International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 735-748
https://doi.org/10.70917/ijcisim-2026-2769
Article

Machine Learning Decision Support for Predicting and Profiling Student Startup Readiness in Higher Education Institutions

Roseanne Jane C. Agustin-Cera[1,4], Gloria M. Ducut[2,4], Roy R. Flores[3,4*]
[1] Department of Business Administration, Pangasinan State University – Bayambang Campus, Bayambang, Pangasinan 2423, Philippines.
Email: roseannejaneagustin@gmail.com
ORCID: 0009-0007-9778-1083
[2] Department of Business Administration, Pangasinan State University – Asingan Campus, Asingan, Pangasinan 2439, Philippines.
Email: gducut.alaminos@psu.edu.ph
ORCID: 0009-0002-6814-4046
[3] Department of Electrical Engineering, Pangasinan State University – Urdaneta City Campus, Urdaneta City, Pangasinan 2428, Philippines.
Email: rflores@psu.edu.ph
ORCID: 0009-0001-0157-6963
[4] DOST–PSU Aligwas Technology Business Incubator, Pangasinan State University – Bayambang Campus, Bayambang, Pangasinan 2423, Philippines.

Abstract: This study develops a machine learning decision support framework for assessing startup readiness among higher education institution students in Pangasinan, Philippines. Survey data from 385 students were transformed into construct level indicators for startup knowledge, policy awareness, career aspiration, entrepreneurial skills, and business resource readiness. The analysis first examined reliability, associations, predictive relationships, and construct differences using Cronbach alpha, Spearman correlation, multiple regression, and the Friedman test. It then trained Logistic Regression, Decision Tree, and Random Forest models to classify practical startup readiness using academic profile variables, startup knowledge, and policy awareness as non-leakage predictors. K-Means clustering was applied to identify student readiness profiles for institutional decision support. The constructs showed acceptable to good reliability. Startup knowledge and policy awareness were stronger than practical readiness dimensions, confirming a relative knowledge action gap. Logistic Regression obtained the best classification result with 0.66 accuracy, 0.66 weighted F1 score, and 0.729 receiver operating characteristic area under the curve. Random Forest feature importance ranked policy awareness and startup knowledge as the strongest predictors, while clustering produced four intervention-oriented profiles. The study contributes a compact analytics workflow that higher education institutions can use to screen readiness, identify student groups, and design targeted entrepreneurship support.
Keywords: startup readiness, machine learning, decision support systems, educational data mining, higher education institutions, student profiling, entrepreneurship education, Random Forest.

1. Introduction

Entrepreneurship is widely treated as a driver of innovation, employment generation, and regional competitiveness (World Bank Group, 2021). In the Philippines, this expectation is reflected in national policy instruments such as the Philippine Innovation Act and the Innovative Startup Act, which recognize higher education institutions as important actors in developing entrepreneurial capability and linking students to startup ecosystems (Ilas-Panganiban & Mitra-Ventanilla, 2020). These policy directions are especially relevant to provincial economies where universities can function as knowledge producers, skills developers, and ecosystem connectors.

The presence of entrepreneurship education, however, does not automatically produce startup-ready graduates. Students may understand business concepts, recognize the value of entrepreneurship, and express general support for innovation without having enough confidence, operational skills, networks, mentors, or resource access to initiate a venture. This mismatch is often described as a knowledge action gap: cognitive preparedness is present, but practical readiness remains constrained (Armas & Jose, 2024; Carpenter & Wilson, 2022; Lyu et al., 2023; Widati et al., 2024). The issue is not simply whether students know entrepreneurship. The harder question is whether universities can identify which students are ready to act, which students are only conceptually prepared, and which students need targeted support before they can enter an incubation or venture creation pathway.

Pangasinan provides a useful setting for examining this problem. The province has a growing higher education sector, students from varied academic programs, and increasing exposure to entrepreneurship related subjects and initiatives (Center for Pangasinan Studies, 2023; Commission on Higher Education Regional Office 1, 2026a, 2026b). Yet the support environment remains uneven across institutions, disciplines, and districts. In such a setting, a conventional mean score can describe readiness, but it cannot easily guide intervention (Ndofirepi, 2022). A program head, incubation manager, or entrepreneurship coordinator needs more operational information: who is likely to be practice-ready, which variables contribute to readiness classification, and what readiness profiles exist among students.

