. AI literacy, therefore, has become a prerequisite for full participation in society. The notion of literacy extends beyond instrumental skills to include critical awareness of how power, ideology, and social structures are embedded in AI technologies and how they affect individuals and communities. The importance of critical understanding in technology has also been emphasized in recent research on Computational Thinking (CT).

Scholars argue that traditional CT competencies (e.g., decomposition, abstraction, automation) must be complemented by a critical perspective on the personal, social, and political consequences of digital technologies. In this broader and more comprehensive approach to CT, students are encouraged to learn about computational systems and infrastructures not only to become better programmers or designers, but also to engage in ethical and political inquiry about technology in real world contexts [6], this perspective positions cognitive understanding of CT concepts and their real-world application as essential to cultivating critical and reflective computational practitioners. In alignment with this critical approach, this study supports the concept of Computational Empowerment (CE), which aims to foster informed decision making and critical understanding of the role of digital technologies in personal and societal contexts [7].

Significant changes have occurred in various industrial sectors, particularly in education, prompting the Indonesian government to implement the Independent Learning Policy to address paradigm shifts in the modern era. However, there remain at least sixteen challenges in the implementation of the independent learning curriculum, especially at the program study level. These include difficulties in managing new learning paradigm activities, scheduling inconsistencies, and quota restrictions. To achieve the vision and mission of national education particularly in realizing the Pancasila Student Profile higher education institutions must improve the quality of learning through innovative models and pedagogical strategies. Although the design of the independent learning policy and the independent learning campus Program is conceptually sound, careful attention is required during implementation. Most students express positive responses to the independent learning policy, but some negative implications must be minimized. To improve educational quality and graduate competency, e learning systems incorporating innovative and adaptive technologies have been proposed. One of the latest developments is the smart adaptive e learning system, which represents the new learning paradigm through AI driven personalization and machine learning mechanisms [8].

The new learning paradigm has shown a measurable positive impact on student competence. Surveys indicate that 78.53% of students reported acquiring additional competencies, 73.08% developed broader problem-solving perspectives, and 73.4% agreed that the programs align with future employability needs [9]. With the integration of e-learning systems, AI based adaptive platforms can personalize learning experiences for individuals with diverse backgrounds using intelligent analytics and ML algorithms [10]. Among these algorithms, K-Nearest Neighbors (KNN) has been widely recognized as a supervised learning technique that classifies instances based on the majority category of their nearest neighbors [11]. The dominant class within the nearest data points becomes the predicted class label for new observations [12]. Intelligent tutoring systems have also been developed for adaptive education environments, leveraging AI to deliver individualized instruction and feedback [13]. The rapid expansion of AI technologies has permeated virtually every field, including education, where AI applications are increasingly used for both instructional delivery and learning assessment. Understanding AI and ML processes such as model construction, algorithmic decision making, and ethical reflection is therefore essential [14], [15].

2.1. Differentiated Learning and the New Learning Paradigm

Differentiated learning is designed to accommodate variations in student interests, talents, abilities, and learning styles, thereby creating an adaptive and inclusive learning environment. This literature review explores the concept, implementation, and impact of differentiated learning within the context of the Independent Learning Curriculum. Findings indicate that differentiated learning in this framework enhances student motivation, active engagement, critical thinking, reduces learning boredom, and improves academic achievement. Effective strategies include differentiation in content, process, product, and learning environment. Teacher support and professional development play crucial roles in its success, although challenges such as limited resources, time constraints, and resistance to change persist. Differentiated learning within the independent learning curriculum has the potential to significantly improve educational quality and responsiveness to individual student needs. However, further research is required to address implementation challenges and develop more effective and sustainable models. The new learning paradigm represents an advanced application of the differentiated learning approach customizing learning processes, materials, and outcomes according to the characteristics and potential of each student. This paradigm emphasizes the optimal development of individual potential, positioning the teacher as a facilitator capable of mapping student needs and creating supportive environments for diverse abilities [16].

The development of the independent learning curriculum is a strategic response to the national learning crisis. It was designed as an innovation focusing on holistic character and competency development, aligning with societal evolution, and positioning students as the center of the learning process. The new learning paradigm embedded in this curriculum highlights its dynamic and adaptive nature, encouraging educational institutions to adopt relevant and innovative models such as differentiated and diagnostic learning. These models promote deeper and more engaging competency development aligned with future needs [17], The implementation of the new learning paradigm is particularly relevant to contemporary educational demands, especially within the independent learning curriculum framework. This approach positions students as the central agents in learning, while teachers act as facilitators, motivators, and learning supporters. Differentiation in learning improves academic outcomes, motivation, and comprehension by adapting teaching methods, materials, and assessments to individual characteristics and needs. Furthermore, this paradigm encourages a more active, creative, and responsive learning process, promoting inclusivity and adaptability to curriculum changes and learner diversity [18].

A recent study examined the application of intelligent adaptive learning systems to enhance students' critical thinking skills in online learning environments. The system detected students' learning styles using the VARK questionnaire and recommended personalized content based on individual profiles. Results revealed a statistically significant improvement in students' critical thinking skills, with a significance value ($P = 0.020$) < 0.050, and higher mean scores in the experimental class compared to the control group. Statistical analysis was conducted using the independent T-Test to compare pre test and post test scores between experimental and control groups. The general formula for the T-Test is:

$$t = \frac{\bar{X}_1 \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{1}$$

$\bar{X}_1, \bar{X}_2$ = mean scores of both groups, S_1^2, S_2^2 = variance of each group, n_1, n_2 = sample sizes of the respective groups

Normality and homogeneity tests were also performed prior to the T-test to ensure data validity. The findings indicate a highly significant difference between students using the intelligent adaptive learning system and those using non adaptive e-learning, with (P = 0.020) < 0.050. Students in the experimental class achieved higher mean scores than those in the control class, demonstrating that the adaptive system effectively improved students' critical thinking and learning outcomes particularly in the Mathematical Logic course. The system employed a Bayesian Network to determine the user model and adopted the Dunn and Dunn Learning Style Model to design interfaces accommodating diverse learning preferences. The Bayesian Network followed the probabilistic model expressed as:

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \tag{2}$$

$P(H|X)$ = conditional probability of hypothesis H given data X, $P(H|X)$ = probability of data X given hypothesis H, $P(H)$ = prior probability of hypothesis H, $P(H)$ = probability of observing data X

Additionally, class intervals were determined using Sturges' formula:

$$K = 1 + 3.344 \, log \, n \tag{3}$$

and the class interval width was calculated as:

$$\text{Class Interval Width} = \frac{Maximum\ Data\ Value - Minimum\ Data\ Value}{Number\ of\ Classes} \tag{4}$$

The results confirmed that the adaptive e-learning system was successfully developed and effectively supported computer network instruction aligned with students' learning styles. The system contained nine functional modules: home page, learning style test, main dashboard, text-based content, visual & auditory materials, image & text content, quizzes, assignments, and evaluation pages. Assessment using the Honey and Mumford's Learning Styles Questionnaire revealed that students prioritized usability, motivation, acceptance, and positive attitudes toward e-learning. The Bayesian Network based user model, using parameters such as perceived ease of use, motivation, attitude, and acceptance, successfully categorized learners according to their learning preferences. The results demonstrated that the system effectively represented and supported students' diverse learning characteristics.

2.2. Application of Machine Learning in Academic Data Processing

Forecasting plays an essential role in planning and decision-making processes aimed at predicting future events [19]. In the context of education, academic performance prediction is particularly important as it enables institutions to make data driven decisions to improve learning outcomes. Among the various techniques, the K-NN algorithm has been widely applied for predicting students' academic achievement. Before model training, data normalization is required to scale attribute values within a smaller range to prevent feature dominance. Feature selection helps eliminate irrelevant attributes, while data cleaning removes outliers that could distort the model's accuracy. Achieving good academic performance is a key indicator of educational quality in universities. One of the determining factors influencing academic success is the utilization of information technology and data analytics [20]. ML has been successfully implemented across various domains, including nutritional data analysis, encompassing metabolomics, obesity prediction, and precision nutrition. Commonly used methods include Random Forest (RF), Gradient Boosting Machine (GBM), cross validation, and hyperparameter optimization techniques such as grid search and random search. Furthermore, model interpretation methods such as SHAP values, partial dependence plots (PDP), and feature importance plots provide high potential for handling complex nutritional data, including high dimensional metabolomic, genomic, and behavioral datasets. Compared to traditional statistical techniques, ML approaches demonstrate superior prediction and clustering capabilities, especially in applications such as obesity prediction and biomarker identification [21].