This paper responds to that decision problem by reframing student startup readiness as an applied educational data analytics task. The study presents descriptive and inferential analyses, and extends them through supervised classification and unsupervised profiling. Logistic Regression, Decision Tree, and Random Forest models are used to classify practical startup readiness, while K-Means clustering is used to reveal readiness profiles. The resulting workflow is intended for higher education decision support rather than for replacing academic judgment. It provides a structured basis for screening, mentoring allocation, incubation referral, and resource planning.

The study contributes in four ways. First, it develops a machine learning decision support formulation for a multidimensional startup readiness instrument. Second, it uses non-leakage predictors for supervised classification by excluding the readiness dimensions used to compute the target variable. Third, it combines explainable model output, particularly Random Forest feature importance, with clustering-based readiness profiles. Fourth, it translates the statistical and machine learning results into intervention pathways that can be used by higher education institutions in entrepreneurship ecosystems.

1.1 RELATED WORK:

1.1.1 Startup Readiness in Higher Education

Entrepreneurial readiness is not a single trait. It combines cognitive, behavioral, motivational, and structural dimensions. Students need to understand what a startup is and how it operates, but they also need aspiration, persistence, problem solving ability, communication skills, risk management, access to mentors, and resource awareness (Iwu et al., 2024; Ng et al., 2023; Panakaje et al., 2025). Prior studies show that entrepreneurship education can improve intention and awareness, especially when it strengthens self-efficacy, alertness, opportunity recognition, and perceived feasibility (Kim et al., 2022; Otache et al., 2023). The effect is weaker when entrepreneurship is treated mainly as a classroom topic rather than as a practice-based learning process.

The readiness instrument used in this study reflects this multidimensional view. It measured knowledge of business startup, career aspirations toward entrepreneurship, policy awareness and perception, entrepreneurial skills and attitudes, and business resource readiness. This structure is important because students may score high in one dimension and lower in another. For example, a student may know the basic requirements for starting a business but have limited access to people who can help with operations, funding, or market validation. A single readiness score may conceal that imbalance (Panakaje et al., 2025).

The literature increasingly emphasizes practice-oriented entrepreneurship education. Incubation, venture simulation, industry immersion, prototyping, mentorship, and market exposure are more likely to connect knowledge to behavior than lecture-based instruction alone (Hagg & Gabrielsson, 2019; Lackeus, 2020; Nabi et al., 2016). In emerging or provincial ecosystems, these mechanisms are not always equally available (Armas & Jose, 2024). Institutional support, policy awareness, and resource access therefore become central to readiness. However, awareness of support programs is not the same as the ability to use them. This distinction motivates the need for analytics that can separate cognitive readiness from operational readiness (Donbesuur et al., 2020).

1.1.2 Educational Data Mining and Readiness Analytics

Educational data mining and learning analytics convert student level records into patterns that can support prediction, segmentation, and intervention design (Romero & Ventura, 2020; Siemens & Long, 2011). In student success applications, these methods are often used to identify learners at risk, recommend support, or evaluate instructional pathways (Gibson, 2017). The same logic can be applied to entrepreneurship readiness. Instead of reporting only aggregate means, institutions can classify practical readiness, identify influential predictors, and create student profiles that suggest different intervention routes.

Supervised learning is useful when a target outcome can be defined. In this study, practical startup readiness was converted into a binary classification target using the overall practical readiness score. Logistic Regression provides a transparent baseline and allows comparison with familiar statistical modelling (Uddin et al., 2022). Decision Tree analysis offers interpretable branching rules and can capture nonlinear thresholds (Quinlan, 1986). Random Forest improves stability by aggregating multiple trees and provides relative feature importance scores (Breiman, 2001). Because the sample size is moderate, these models are appropriate as a compact benchmark set rather than as a high complexity predictive system.

Unsupervised learning addresses a different question. K-Means clustering can group students according to similarity across readiness dimensions (MacQueen, 1967). This is useful when the institution does not merely want to know whether a student is ready or not, but wants to understand what type of support the student may need. Silhouette scores can be used to assess how clearly the clusters are separated (Rousseeuw, 1987). In practice, interpretability matters alongside the score because a weakly separated but meaningful cluster structure may still support intervention planning.

Taken together, the literature shows that startup readiness is both educational and operational. It involves what students know, what they can do, what they are motivated to pursue, and what support they can access. On this basis, the present study treats startup readiness not only as a survey construct, but also as a decision support problem that can be examined through statistical testing, supervised classification, and student profiling.

2. MATERIALS AND METHODS

2.1 RESEARCH DESIGN, LOCALE, AND PARTICIPANTS:

Guided by this decision-support objective, the study used a descriptive-correlational quantitative design extended with machine learning classification and clustering. The dataset consisted of 385 students from selected public and private higher education institutions across the six congressional districts of Pangasinan, Philippines. Respondents represented different sex groups, age groups, academic programs, and year levels. Inclusion criteria required that respondents were currently enrolled students with exposure to entrepreneurship related subjects or activities.

The demographic structure of the sample is important for model interpretation. Male respondents comprised 56.1% of the sample while female respondents comprised 43.9%. The largest age group was 21 to 25 years old, followed by 18 to 20 and 26 to 30. BS Business Administration students formed the largest program group, with additional representation from BS Information Technology, Bachelor of Technology and Livelihood Education, and other programs. First year and fourth year students were strongly represented, while second year students formed the smallest year level group. Table 1 summarizes the main respondent profile variables used in the study.

Table 1: Respondent profile used in the analytics dataset

Variable	Category	Frequency	Percentage
Sex	Male	216	56.10
	Female	169	43.90
Age	18 to 20 years old	124	32.21
	21 to 25 years old	148	38.44
	26 to 30 years old	113	29.35

Variable	Category	Frequency	Percentage
Program	BS Business Administration	179	46.49
	BS Information Technology	88	22.86
	B Technology and Livelihood Education	78	20.26
	Other programs	40	10.39
Year Level	1st Year	124	32.21
	2nd Year	46	11.95
	3rd Year	102	26.49
	4th Year	113	29.35

2.2 INSTRUMENT AND CONSTRUCT SCORES:

Data were collected using a structured questionnaire validated through expert review and reliability testing. The instrument contained 28 Likert type items rated on a four-point scale. Five items measured knowledge of business startup, five measured career aspirations toward entrepreneurship, four measured policy awareness and perception, nine measured entrepreneurial skills and attitudes, and five measured business resource readiness. The four-point scale was interpreted as follows: 3.26 to 4.00, very high; 2.51 to 3.25, high; 1.76 to 2.50, moderate; and 1.00 to 1.75, low (Salleh et al., 2026).

Composite construct scores were computed as means of the corresponding items. 'Knowledge of business startup' was coded as KNO and computed from Items 1 to 5. 'Career aspirations' was coded as CAR and computed from Items 6 to 10. 'Policy awareness and perception' was coded as POL and computed from Items 11 to 14. 'Entrepreneurial skills and attitudes' was coded as SKL and computed from Items 15 to 23. 'Business resource readiness' was coded as RES and computed from Items 24 to 28. Overall practical readiness, coded as ORD, was computed as the mean of CAR, SKL, and RES. Table 2 shows the construct definitions used for statistical and machine learning analysis.

Table 2: Construct Definitions and Analytics Role

Code	Construct	Items	Role in Study
KNO	Knowledge of business startup	1 to 5	Cognitive predictor and clustering variable
CAR	Career aspirations toward entrepreneurship	6 to 10	Practical readiness component
POL	Policy awareness and perception	11 to 14	Cognitive and institutional predictor
SKL	Entrepreneurial skills and attitudes	15 to 23	Practical readiness component
RES	Business resource readiness	24 to 28	Practical readiness component
ORD	Overall practical readiness	CAR, SKL, RES	Classification target basis

The supervised classification target was derived from ORD. Students with ORD greater than or equal to 3.00 were coded as having strong practical startup readiness, while those below 3.00 were coded as having limited practical startup readiness. The threshold was researcher-defined and set above the lower bound of the high category to create a more conservative distinction between generally positive responses and stronger operational readiness.

2.3 STATISTICAL ANALYSIS:

Cronbach alpha was used to assess internal consistency of each multi-item scale (Santos, 1999). Spearman rank correlation was used because the source data were Likert type responses and the study examined monotonic associations among construct scores (Franzese & Iuliano, 2019; Sullivan & Artino, 2013). Multiple linear regression was used to test whether KNO and POL predicted CAR, SKL, RES, and ORD. Regression diagnostics included checks for multicollinearity using variance inflation factor values. The Friedman test was used as a nonparametric repeated measures comparison of the five related construct scores to evaluate the relative knowledge action gap (Hancock et al., 2018). Post hoc pairwise comparisons were conducted using Wilcoxon signed-rank tests with Bonferroni adjustment.