The field of nutritional informatics is becoming increasingly complex and high dimensional, requiring novel analytical approaches such as machine learning. Model validation techniques, including cross validation, nested cross-validation,

ISSN 2723-6471

and ensemble methods such as gradient boosting and random forest, have shown great promise in managing large scale, heterogeneous datasets. ML techniques can also process unstructured and multimodal data, such as textual dietary logs using Natural Language Processing (NLP). However, a recurring limitation is the lack of generalizability due to homogeneous training data. While ML algorithms can identify significant features, their feature importance assessments are not always precise [22]. A comparative study involving five ML algorithms Multinomial Logistic Regression, Artificial Neural Networks (ANN), Random Forest, Gradient Boosting, and Stacking methods demonstrated that the stacking algorithm achieved the best overall performance based on precision, recall, and F1 score metrics. The study also identified lifestyle, stress level, obesity, family size, and parental marital status as key factors influencing academic performance, while gender and academic level were found to have no significant impact. The predictive model employs a logistic regression function, which estimates constant parameters and regression coefficients to fit the data. For a student indexed as i, the probability that the student belongs to a particular Grade Point Average (GPA) category is given by the multinomial logistic regression function, expressed mathematically as:

$$P(Y=i \mid X)=\frac{e^{\beta_i X}}{\sum_k e^{\beta_k X}} \tag{5}$$

βi represents the regression coefficient for category i and X denotes the input feature vector. The stacking algorithm exhibited the highest predictive performance, followed by gradient boosting and random forest. Multinomial logistic regression yielded the fastest computational time but the weakest predictive performance. The most significant predictors of academic performance were physical activity, stress level, family size, parental marital status, dietary intake, previous academic performance, and obesity. Conversely, factors such as gender, age, academic level, average household income, and parental education showed minimal influence. Healthy lifestyle factors, including regular physical activity and a positive mindset, demonstrated strong positive correlations with academic achievement, while stress exerted a significant negative effect. These findings support the development of healthy lifestyle programs and early monitoring systems to identify at risk students [23]. Another study combined GPA with external variables such as parental occupation and education, address, gender, and extracurricular activities to analyze and predict academic performance using several ML algorithms, including Decision Tree, Random Forest, KNN, Support Vector Classifier (SVC), Naïve Bayes, and Gaussian models. Results indicated that the Decision Tree algorithm achieved the highest accuracy, followed by Random Forest and KNN, with all three models performing similarly [24].

The Decision Tree method partitions data recursively based on entropy or Gini index to determine optimal node splits. Random Forest aggregates multiple decision trees and determines the final classification by majority voting. KNN computes distances (typically Euclidean) between data points to classify based on proximity to the nearest neighbors. Support Vector Machines (SVM) seek the optimal hyperplane that maximizes the margin between classes, while Naïve Bayes calculates probabilities based on feature independence assumptions [20]. Various machine learning techniques such as logistic regression and decision trees can effectively uncover complex relationships among nutrition, lifestyle factors, and academic performance. Prior to modeling, data preprocessing ensures suitability for analysis. The theoretical basis of logistic regression involves the sigmoid function, which models class probability as:

$$P(y=i \mid X)=\frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\ldots+\beta_n X_n)}} \tag{6}$$

Meanwhile, Decision Tree classifiers recursively divide the dataset to maximize separation between classes using measures such as information gain or Gini impurity. According to [25], machine learning can also be applied to predict student performance using deep learning models such as Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Long Short-Term Memory with Attention (LSTM Attention). The performance of these models is typically evaluated using regression-based metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), Mean Squared Logarithmic Error (MSLE), and Mean Absolute Percentage Error (MAPE). The evaluation metrics are defined as follows: MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), R^2 (Coefficient of Determination), MSLE (Mean Squared Logarithmic Error), MAPE (Mean Absolute Percentage Error) [26].

ISSN 2723-6471

3. Methodology

To optimize the outcomes of the new learning paradigm program as a strategic effort to enhance educational quality, this study develops a classification and correlation model for the implementation of the freedom to learn program. To operationalize the optimization of independent learning outcomes, this study implements a structured research process, which is illustrated in figure 1.

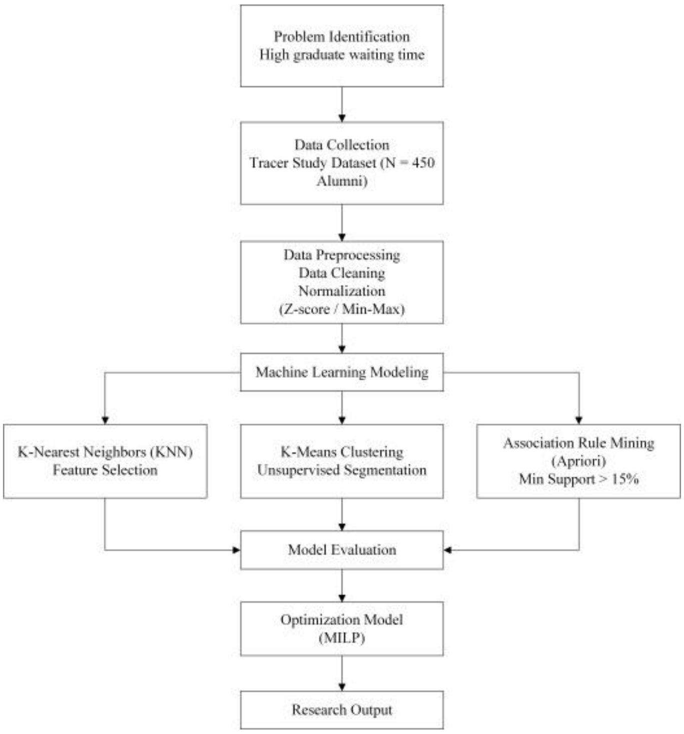

Figure 1. Association Rule Graph

The freedom to learn program initiative represents an integral component of the new learning paradigm, embodying Indonesia's educational reform to enhance graduate competencies and cultivate highly skilled human resources. Despite its progressive objectives, the implementation of independent learning campus program continues to encounter several operational and systemic challenges. Therefore, optimization through empirical research is necessary to strengthen its effectiveness and alignment with desired learning outcomes. This study employs a machine learning based modeling approach to construct and evaluate classification and association models and dataset uses alumni data from students who graduated from the independent learning class reference learning program. The research process consists of four primary stages: data collection, data processing, model construction, and model testing and validation. Data collection involved tracing alumni of independent learning programs, which included information on graduate waiting periods, academic achievement (GPA), field of employment, and income levels.

Data processing involved data cleaning, normalization, and transformation into a structured format suitable for computational analysis. Model construction using machine learning algorithms was applied to model classification and association patterns. The algorithms selected included K-Means Clustering for unsupervised clustering and KNN for classification based on similarity proximity. The number of nearest neighbors (K) determines how many closest training instances are considered in the classification of a test sample. To identify the optimal value of K, this study applies a cross-validation technique by evaluating several candidate values, namely K = 3, 5, 7, and 9. The value of K that yields the highest classification accuracy while maintaining balanced performance across classes is selected as the final K for the model. This approach ensures that the KNN model exhibits strong generalization capability when applied to previously unseen graduate data. Association Rule Mining for discovering correlations among program participation and graduate outcomes. Model testing and validation were performed by evaluating each model using support, confidence, and lift metrics to assess the strength of the relationship between variables while clustering validity was assessed using intra and inter cluster distance measures.

The model is evaluated using Precision, Recall, and F1-Score metrics. The use of Precision, Recall, and F1-Score enables a more comprehensive evaluation of the KNN model that is closely aligned with the research objective, namely measuring the impact of the new learning paradigm on graduate quality. This metric-based evaluation ensures that the model is not only statistically accurate, but also reliable and informative in supporting strategic decision making in higher education institutions related to the selection and optimization of the independent learning program. All data processing and modeling procedures were conducted on a Python based computational platform to enable seamless integration of data preprocessing, modeling, evaluation, and interactive visualization of analytical results for effective model interpretation. The scikit learn library was employed to support comprehensive model evaluation, including the implementation of KNN and K-Means clustering, as well as the assessment of classification performance using robust evaluation metrics, particularly for datasets exhibiting class imbalance.