The statistical analysis addressed two hypotheses. H1 stated that knowledge and policy awareness would have positive associations with students' startup readiness. H2 stated that cognitive readiness would be significantly higher than practical startup readiness, indicating a relative gap between knowledge and action. This emphasizes relative difference rather than absolute deficiency.

2.4 MACHINE LEARNING WORKFLOW:

After the statistical procedures established the structure of the readiness gap, the machine learning stage translated the same dataset into decision-support outputs for screening and profiling. For the supervised models, the predictors were Program, Year Level, KNO, and POL. CAR, SKL, RES, and ORD were excluded from the predictor set because they were used to compute the target variable. This avoided data leakage and framed the readiness classification problem more conservatively. The models were implemented using standard machine learning procedures consistent with scikit-learn-based educational data mining workflows (Pedregosa et al., 2011). Categorical variables were encoded, numeric variables were standardized where required, and the dataset was divided into training and testing sets using a 70:30 stratified split. The classifiers were trained using standard scikit-learn implementation settings, and no hyperparameter tuning was performed because the purpose was to establish a compact baseline decision-support workflow.

Three supervised classifiers were trained: Logistic Regression, Decision Tree, and Random Forest. Performance was evaluated using accuracy, weighted precision, weighted recall, weighted F1 score, confusion matrix, and receiver operating characteristic area under the curve (Imani et al., 2026). Weighted metrics were used because the readiness classes were not assumed to be perfectly balanced (Khan et al., 2024). Random Forest feature importance was then examined to identify the relative contribution of POL, KNO, Program, and Year Level to readiness classification. These importance scores were interpreted as predictive contributions, not causal effects (Breiman, 2001; Pedregosa et al., 2011).

For student profiling, K-Means clustering was applied using the five construct scores: KNO, POL, CAR, SKL, and RES. The variables were standardized before clustering. Cluster solutions from two to five groups were compared using silhouette scores and substantive interpretability. A four-cluster solution was retained because it provided clearer intervention-oriented profiles even though the silhouette value indicated modest separation, and because clustering was used for decision-support interpretability rather than definitive typology (Rousseeuw, 1987). The clustering results were therefore interpreted as decision support categories rather than as rigid psychological types.

3. RESULTS

3.1 RELIABILITY AND CONSTRUCT LEVEL READINESS:

The five measurement scales showed acceptable to good internal consistency. KNO produced a Cronbach alpha of 0.744, which indicates acceptable reliability. CAR, POL, SKL, and RES produced alpha values of 0.848, 0.812, 0.848, and 0.803, respectively, indicating good reliability. Corrected item total correlations were generally above the accepted threshold, and no item removal was necessary on conceptual grounds. These results support the use of the construct scores for subsequent statistical and machine learning analysis.

The construct means showed a consistent pattern. KNO had the highest mean at 3.60 and was interpreted as very high. POL had a mean of 3.24 and was interpreted as high, close to the very high boundary. CAR, SKL, and RES were also interpreted as high, with means of 2.92, 3.03, and 2.91, respectively. ORD had a mean of 2.95. The pattern suggests that students were not weak in practical readiness, but they were stronger in cognitive and policy related dimensions than in action-oriented dimensions. Table 3 summarizes these results.

Table 3: Reliability and Construct Level Readiness

Construct	Items	Alpha	Mean	Interpretation
KNO	5	0.744	3.60	Very High
POL	4	0.812	3.24	High
CAR	5	0.848	2.92	High
SKL	9	0.848	3.03	High
RES	5	0.803	2.91	High
ORD	CAR, SKL, RES	NA	2.95	High

Several item-level patterns from this analysis clarify the construct results. Students reported strong understanding of basic startup requirements and the role of entrepreneurship education. Policy awareness was positive, but awareness of specific government youth entrepreneurship programs showed greater variability than other policy items. In the practical dimensions, students reported relatively stronger communication, teamwork, persistence, and risk-informed decision making, but lower confidence in setting clear business goals, access to operational assistance, seeking learning resources, and emotional preparedness for business stress. These details explain why the practical constructs remained high overall but lower than KNO.

3.2 CORRELATION, REGRESSION, AND FRIEDMAN RESULTS:

Spearman correlation results showed that KNO was not significantly associated with CAR, SKL, RES, or ORD. POL showed weak but statistically significant positive correlations with SKL, RES, and ORD, but not with CAR. These findings partially support H1: policy awareness had a positive association with selected readiness dimensions, while knowledge did not. Table 4 presents the correlation coefficients and p values.

Table 4: Spearman Correlation Between Cognitive Predictors and Readiness

Predictor	CAR	SKL	RES	ORD
KNO	-0.090 (0.079)	0.076 (0.135)	0.043 (0.400)	0.017 (0.736)
POL	-0.015 (0.775)	0.137 (0.007)	0.165 (0.001)	0.168 (<0.001)

Regression results were consistent with the correlation analysis. For ORD, the model was statistically significant but had limited explanatory power, with R2 equal to 0.034. POL was a significant positive predictor, while KNO was not. The same general pattern appeared for SKL and RES. The CAR model was not statistically significant, and neither KNO nor POL significantly predicted career aspirations. Variance inflation factor values were all 1.000, indicating no multicollinearity concern. The results show that cognitive and policy variables alone cannot explain practical startup readiness. Table 5 provides the compact regression summary.

The Friedman test showed significant differences among KNO, POL, CAR, SKL, and RES, with chi square of 362.018 and $p < 0.001$. Post hoc comparisons indicated that KNO was significantly higher than all other constructs and that POL was significantly higher than the practical readiness dimensions. Several practical readiness dimensions were closer to each other. This confirms H2 as a relative knowledge action gap: students are cognitively and policy ready to a greater degree than they are operationally ready.

Table 5: Multiple Linear Regression Summary

Outcome	Predictor	B	Beta	p	VIF
ORD	KNO	0.017	0.015	0.772	1.000
ORD	POL	0.130	0.183	<0.001	1.000
CAR	KNO	-0.150	-0.099	0.054	1.000

Outcome	Predictor	B	Beta	p	VIF
CAR	POL	-0.009	-0.009	0.858	1.000
SKL	KNO	0.100	0.058	0.253	1.000
SKL	POL	0.171	0.160	0.002	1.000
RES	KNO	0.101	0.048	0.336	1.000
RES	POL	0.229	0.179	<0.001	1.000

From a decision support perspective, this statistical layer is important because it identifies the structure of the problem. The issue is not a lack of entrepreneurship awareness. Rather, the evidence points to a translation problem: students know about entrepreneurship and generally recognize support mechanisms, but readiness depends on the less visible combination of skill confidence, resource access, aspiration, and ecosystem connection.

3.3 MACHINE LEARNING CLASSIFICATION RESULTS:

After the construct-level pattern was established, the supervised models examined whether non-leakage predictors could classify students into strong or limited practical startup readiness. The resulting target variable contained 173 (44.9%) students classified as strong practical startup readiness and 212 (55.1%) students classified as limited practical startup readiness. The supervised models produced moderate classification performance. Logistic Regression performed best overall, with accuracy of 0.66, weighted precision of 0.67, weighted recall of 0.66, weighted F1 score of 0.66, and receiver operating characteristic area under the curve of 0.729. The Decision Tree achieved 0.61 accuracy and 0.725 area under the curve, suggesting useful ranking ability despite weaker classification performance. Random Forest obtained 0.62 accuracy, 0.62 weighted F1 score, and 0.694 area under the curve. Table 6 summarizes the model comparison.

Table 6: Machine Learning Model Performance

Model	Accuracy	Weighted Precision	Weighted Recall	Weighted F1	Area Under Curve
Logistic Regression	0.66	0.67	0.66	0.66	0.729
Decision Tree	0.61	0.61	0.61	0.60	0.725
Random Forest	0.62	0.63	0.62	0.62	0.694

The goal of the model comparison was not to claim a high accuracy predictive system, but to test whether a compact institutional dataset could support readiness screening. The moderate scores are reasonable given that the supervised models used only program, year level, KNO, and POL as predictors. More direct predictors of readiness, such as self-efficacy, prior entrepreneurial experience, family business exposure, access to capital, and actual incubation participation, were not considered in the dataset. The result therefore supports a cautious but useful decision support role: the model can assist screening, but should not be used as the sole basis for student startup readiness.

Random Forest feature importance ranked POL first, KNO second, Program third, and Year Level fourth. This result is consistent with the statistical analysis in one respect: policy awareness had more predictive value than knowledge in explaining practical readiness. However, KNO remained important in the classification model, suggesting that it may contribute through nonlinear or interaction patterns that were not captured by linear regression. Table 7 provides the feature importance values.