3.1. Notation, Sets, and Parameters

Model 1. KNN–Apriori-Based Student Assignment Optimization Model

$\{1, \ldots, N\}$: index of students.
K= {1,…,10}: index of new paradigm programs (optionally, k=0 denotes baseline/non-independent learning campus program)
K= {1,…,10}: index of new paradigm programs
G, T, S, F: indices for GPA, Waiting time, Salary, and Job field categories.
Parameters:

$p_g[i,k,g]$: KNN pedicted probability that student i in program k belongs to GPA category g
$p_t[i,k,t]$: probability for Waiting Time category t.
$p_s[i,k,s]$: probability for Salary category s
$p_f[i,k,f]$: probability for Job Field f (optional)
v_g, m_t, w_s: numerical utility values for GPA, Witing Time (months), and Salary (median org log −median)R: set of association rules (from Apriori). Each rule r $\in \boldsymbol{R}$ has antecedent A_r, consequent C_r, *support* sup_r, *confidence* $con\,f_r$, *and lift* L_r.
$b_r = (L_r - 1) \times con\,f_r \times sup_r$: association weight.
Q_k: program quota; c_k: per − participant cost; $\boldsymbol{B}$: total budget.
Decision variables:
$x[i,k] \in \{0,1\}$:1 if student i is assigned to program k
$a[i,r] \in \{0,1\}$: indicator of antecedent rule r matching student i (computed, not decision)

3.2.KNN Based Outcome Estimation:

The expected outcomes for each(i,k) pair are computed as:

$$\begin{aligned} E_{GPA}[i,k] &= \sum_g (v_g \times p_g[i,k,g]) \\ E_{Wait}[i,k] &= \sum_t (m_t \times p_t[i,k,t]) \\ E_{Salary}[i,k] &= \sum_s (w_s \times p_s[i,k,s]) \end{aligned} \tag{7}$$

Role: these equations transform probabilistic KNN outputs into expected numerical values of GPA, employment waiting time, and salary for each student under each program.

Normalization (scale 0-1):

$$\hat{g}[i,k] = \frac{E_{GPA}[i,k] - g_{min}}{g_{max} - g_{min}} \tag{8}$$

$$\hat{t}[i,k] = \frac{E_{Wait}[i,k] - t_{min}}{t_{max} - t_{min}}$$

$$\hat{s}[i,k] = \frac{\log(E_{Salary}[i,k]) - \log(S_{minh})}{\log(S_{max}) - \log(S_{\min})}$$

Role: normalization places GPA, waiting time, and salary on a common scale, preventing dominance of any single variable in the optimization process.

3.3. Association Rule Translation

Rule definition and activation

Let each association rule $r \in R$ defined as $A_r \rightarrow C_r$, Where A_r denotes the antecedent and C_r denotes the consequent. For each student i a rule activation indicator.

Rule weighting, each rule is assigned a weight to reflect its strength:

$$b_r = (Lift_r - 1) \times conf_r \times sup_r \tag{9}$$

This formulation ensures that only rules with lift greater than one (associations stronger than random chance) contribute positively to the optimization process, while confidence and support reflect reliability and prevalence, respectively. The consequents of association rules are translated into optimization signals in two ways: Program level consequent, If C_r corresponds to a specific program k, the rule contributes a bonus or penalty to assigning student i to program k. Outcome level consequent, If C_r corresponds to an outcome category (low salary or long waiting time), the rule contributes a bonus or penalty directly to the utility of the predicted outcome. The cumulative contribution of rules is incorporated into the objective function through weighted bonuses and penalties, enabling the integration of empirical association patterns into the optimization framework.

3.4. Weighted Scalar Objective Function

$$\text{Maximize } Z = \sum_i \sum_k x[i,k] \times \left(w_g \hat{g}[i,k] + w_s \hat{s}[i,k] - w_t \hat{t}[i,k] + n \sum_{r \in Rk^+} a[i,r] b_r - \mu \sum_{r \in Rk^-} a[i,r] b_r \right) \tag{10}$$

$x[i,k] \in \{0,1\}$: binary decision variable indicating whether student i is assigned to program k.

$a[i,r] \in \{0,1\}$: indicator variable equal to 1 if student i satisfies the antecedent of rule r

b_r : weight of association rule r, computed from its lift, confidence, and support.

R^+_k : set of rules associated with positive recommendations or favorable outcomes for program k.

R^-_k : set of rules associated with negative outcomes for program k.

w_g, w_s, w_t : relative importance weights for GPA, salary, and waiting time, respectively, satisfying:

$\eta \mu$: scaling parameters controlling the influence of positive and negative association rules.

Role: the objective function maximizes overall graduate quality by increasing GPA and salary while minimizing employment waiting time, with adjustments based on empirical association patterns extracted from historical data.

3.5. Constraints

Each student can joint at most one program:

$$\sum_k x[i,k] \leq 1 \; \forall i \tag{11}$$

Program quota:

ISSN 2723-6471

$$\sum_{k} c_k x[i,k] \leq Q_k \ \forall k \tag{12}$$

Budget limit:

$$\sum_{i,k} x[i,k] \leq B \tag{13}$$

Minimum average GPA target (optional):

$$\sum_{i,k} E_{GPA}[i,k] x[i,k] \geq G_{bar} \sum_{i,k} x[i,k] \tag{14}$$

Maximum average waiting time target:

$$\sum_{i,k} E_{Wait}[i,k] x[i,k] \geq T_{bar} \sum_{i,k} x[i,k] \tag{15}$$

Minimum average salary target:

$$\sum_{i,k} E_{Salary}[i,k] x[i,k] \geq S_{bar} \sum_{i,k} x[i,k] \tag{16}$$

Optional fairness constraints by major, region, or gender.

3.6. Multi Objective Variant (Epsilon Constraint Formulation)

Model A (Maximize quality):

$$\max \sum x[i,k] \widehat{(g}[i,k] + \hat{S}[i,k]) \text{ s.t.} \bar{T} \leq T_{bar} \tag{17}$$

Model B (Minimize waiting time):

$$\min \sum x[i,k] \hat{t}[i,k] \ \text{ s.t.} \bar{G} \geq T_{bar}, \bar{S} \geq S_{bar} \tag{18}$$

3.7. KNN Training Details

Features: academic records, major, domicile, experience, and field preferences. Standardization: z-score normalization k selection via cross validation$(\text{e.g.}, 5, 7, \text{or } 9)$. Euclidean distance metric, Either per program KNN or unified model with indicator and probability calibration for output normalization.

4. Results and Discussion

The optimization model is formulated to assign students to independent learning campus program with the goal of maximizing learning outcomes (GPA and salary) while minimizing employment waiting time. The expected outcomes for each student program pair are estimated using two machine learning approaches: KNN to predict the probabilistic distribution of outcomes and Association Rule Mining to extract strong patterns (lift > 1), which are then translated into penalty or soft constraints in the objective function. Although only association rules with lift > 1 are considered to indicate non-random relationships, the optimization model does not assume that all such rules are uniformly beneficial. To resolve conflicting or redundant rules, the framework separates rules into positive groups, which are incorporated into the objective function with opposite effects. Each rule contribution is further moderated using a composite weight based on lift, confidence, and support, ensuring that overlapping or weak rules have limited influence. A regularization parameter is applied to prevent excessive penalties from multiple negative rules, while conflicts among activated rules for the same student program pair are resolved implicitly through net utility maximization. This design enables the optimization process to balance, attenuate, and reconcile competing rules in a robust and interpretable manner rather than treating all lift > 1 rules as equally advantageous. The main optimization model is formulated as a Mixed Integer Linear Programming (MILP) problem.

4.1. Correlation Model of the New Learning Paradigm using the Apriori Algorithm

This study investigates the association patterns between student participation across ten Independent Learning Campus Program Internship, Independent Study, Teaching Assistance, Indonesian International Student Mobility Awards (IISMA), Student Exchange, Community Service, Humanitarian Project, Research, Entrepreneurship, and Independent Program and the graduate outcome variables: GPA, employment waiting period, initial monthly salary, and employment field. The primary focus of this model is to explore associative relationships rather than causal effects, leveraging an Apriori based association rule mining approach. The alumni tracer dataset used in this analysis contains binary indicators of participation in each independent learning campus program (1 = participated, 0 = not participated), along with the four aforementioned output variables. To prepare the data for algorithmic processing, continuous variables were discretized into categorical classes as GPA includes GPA_4 (≥ 3.50), GPA_3 (3.25–3.49), Waiting Time: waiting_time_1 (≤ 2 months), waitingtime_3 (≥ 6 months), while salary_1 (lowest income tertile).