One possible explanation is that a student's program and year level reflect distinct differences in curriculum exposure, maturity, disciplinary orientation, and proximity to graduation. Business students may encounter entrepreneurship content more directly, while technology and livelihood education students may bring different forms of practical or technical readiness. Year level may capture cumulative academic exposure and the urgency of post-graduation career planning. These interpretations require further validation, but they are useful for designing more differentiated entrepreneurship support.

Table 7: Random Forest Feature Importance

Rank	Predictor	Importance Score
1	POL	0.316
2	KNO	0.292
3	Program	0.206
4	Year Level	0.186

One possible explanation is that a student's program and year level reflect distinct differences in curriculum exposure, maturity, disciplinary orientation, and proximity to graduation. Business students may encounter entrepreneurship content more directly, while technology and livelihood education students may bring different forms of practical or technical readiness. Year level may capture cumulative academic exposure and the urgency of post-graduation career planning. These interpretations require further validation, but they are useful for designing more differentiated entrepreneurship support.

3.4 K-MEANS READINESS PROFILES:

While supervised classification addressed whether students were likely to show stronger practical readiness, clustering addressed a different institutional question: what types of readiness profiles exist among students, and what support routes do those profiles suggest? The clustering stage identified four readiness profiles. Silhouette scores were 0.181 for two clusters, 0.169 for three clusters, 0.161 for four clusters, and 0.168 for five clusters. The numerical separation was modest, but the four-cluster solution was retained because it produced the clearest intervention-oriented interpretation. Table 8 reports the cluster means and sizes.

Table 8: K-Means Readiness Profiles

Cluster	Size	KNO	POL	CAR	SKL	RES
0	97	3.77	2.89	2.96	3.29	3.31
1	94	3.17	3.15	3.05	2.96	2.69
2	86	3.77	2.97	2.80	2.54	2.33
3	108	3.69	3.83	2.86	3.23	3.19

Cluster 0 is interpreted as a practice-ready but policy under connected group. Students in this cluster had high KNO, SKL, and RES, but lower POL. They appear capable and resource ready, but may need clearer connection to government programs, institutional support mechanisms, and startup policy pathways. Cluster 1 is interpreted as an aspiration positive but resource constrained group. These students had the highest CAR among the clusters, but weaker RES and moderate SKL. They may benefit from mentoring, operational assistance, funding literacy, and access to incubation facilities.

Cluster 2 represents the clearest knowledge ready but action limited group. These students had high KNO but the lowest CAR, SKL, and RES. Their profile is central to the knowledge action gap. They may understand entrepreneurship but lack confidence, practical exposure, or perceived feasibility. Cluster 3 is interpreted as a policy-aware and support responsive group. These students had the highest POL and relatively strong SKL and RES, but their CAR was lower than Cluster 1. They may benefit from challenge-based learning, venture ideation, and opportunity recognition activities that convert support awareness into entrepreneurial commitment.

Table 9 translates the clusters into intervention options. This conversion shows that the machine learning output is not merely descriptive. It becomes a lightweight decision support map for HEI entrepreneurship offices, technology business incubators, and academic program heads.

Table 9: Cluster-Based Decision Support Interventions

Cluster	Profile Label	Primary Gap	Suggested Intervention
0	Practice-ready but policy under connected	Policy linkage	Policy orientation, startup grant briefing, referral to support programs
1	Aspiration positive but resource constrained	Resources and operations	Mentoring, resource mapping, incubation onboarding, finance bootcamp
2	Knowledge ready but action limited	Skills, aspiration, feasibility	Venture simulation, confidence building, ideation clinics, team formation
3	Policy-aware and support responsive	Action translation	Challenge based projects, prototype mentoring, market validation activities

4. DISCUSSION

Across the statistical and machine learning layers, the results point to a consistent pattern: students were generally aware of entrepreneurship and policy support, but practical readiness depended on less visible behavioral, resource, and institutional factors. This challenges the assumption that improving entrepreneurship content alone will automatically increase venture readiness (Armas & Jose, 2024; Carpenter & Wilson, 2022; Lyu et al., 2023; Widati et al., 2024). The more defensible interpretation is that knowledge is necessary but incomplete (Hagg & Gabrielsson, 2019; Lackeus, 2020; Nabi et al., 2016). It must be connected to repeated practice, feedback, mentoring, social support, and access to resources.

Policy awareness was more informative than knowledge, but its effect remained weak. This is useful because it clarifies the difference between awareness and access. Students may agree that government policies influence startup success or that support makes entrepreneurship easier, but they may not know how to access a program, prepare requirements, find mentors, or convert a business idea into a fundable proposal. In other words, policy awareness is a bridge, but it is not the destination (Armas & Jose, 2024; Quimba et al., 2024). Higher education institutions should therefore translate policy information into guided processes: orientation, eligibility mapping, application clinics, grant writing support, incubation referral, and networking with local ecosystem actors.

The machine learning results add a different layer of insight. Logistic Regression outperformed the tree-based models in overall performance, which suggests that the available predictors may have had relatively simple decision boundaries. This is not a weakness. For institutional use, a transparent baseline model with moderate predictive value can be more practical than a complex model that is difficult to explain (Pedregosa et al., 2011; Romero & Ventura, 2020; Siemens & Long, 2011). Decision Tree results remain valuable because they can generate rules, while Random Forest contributes feature importance rankings. Together, the models provide complementary forms of decision support: classification performance, interpretability, and predictor ranking.

The Random Forest feature importance results are also consistent with the broader narrative. POL and KNO were the two most influential predictors, but Program and Year Level were not negligible. This suggests that readiness classification is shaped by both cognitive institutional variables and academic context. It would be premature to infer causality from these scores, but they can guide practical questions for program improvement. For example, if program membership contributes to classification, institutions may examine whether entrepreneurship exposure differs across curricula. If year level matters, institutions may investigate whether readiness support should be staged across the academic lifecycle instead of concentrated in a single course.

The cluster solution provides the most direct intervention value. Averages alone might suggest that students are generally ready because all practical constructs were interpreted as high. Clustering shows a more refined picture. Some students appear practice-ready but insufficiently linked to policy support. Others are aspiration positive but resource constrained. Another group is knowledge ready but action limited, which is the most visible form of the knowledge action gap. A fourth group is highly policy-aware and may respond well to action translation activities. These distinctions allow institutions to avoid one size fits all entrepreneurship programming.

For an HEI technology business incubator, the workflow can be operationalized as a screening and advising tool. Students can complete the readiness instrument at the beginning of an entrepreneurship course, innovation challenge, or incubation pre-selection period. The system can compute construct scores, classify practical readiness, assign a readiness profile, and recommend an intervention track. Students in Cluster 2, for instance, may be routed to ideation and confidence building before incubation. Students in Cluster 0 may be fast tracked to policy linkage and program application support. Students in Cluster 1 may need mentoring and resource mapping, while Cluster 3 may benefit from market validation and prototype development.

The study covers readiness assessment up to applied educational analytics. The computational innovation practically lies in converting validated survey constructs into a classification and profiling workflow that is readable, implementable, and useful for institutional decision makers. The work demonstrates how machine learning and decision support can be applied to higher education entrepreneurship development.

4.1 IMPLICATIONS FOR SYSTEM DESIGN AND INSTITUTIONAL DEVELOPMENT:

Because the study is framed as decision support, the findings can be translated into a lightweight system design for entrepreneurship education, incubation preparation, and institutional planning. At the input layer, the system collects demographic data and questionnaire responses. At the processing layer, it computes construct scores, updates reliability checks when new cohorts are added, classifies practical startup readiness, estimates feature contributions, and assigns students to readiness clusters. At the output layer, it can present a dashboard showing each student's readiness class, cluster profile, priority support needs, and recommended intervention pathway; Table 10 shows the decision support outputs and the corresponding institutional users. For administrators, aggregate views may show readiness distribution by program, year level, district, or cohort. For entrepreneurship coordinators and incubation staff, individual or group-level views can guide mentoring, course activities, and pre-incubation preparation. This institutional use case is consistent with recent work emphasizing the role of higher education institutions and university ecosystems in supporting innovation, entrepreneurship, and startup development (Astuty et al., 2022; Quimba et al., 2024).