The Apriori algorithm was applied with a minimum support threshold of approximately 15% to filter out infrequent itemsets. A lift value greater than 1 indicates that the association between antecedent and consequent occurs more frequently than random chance, implying a strong positive correlation. The higher the lift value, the stronger and more meaningful the association within the dataset. The correlation model results provide an in-depth overview of how student participation across freedom to learn program correlates with graduate performance indicators GPA, employment waiting time, and initial monthly salary. To ensure interpretability, all continuous variables were discretized before the application of the Apriori association rule mining algorithm. The derived association rules were evaluated based on support, confidence, and lift metrics. A lift value greater than one (Lift > 1) indicates that the relationship between the antecedent and the consequent occurs more frequently than would be expected by random chance, meaning that the association is statistically significant and stronger with increasing lift values. The generated association rules, which provide a quantitative basis for our findings, are visualized in figure 2. This graph highlights the relationships between program participation and graduate outcomes, which are further analyzed in table 1.

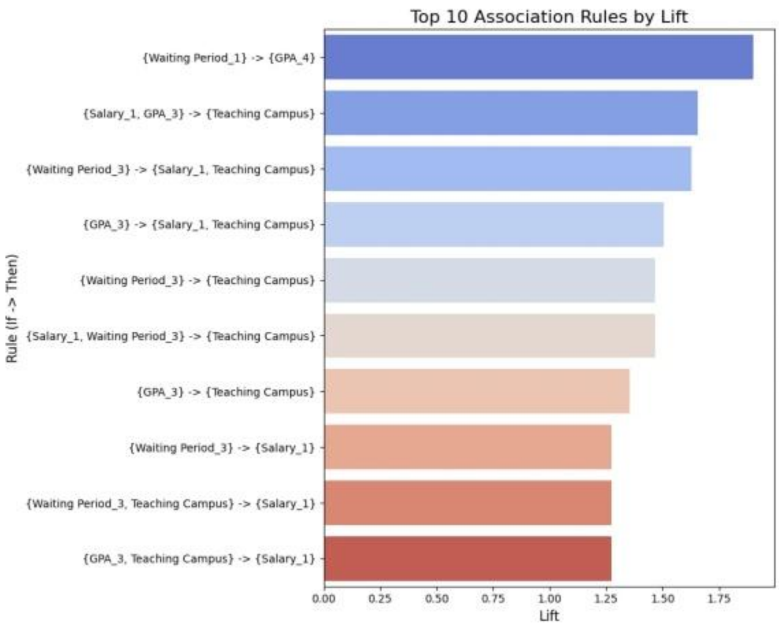

Figure 2. Association Rule Graph

Based on the association rule results in figure 2, the top association rules are produced, which are shown in table 1.

Table 1. Top Association Rules

No	Association Rule *(If → Then)*	Support (≈)	Confidence (≈)	Lift (≈)	Explanation
1	{WaitingPeriod_1} → {GPA_4}	0.2	1	1.9	Graduates who obtain employment within ≤2 months frequently have a GPA ≥ 3.50 well above random chance. This indicates that academic excellence and appropriate independent learning

					campus program placement should be prioritized to sustain this trend.
2	{Salary_1, GPA_3} → {Teaching Campus}	0.18	0.89-0.90	1.6-1.7	The combination of a lower initial salary and moderate GPA (3.25–3.49) often corresponds to participation in the Teaching Campus program. For this profile, alternative pathways such as internships or independent projects should be offered for non teaching career goals.
3	{WaitingPeriod_3} → {Salary_1, Teaching Campus}	0.18	0.73	1.6	Waiting periods ≥6 months are frequently associated with low starting salaries and Teaching Campus participation—likely due to recruitment cycles in the education sector. Implementation of service level agreements (SLA) and targeted career coaching is recommended.
4	{GPA_3} → {Salary_1, Teaching Campus}	0.24	0.84	1.5	Students with GPA 3.25-3.49 often begin their careers in education-related fields with lower initial salaries.
5	{WaitingPeriod_3} → {Teaching Campus}	0.3	0.92	1.5	Graduates with longer job search durations frequently have Teaching Campus experience. Strengthening recruitment networks with schools and EdTech companies can reduce time to hire.
6	{Salary_1, WaitingPeriod_3} → {Teaching Campus}	0.23-0.25	0.82	1.5	The combination of low starting salary and long waiting time tends to direct graduates toward Teaching Campus participation. Salary negotiation and digital teaching certification programs are needed as bridging strategies.
7	{GPA_3} → {Teaching Campus}	0.25	0.79	1.4	Moderate GPA (3.25-3.49) shows a positive association with Teaching Campus participation. This program can serve as an educational job matching pathway, or alternatively, students may be directed toward industrial internships.
8	{WaitingPeriod_3} → {Salary_1}	0.25	0.65	1.3	Long waiting periods often lead to lower initial salaries. Strengthening industry internships and certification programs can help accelerate salary growth.
9	{WaitingPeriod_3, Teaching Campus} → {Salary_1}	0.17-0.18	0.73	1.3	The combination of long waiting time and Teaching Campus participation is associated with lower starting salaries. A bundled approach combining Teaching Campus with micro credentials (EdTech, learning content, or school data) is suggested.
10	{GPA_3, Teaching Campus} → {Salary_1}	0.17-0.18	0.73	1.3	Students with moderate GPA (3.25-3.49) participating in Teaching Campus tend to start with lower salaries. Providing non teaching pathways (academic operations, EdTech roles) or sequential Teaching + Internship programs can improve career outcomes.

To strengthen the resulting rule, the next step is to determine support and confidence. To further evaluate the statistical robustness of the discovered associations, we analyzed the distribution of support and confidence values. Figure 3 presents this distribution, allowing us to identify specific rules that surpass the minimum support threshold and demonstrate meaningful lift for decision-making purposes.

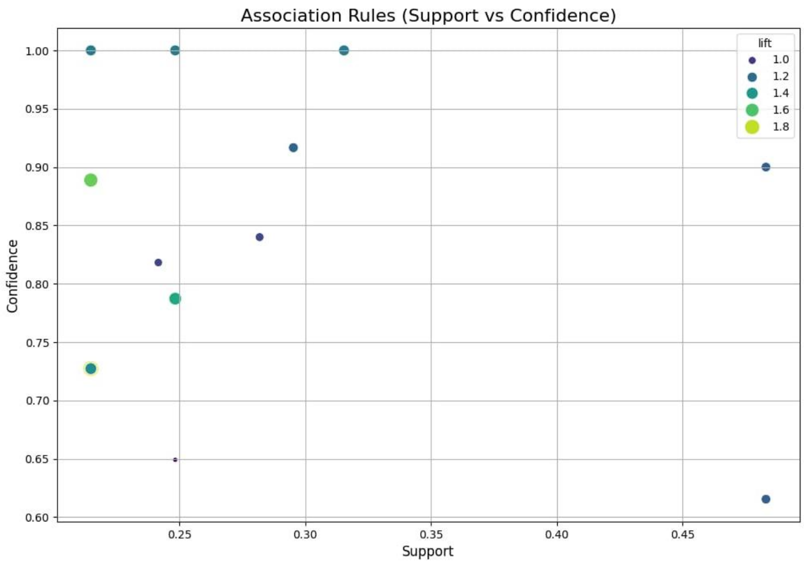

Figure 3. Rule Support and Confidence

The items GPA, Waiting Time, and Salary were discretized into categorical bins prior to rule extraction that categorized GPA: GPA_4: ≥ 3.50, GPA_3: 3.25-3.49, WaitingTime_1: ≤ 2 months, WaitingTime_3: ≥ 6 months and Salary_1: lowest income tertile (bottom 33%). The minimum support threshold was set to approximately 0.15, and only rules with Lift > 1 were retained, representing associations that are statistically stronger than random co occurrence. The primary observed patterns from the resulting association graph is students categorized as "Quick Employment" (≤ 2 months) exhibit a strong positive association with high GPA (≥ 3.50), indicating that academic performance remains a consistent predictor of employability. Several rules associate participation in the teaching assistance program with longer employment waiting times and or lower initial salary tiers. This finding reflects the sample composition and labor market characteristics of the education sector, rather than a causal relationship and the GPA_3 (3.25-3.49) group frequently co occurs with teaching assistance participation and lower starting salaries, suggesting that this profile may serve as a target group for policy or intervention strategies such as micro credential programs or alternative placement schemes to improve employability outcomes.

These findings emphasize that association rule mining can uncover latent relational structures in higher education datasets that are not directly observable through traditional statistical techniques. By mapping frequent co occurrences between program participation and graduate outcomes, institutions can identify which learning pathways yield optimal post graduation performance and which require further policy support or curriculum redesign. Two graduate personas have been identified. Persona A faster Employment with high GPA, this group represents graduates who secure jobs within a short waiting period (≤ 2 months) and achieve a high GPA (≥ 3.50). Policy emphasis should focus on maintaining high academic quality standards and aligning them with internship or research programs supported by SLA in recruitment processes, ensuring that strong academic performers also obtain competitive starting salaries, approximate proportion 20%. Persona B is more focused education oriented / teaching campus pathway, this group (as reflected in Rules #2-#7) tends to experience longer waiting periods and lower starting salaries. To improve their employability outcomes, institutions should strengthen recruitment networks with schools, educational technology (EdTech) companies, and corporate social responsibility (CSR) initiatives. Additionally, providing targeted micro credentials and salary negotiation training will be essential to enhance career readiness and compensation potential for this segment, approximate proportion 25-30%

Although the classification models demonstrate reasonably good performance for majority classes particularly for medium salary levels and mid to high GPA categories, the evaluation results also reveal significant limitations in identifying minority classes, such as high salary graduates and the longest employment waiting periods, as reflected by low recall values. This issue is primarily driven by class imbalance in the tracer study dataset, where observations in minority classes are substantially fewer than those in majority classes. Therefore, the reported model success should not be interpreted as uniformly high predictive performance across all outcome categories, but rather as evidence that the proposed KNN and association rule approaches are effective in capturing dominant patterns and directional correlations between participation in the new learning paradigm programs and graduate outcomes.

Consequently, the model is better positioned as a decision-support and analytical tool for policy evaluation rather than a final predictive system. These findings also highlight opportunities for future improvements, including the application of imbalance aware techniques such as resampling, cost sensitive learning, or longitudinal data integration to enhance sensitivity toward minority outcome classes. The terms Persona A and Persona B do not represent outcomes of an independent clustering process or subjective manual categorization. Instead, they are analytical abstractions derived systematically from empirical patterns identified through association rule mining and KNN based outcome distributions. Persona A (Fast Employment with High GPA) is derived from high support and high-confidence association rules, particularly patterns such as {Wait_1} → {GPA_4}, which consistently indicate rapid employment among graduates with strong academic performance. Persona B (education oriented / teaching campus pathway) is synthesized from a set of dominant association rules that repeatedly link participation in the teaching campus program with longer employment waiting periods and lower initial salaries.

4.2. Learning Correlation Model Using K-NN

This section presents the results of data processing using the KNN classification method, illustrating the relationship between student participation in the independent learning campus program and graduates' initial salary categories. The KNN algorithm was applied for classification using features such as independent learning campus program participation (yes/no), GPA, and job waiting period. After running the model, the salary category distribution was analyzed between independent learning campus program participants and non participants to identify potential differences. The salary categories on the X-axis (0, 1, 2, 3, 4, 5) are ordinal, with higher values representing higher income brackets. These graphs are descriptive and do not establish causality but instead serve as exploratory visualizations to identify directional patterns later validated using KNN metrics and statistical tests. To analyze the impact of the program on initial income, we evaluated the salary distributions of both participants and non participants. Figure 4 illustrates these distributions, revealing a distinct concentration of entry level salaries across all groups.

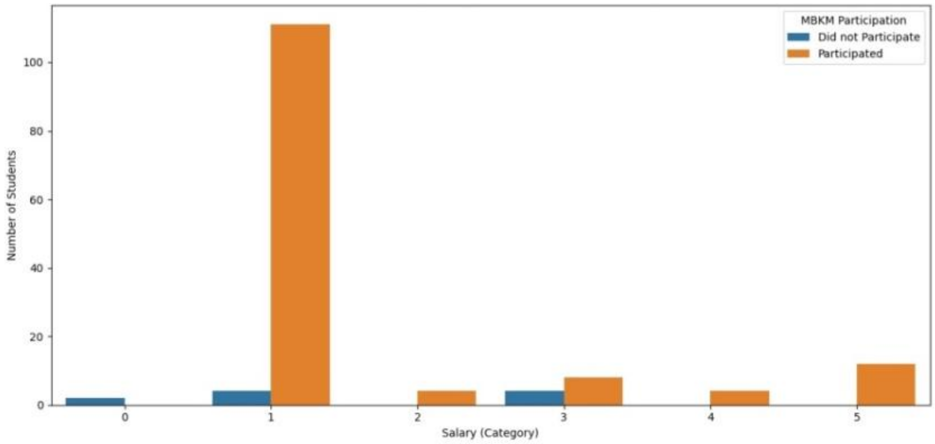

Figure 4. Salary Distribution

Figure 4 shows the dominance of category 1 (entry level salaries), independent learning campus program participants (orange bars) are heavily concentrated in salary category 1, representing entry level income. Non participants (blue bars) are also present in category 1 but in smaller absolute numbers, reflecting the smaller sample size of this group. Higher salary categories (4–5) appear exclusively among independent learning campus program participants, only freedom to learn program participants achieved salary categories 4 and 5, while non participants did not appear in the two highest salary groups. This indicates that, within the dataset, the probability of achieving a high starting salary (>0) exists exclusively among Freedom to learn program participants. Therefore, we can conclude the visualization suggests that (i) salary less or unemployed cases are more common among non participants, (ii) higher salary outcomes are only observed among Freedom to learn program participants, and (iii) category 1 remains the dominant entry point for all graduates. Therefore, strategies to increase early career income should focus on industry-oriented Freedom to Learn program tracks, recruitment SLA based partnerships, and micro credential programs that accelerate early career progression.

Next, the relationship between independent learning campus program participation (yes/no) and graduates' GPA distribution is presented. Academic achievement remains a core metric for assessing program effectiveness. Figure 5 depicts the distribution of GPA across different cohorts, illustrating how independent learning initiatives correlate with academic outcomes and helping to visualize performance disparities.

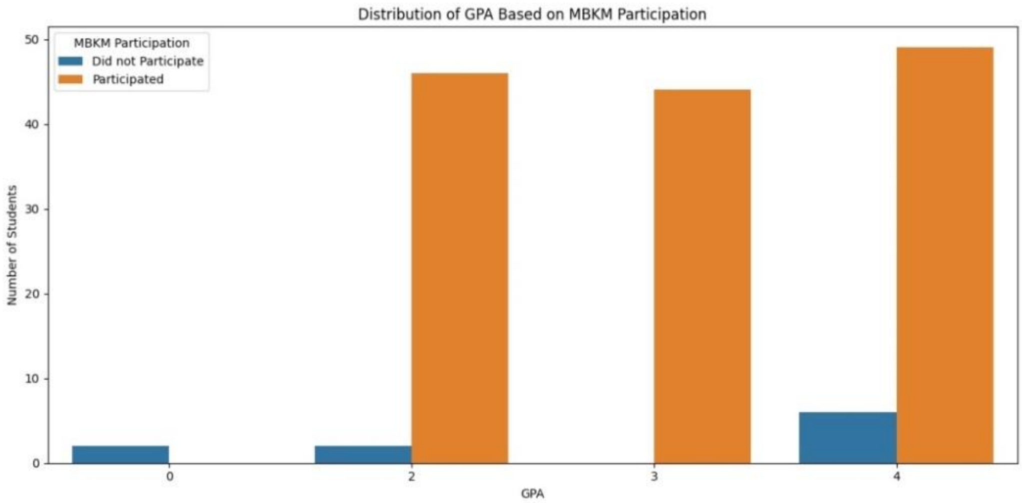

Figure 5. GPA Distribution

This visualization presents the relationship between independent learning campus program participation (yes/no) and graduates' GPA distribution. The KNN model displays the number of graduates per GPA category (X-axis: 0, 2, 3, 4 corresponding approximately to GPA_2 = 3.00–3.24, GPA_3 = 3.25–3.49, GPA_4 ≥ 3.50). Some records appear at 0, likely representing missing or outlier data. Dominance of independent learning campus program participants in higher GPA categories, the orange bars (independent learning campus program participants) dominate GPA_3 and GPA_4 categories, indicating that academically stronger profiles are more prevalent among independent learning campus program participants. Non participants are fewer and spread across extremes, the blue bars (non participants) are sparse and appear at GPA_2, GPA_4, and occasionally at GPA_0, reflecting class imbalance. The chart indicates a positive correlation between independent learning campus program participation and higher GPA distribution (3–4 categories). Thus, GPA is an informative feature for the KNN model. However, due to sample imbalance between participant and non participant groups, model evaluation should include robustness checks to ensure fairness and accuracy**.**

To examine the relationship between independent learning campus program participation and job waiting period categories. The transition from graduation to employment is a critical indicator of employability. Figure 6 categorizes job waiting periods to determine whether program participants experience faster or slower transitions into the workforce compared to non-participants, providing insights into the efficiency of specific learning pathways.

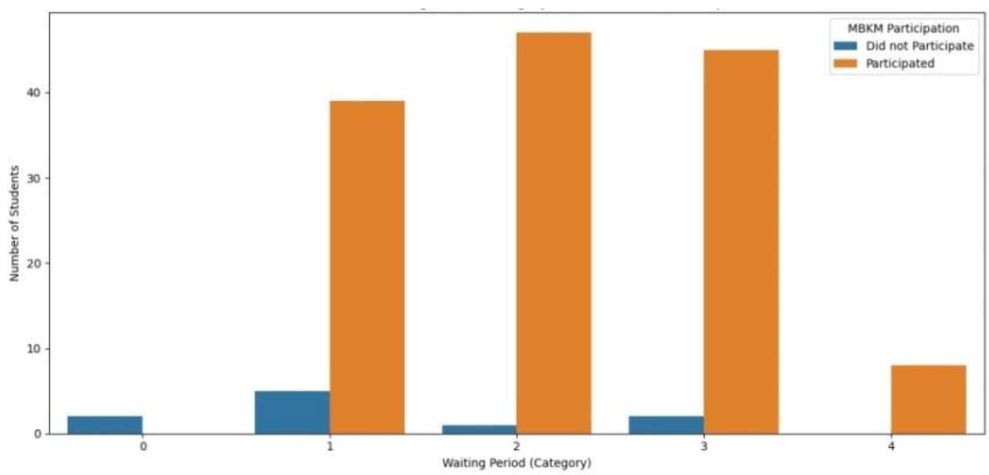

Figure 6. Waiting Period Distribution

This chart illustrates the relationship between independent learning campus program participation and the job waiting period category (X-axis: 0–4, where larger numbers indicate longer waiting periods). Broader and dominant distribution among independent learning campus program participants (Categories 2–3), orange bars (participants) show higher frequencies in categories 2 and 3, indicating many independent learning campus program graduates wait a moderate to long duration before securing employmen possibly influenced by program type. Smaller non participant group with some zero wait cases, blue bars (non participants) are few but include “0 month” cases, potentially representing pre arranged employment, family businesses, or immediate placements. Consistency with Apriori findings, previous Apriori analysis indicated that WaitingPeriod_3 was strongly associated with the teaching campus program. The dominance of orange bars in higher categories aligns with the recruitment calendar of the education sector, not a causal effect of independent learning campus program prolonging job search. Informative feature for prediction, independent learning campus program participation is a useful predictive feature for waiting period classification due to its distinct distribution pattern. Policy Implication is targeting quick win outcomes (≤2 months), Institutions should focus on improving the odds of "waiting ≤2 months" by fostering certified internships with industry partners that historically offer faster job placements.

To evaluate the classification accuracy and predictive reliability of our KNN model, we constructed confusion matrices to visualize the distribution of true versus predicted outcome categories. Figure 7 presents the confusion matrix for salary prediction, providing insights into the model's sensitivity and specific class misclassifications.

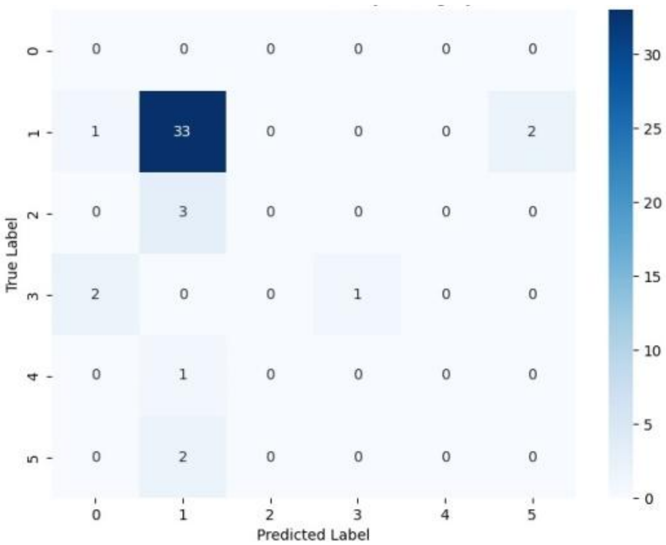

Figure 7. Confusion Matrix for Salary Prediction

Based on the result of confusion matrix for salary the following explanations, TP (True Positive): Correctly classified salary category, FN (false negative) means actual salary level misclassified as lower, FP (false positive) means incorrectly classified as higher salary, TN (true negative) means correctly excluded cases from unrelated categories. Table 2 displays the confusion matrix data for salary categories.

Table 2. Confusion Matrix for Salary Categories

	Predicted Low	Predicted Medium	Predicted High
Actual Low	TP (True Positive)	FN (False Negative)	FP (False Positive)
Actual Medium	FP (False Positive)	TP (True Positive)	FN (False Negative)
Actual High	FN (False Negative)	FP (False Positive)	TP (True Positive)

Implications for independent learning campus program is increased salary potential among independent learning campus program participants; independent learning campus program participants are more represented in the medium and high salary classes. An effective model should maximize TP rates in high salary categories while minimizing FP errors. Model refinement needs, misclassification of high performing graduates into lower salary categories suggests additional variables (e.g., independent learning campus program program type, skill certification) are required. Future data enhancement, incorporating additional factors such as technical competencies, internship experience, and industrial exposure will enhance model accuracy and salary prediction reliability. Next, we present the salary classification model data and its performance in determining its effectiveness in capturing dominant salary outcomes. table 3 displays the salary category classification report, and table 4 displays a summary of classification performance metrics for salary categories.

ISSN 2723-6471

Table 3. Salary Category Classification Report

Salary Category	Precision	Recall	F1-Score	Support
0	0	0	0	0
1	0.85	0.92	0.88	36
2	0	0	0	3
3	1	0.33	0.5	3
4	0	0	0	1
5	0	0	0	2

Based on table 3 and table 4, the KNN based salary classification model achieves an overall accuracy of 76%, indicating a satisfactory ability to predict graduate salary categories. The model performs best on salary category 1 (entry level salary), which dominates the dataset (support = 36). This class attains high precision (0.85), recall (0.92), and F1-score (0.88), demonstrating reliable identification of common early career income levels. Performance on higher salary categories (categories 2–5) is limited due to severe class imbalance, as reflected by very low support values. Consequently, precision and recall for these classes are near zero, with partial discrimination observed only in category 3. This imbalance causes the model to favor the majority class; a behavior also reflected in the gap between macro average F1-score (0.23) and weighted average F1-score (0.74). Overall, the results confirm that the proposed model effectively captures dominant salary outcomes of independent learning campus program participants, while highlighting the need for additional data and imbalance handling techniques to improve predictions for medium and high salary categories.

Table 4. Summary of Classification Performance Metrics for Salary Categories

Metric	Precision	Recall	F1-Score	Support
Macro Average	0.31	0.21	0.23	45
Weighted Average	0.74	0.76	0.74	45

To further examine the impact of the Independent Learning campus program framework paradigm on GPA. Building on the salary classification performance, we applied the same validation approach to academic outcomes. Figure 8 presents the confusion matrix for GPA prediction, which helps us identify which GPA categories are well predicted by the model and where classification errors are most frequent.

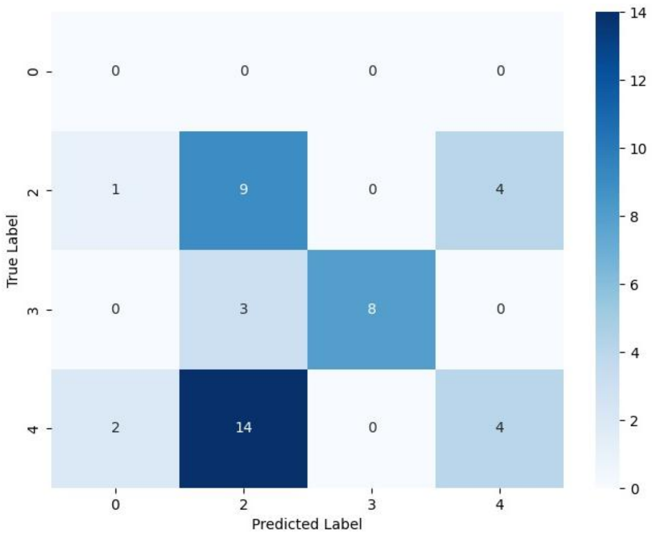

Figure 8. Confusion Matrix for GPA Prediction

Based on the result of confusion matrix for GPA categories the following explanations, True labels: Actual GPA categories, predicted labels means model predicted GPA categories. Correct classifications occur along the diagonal (TP), the model performs well in GPA 2 and 3 but struggles distinguishing GPA 4 (high achievers), often misclassifying them as GPA 2. Policy implications for freedom to learn program is focus on GPA 2 Students: Provide additional mentoring and skill based independent learning campus program (e.g., Certified Internships, Humanitarian Projects) to improve employability and income. Enhance placement for GPA 4 Graduates means strengthen career counseling and industry partnerships to improve job placements and salaries for high achieving students, who often enter the education sector with lower initial pay and future model improvement is include external features (e.g., job field, specific independent learning campus program track) to better capture labor market dynamics. To illustrate the classification of GPA against academic achievement levels, table 5 displays the GPA category classification report, and table 6 displays the summary of classification performance metrics for GPA categories.

Table 5. GPA Category Classification Report

GPA Category	Precision	Recall	F1-Score	Support
0	0	0	0	0
2 (≈ 3.00–3.24)	0.35	0.64	0.45	14
3 (≈ 3.25–3.49)	1	0.73	0.84	11
4 (≥ 3.50)	0.5	0.2	0.29	20

Based on table 5 and table 6, the classification results for GPA categories indicate moderate overall performance, with an accuracy of 47%, reflecting the increased complexity of distinguishing academic achievement levels compared to salary classification. The model demonstrates strong discriminative ability for the GPA 3 category, achieving a precision of 1.00 and a recall of 0.73, which suggests that students in the upper middle GPA range are relatively well identified. In contrast, the GPA 4 category (≥3.50) exhibits lower recall (0.20), indicating that a substantial proportion of high GPA graduates are misclassified into lower GPA categories. This pattern is likely influenced by class imbalance and overlapping feature distributions between GPA 3 and GPA 4 groups, particularly among participants of the new learning paradigm programs. The discrepancy between macro average (F1 = 0.39) and weighted average (F1 = 0.47) scores further confirms the effect of class imbalance, where model performance is dominated by majority classes. These findings suggest that while the KNN model captures general GPA trends, additional discriminative features such as learning intensity, program type, and academic workload are required to improve classification accuracy for high GPA graduates.

Table 6. Summary of Classification Performance Metrics for GPA Categories

Metric	Precision	Recall	F1-Score	Support
Macro Average	0.46	0.39	0.39	45
Weighted Average	0.57	0.47	0.47	45

The GPA confusion matrix (Fig. 7) indicates that the model demonstrates limited capability in accurately predicting high performing students (GPA ≥ 3.50). Although the model performs reasonably well for intermediate GPA categories, the recall for the highest GPA class remains relatively low. This limitation suggests that the current feature set primarily consisting of program participation, salary category, and waiting period may not sufficiently capture the multidimensional determinants of high academic performance. Several potential factors may explain this result. First, GPA is influenced by internal academic variables such as learning behavior, cognitive ability, study consistency, and assessment structure, which are not fully represented in the dataset. Second, possible class imbalance and overlap between GPA categories may reduce discriminative power in KNN classification. To address this limitation, future model refinement should incorporate additional predictive features, including technical competency indicators, certification records, project portfolios, participation intensity within programs, and longitudinal academic trajectories. Moreover, applying class balancing techniques and experimenting with alternative classification algorithms may improve predictive robustness for high achievement categories.

To assess the accuracy of graduates waiting periods after participating in the new paradigm learning program in minimizing graduate unemployment. Finally, we assessed the accuracy of predicting post-graduation employment latency. Figure 9 presents the confusion matrix for the waiting period, highlighting the model's performance in capturing both short-term job placements and potential delays in the employment transition."

ISSN 2723-6471

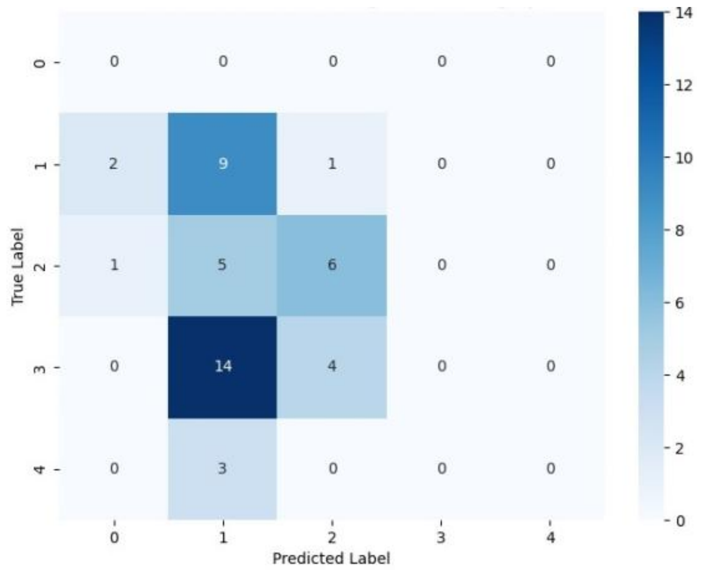

Figure 9. Confusion Matrix for Waiting Period Prediction

The model accurately predicts categories with short waiting times (1–2 months) but exhibits significant misclassifications in the longest waiting category (≥6 months). Policy Implication, to minimize graduate unemployment, independent learning campus program programs should prioritize industry-oriented partnerships, SLA based placements, and career mentoring schemes to shorten post graduation waiting periods and improve employability outcomes. Using the confusion matrix for waiting period prediction, we next examine the classification results for the time categories in predicting post graduation waiting time. Table 7 displays the waiting period classification report, and table 8 displays the summary of classification performance metrics for waiting period categories.

Table 7. Waiting Period Classification Report

Class (Category)	Precision	Recall	F1-Score	Support
0	0	0	0	0
1	0.29	0.75	0.42	12
2	0.55	0.5	0.52	12
3	0	0	0	18
4	0	0	0	3

Based on table 7 and table 8, the classification results for employment waiting time categories indicate a relatively modest overall performance, with an accuracy of 0.33. This outcome reflects the inherent difficulty in predicting post graduation waiting periods, which are influenced by multiple external factors beyond academic and program participation variables. The model demonstrates its strongest performance in Category 1 (short waiting period), achieving a recall of 0.75, suggesting that graduates who enter employment relatively quickly are detected reasonably well. Category 2 (moderate waiting period) shows balanced precision (0.55) and recall (0.50), indicating moderate discriminative capability for intermediate waiting times.

Table 8. Summary of Classification Performance Metrics for Waiting Period Categories

Class (Category)	Precision	Recall	F1-Score	Support
Macro Avg.	0.17	0.25	0.19	45
Weighted Avg.	0.22	0.33	0.25	45

However, the model exhibits poor performance in longer waiting time categories (Categories 3 and 4), where both precision and recall approach zero. This limitation is largely attributed to class imbalance, as these categories contain fewer samples, and to the complex, non academic factors affecting employment delays, such as labor market conditions, recruitment cycles, and individual job search strategies. The low macro average F1-score (0.19) further highlights uneven classification performance across categories, while the weighted F1-score (0.25) suggests that overall performance is dominated by majority classes. These findings imply that, although KNN captures short term employment trends reasonably well, predicting prolonged waiting periods remains challenging and requires richer feature representations, such as industry demand indicators, recruitment seasonality, and individual job seeking behaviors.

5. Conclusion

Graduates who participated in the new learning paradigm and achieved a GPA $\geq$ 3.50 were significantly more likely to secure employment within $\leq$ 2 months, with a support value of 20% and confidence level of 100%, indicating a strong and consistent relationship between high academic performance and faster employment outcomes. The Teaching Campus program tends to be associated with a longer job waiting period ($\geq$ 6 months) and lower initial income compared to other new learning paradigm programs, likely due to the distinct characteristics of the education sector's recruitment cycle. Graduates engaged in new learning paradigm programs demonstrated a positive influence on both income and GPA, showing higher representation within the medium and high salary classes compared to non participants. This suggests that participation in such programs contributes to better academic and career performance. The Certified Internship and Independent Study programs were observed to have a greater impact on graduate income, reflecting their alignment with industry-oriented learning outcomes and employability skill development.

6. Declarations

6.1. Author Contributions

Conceptualization: M.B.Y., R.B., and N.N.; Methodology: M.B.Y.,R.B.,Y.A.; Software: M.B.Y.,R.B.,Y.A.; Validation: M.B.Y., R.B., and N.N.; Formal Analysis: M.B.Y., R.B., and N.N.; Investigation: Y.A.; Resources: R.B.,Y.A.; Data Curation: N.N.; Writing Original Draft Preparation: R.B., Y.A., and N.N.; Writing Review and Editing: M.B.Y., R.B., and Y.A.; Visualization: N.N.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

Ministry of Higher Education, Science and Technology

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

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

References

[1] N. Wulandari, S. Wijayanti, I. Untari, and A. Kailani, "The Impact of the Merdeka Belajar-Kampus Merdeka (MBKM) Program in Improving Student Competence," *Jurnal Ilmiah Kampus Mengajar*, vol. 3, no. 2, pp. 146–162, 2023. doi: 10.56972/jikm.v3i2.133.

[2] M. R. Bilad, S. Zubaidah, and S. Prayogi, "Addressing the PISA 2022 Results: A Call for Reinvigorating Indonesia's Education System," vol. 3, no. 1, pp. 1–12, 2024. doi: 10.36312.

[3] I. Gligorea, M. Cioca, R. Oancea, A. T. Gorski, H. Gorski, and P. Tudorache, "Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review," *Education Sciences*, vol. 13, no. 12, pp. 1–27, 2023. doi: 10.3390/educsci13121216.

[4] A. Nurkhin, J. T. B. Santoso, and H. Harsono, "The Use of Artificial Intelligence in Accounting Classes: *Behavioral Insights from Students*," vol. 6, no. 3, pp. 1865–1875, 2025. doi: 10.47738/jads.v6i3.781.

[5] R. B. Buaton, Novriyenni, A. Sihombing, “Clustering and Association Rule Data Mining for Student Performance Optimization in Free Nutritious Meal Programs Using the Apriori Algorithm and K-Means,” *2025 Tenth International Conference on Informatics and Computing (ICIC), Lampung, Indonesia*, vol. 2025, no. December, pp. 1–6, 2025. doi: 10.1109/ICIC68054.2025.11309371.

[6] M. Van Mechelen, M.-L. Wagner, G. E. Baykal, R. C. Smith, and O. S. Iversen, “Digital Design Literacy in K-9 Education: Experiences from Pioneer Teachers,” in *Proceedings of the 20th Annual ACM Interaction Design and Children Conference*, vol. 2021, no. June, pp. 32–42, 2021, doi: 10.1145/3459990.3460696.

[7] O. S. Iversen, R. C. Smith, and C. Dindler, “From Computational Thinking to Computational Empowerment: A 21st Century PD Agenda,” in *Proceedings of the 15th Participatory Design Conference: Full Papers*, vol. 2018, no. August, pp. 1–6. doi: 10.1145/3210586.3210592.

[8] F. Amastini, D. Kinanti, S. Sekarhati, and M. Puspitasari, “Advancing Adaptive and Personalized E-Learning Systems: A Systematic Literature Review,” *Advances in Sustainable Science, Engineering and Technology*, vol. 7, no. 2, pp. 1–14, 2025. doi: 10.26877/m82zg802.

[9] J. H. L. C. Siahaan, S. Simatupang, F. N. Simanjuntak, L. N. Iswari, and S. R. Waruwu, “The Impact and Benefits of the Independent Curriculum Learning Independent Campus During Online Learning in the COVID-19 Period,” *Tadbir: Jurnal Studi Manajemen Pendidikan*, vol. 7, no. 1, pp. 48–64, 2023. doi: 10.29240/jsmp.v7i1.5567.

[10] R. Buaton, A. Fauzi, and M. B. Yel, “New Paradigm E-Learning Model Based on Artificial Intelligence,” *Journal of Artificial Intelligence Engineering Applications*, vol. 2, no. 1, pp. 1–4, 2022. doi: 10.59934/jaiea.v2i1.116.

[11] I. Handayani, “Application of K-Nearest Neighbor Algorithm on Classification of Disk Hernia and Spondylolisthesis in Vertebral Column,” *Indonesian Journal of Information Systems*, vol. 2, no. 1, pp. 57-66, 2019. doi: 10.24002/ijis.v2i1.2352.

[12] N. S. Altman, “An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression,” *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992. doi: 10.1080/00031305.1992.10475879.

[13] A. Klašnja-Milićević, B. Vesin, M. Ivanović, and Z. Budimac, “E-Learning Personalization Based on Hybrid Recommendation Strategy and Learning Style Identification,” *Computers & Education*, vol. 56, no. 3, pp. 885–899, 2011. doi: 10.1016/j.compedu.2010.11.001.

[14] M. H. Kaspersen, K.-E. K. Bilstrup, M. Van Mechelen, A. Hjort, N. O. Bouvin, and M. G. Petersen, “High School Students Exploring Machine Learning and Its Societal Implications: Opportunities and Challenges,” *International Journal of Child-Computer Interaction*, vol. 34, no. December, p. 100539, pp. 1-12. 2022. doi: 10.1016/j.ijcci.2022.100539.

[15] A. Sulaiman, T. B. Kurniawan, D. A. Dewi, and M. Alqudah, “Utilizing Sentiment Analysis for Reflect and Improve Education in Indonesia,” *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 189–200, 2025. doi: 10.47738/jads.v6i1.527.

[16] Y. Mulyawati, M. S. Zulela, and E. Edwita, “Differentiation Learning to Improve Students’ *Potential in Elementary School*,” vol. 6, no. 1, pp. 68–78, 2022. doi: 10.55215/pedagonal.v5i2.4485.

[17] D. Digna and C. Widyasari, “Teachers’ Perceptions of Differentiated Learning in Merdeka Curriculum in Elementary Schools,” *International Journal of Elementary Education*, vol. 7, no. 2, pp. 255–262, 2023. doi: 10.23887/ijee.v7i2.54770.

[18] D. M. Sari, F. Maulida, J. Puspita, and N. Khoirunnisa, “A Literature Review of the Implementation of Differentiated Learning in Indonesian Education Units,” *Jurnal Ilmiah Pendidikan Fisika*, vol. 7, no. 2, pp. 250–264, 2023. doi: 10.20527/jipf.v7i2.8429.

[19] M. Ula, A. F. Ulva, Mauliza, M. A. Ali, and Y. R. Said, “Application of Machine Learning in Predicting Children’s Nutritional Status With Multiple Linear Regression Models,” *Multica Science and Technology Journal*, vol. 2, no. 2, pp. 124–130, 2022. doi: 10.47002/mst.v2i2.363.

[20] M. H. Arief and M. K. Fadhil, “Educational Data Mining for Student Academic Performance Analysis,” *Jurnal Teknologi Informasi dan Terapan (J-TIT)*, vol. 11, no. 2, p. 83, pp. 83-90, 2024. doi: 10.25047/jtit.v11i2.434.

[21] D. Kirk, E. Kok, M. Tufano, B. Tekinerdogan, E. J. M. Feskens, and G. Camps, “Machine Learning in Nutrition Research,” *Advances in Nutrition*, vol. 13, no. 6, pp. 2573–2589, 2022. doi: 10.1093/advances/nmac103.

[22] Q. Tan, “Enhancing Loan Approval Prediction Using Ensemble Machine Learning Techniques Through Comprehensive Model Comparison and Performance Evaluation Analysis”, *J. Digit. Mark. Digit. Curr.*, vol. 3, no. 2, pp. 144–158, May 2026.

[23] S. Rajendran, S. Chamundeswari, and A. A. Sinha, “Predicting the Academic Performance of Middle- and High-School Students Using Machine Learning Algorithms,” *Social Sciences & Humanities Open*, vol. 6, no. 1, p. 100357, pp. 1-15, 2022. doi: 10.1016/j.ssaho.2022.100357.

[24] B. T. Kobone and T. V. Montshiwa, "Impact of Sample Size on the Robustness of Machine Learning Algorithms for Detecting Loan Defaults Using Imbalanced Data," *Journal of Applied Data Sciences*, vol. 6, no. 3, pp. 1830–1849, 2025. doi: 10.47738/jads.v6i3.713.

[25] G. Airlangga, "Predicting Student Performance Using Deep Learning Models: A Comparative Study of MLP, CNN, BiLSTM, and LSTM with Attention," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1561–1567, 2024. doi: 10.57152/malcom.v4i4.1668.

[26] G. Airlangga, "A Comparative Analysis of Machine Learning Models for Predicting Student Performance: Evaluating the Impact of Stacking and Traditional Methods," *Brilliance: Research of Artificial Intelligence*, vol. 4, no. 2, pp. 491–499, 2024. doi: 10.47709/brilliance.v4i2.4669.