Table 10: Decision Support Outputs and Institutional Users

Output	Primary User	Decision Supported
Readiness class	Faculty and program heads	Identify students needing additional practical preparation
Feature importance	Curriculum planners	Identify variables that contribute to readiness classification
Cluster profile	Incubation and mentoring teams	Assign differentiated intervention pathways
Program level summary	Deans and administrators	Compare readiness patterns across academic programs
Year level summary	Course coordinators	Stage entrepreneurship support across the student lifecycle
Policy awareness summary	Agency partners and TBI staff	Plan briefings, grant clinics, and ecosystem orientation

The value of the system lies in its ability to translate readiness data into differentiated institutional action (Astuty et al., 2022; Gibson, 2017; Quimba et al., 2024; Romero & Ventura, 2020; Siemens & Long, 2011). At the course level, instructors may use the readiness profile at the beginning of an entrepreneurship subject to adjust learning activities. A class dominated by knowledge-ready but action-limited students may need simulations, team formation, opportunity discovery, and confidence-building tasks before moving to full business planning. A class with many resource-constrained students may require stronger exposure to mentors, suppliers, funding sources, local enterprises, and technology business incubator services. In incubation preparation, the model can serve as a pre-screening tool that identifies whether students need ideation support, market discovery, prototype coaching, policy navigation, or resource mapping. It does not replace interviews or expert judgment; rather, it provides a structured starting point and reduces the tendency to treat all students as if they had the same readiness problem.

The framework also supports program evaluation and ecosystem coordination (Astuty et al., 2022; Quimba et al., 2024). Because construct scores can be tracked by cohort, program, and year level, higher education institutions can determine whether curricular revisions, mentoring programs, or entrepreneurship initiatives are associated with improvements in practical readiness. For example, consistently high knowledge scores but low resource readiness would indicate that additional lectures are unlikely to solve the main constraint. A more appropriate response may involve mentorship, funding orientation, industry linkage, prototyping support, or access to incubation facilities. Beyond the university, local government units, regional agencies, industry partners, financial institutions, alumni entrepreneurs, and incubators can use aggregate readiness patterns to identify where support is most needed. Low policy awareness may call for targeted orientation from agencies. Weak resource readiness may require stronger industry and mentor engagement. Low skills readiness may require redesigning experiential learning activities.

In the intervention logic proposed in this study, cognitive support addresses what students know; behavioral support addresses what students can do; structural support addresses what students can access; and motivational support addresses whether entrepreneurship appears feasible and personally meaningful to them. A student with high knowledge but low skills and resource readiness does not require the same intervention as a student with high policy awareness and resource access but low career aspiration. The clustering results provide a basis for making these distinctions using patterns observed in the dataset.

For deployment, the framework should be embedded in an iterative improvement cycle. The first cycle establishes baseline readiness and assigns intervention tracks. The second records which interventions were delivered. The third measures readiness again and compares changes across clusters. The fourth links readiness changes to behavioral indicators such as participation in pitch events, prototype development, customer interviews, business registration, or incubation entry. This sequence converts the model from a one-time classifier into a learning system for entrepreneurship support.

Several safeguards are necessary. The model should remain advisory, not determinative (Romero & Ventura, 2020; Siemens & Long, 2011). Students should not be excluded from entrepreneurship opportunities solely because the system predicts limited readiness. Model outputs should be reviewed with human judgment, especially when students possess contextual strengths not captured by the questionnaire. The model should also be recalibrated as new cohorts are added, and institutions should monitor whether interventions actually improve readiness scores or startup-related outcomes. Data privacy must be protected throughout deployment. Identifiable student records should be accessible only to authorized personnel, while aggregate reporting is sufficient for most policy, curriculum, and ecosystem decisions.

A minimum viable implementation may use a spreadsheet, Python notebook, or simple web dashboard. A more mature version may be integrated into a student entrepreneurship portal, learning management system, or technology business incubator database. Readiness assessment should also be timed strategically. Administering the instrument only at the end of a program limits its usefulness because few opportunities remain for intervention. A better approach is to administer it at entry, before entrepreneurship course sequences, before incubation recruitment, and after major experiential activities. This allows institutions to observe readiness development over time rather than merely describe readiness at a single point.

Lastly, the framework can support equity-oriented planning. If future datasets show that certain groups consistently receive lower readiness classifications, the institution should examine whether the difference reflects genuine support needs, measurement bias, uneven access to entrepreneurship activities, or program-level opportunity gaps. Such analysis must be handled carefully. The system should expose barriers, not reproduce them (Mehrabi et al., 2021). Transparent reporting, human review, periodic validation, and responsible use are therefore essential to institutional deployment.

4.2 LIMITATIONS AND FUTURE WORK:

These implications should be interpreted within the methodological limits of the study. First, the data are cross sectional and self-reported; therefore, the findings cannot establish causality or prove that readiness scores lead to actual venture creation. Second, the sample was limited to selected institutions in Pangasinan, so external validation is needed before applying the model to other provinces or national student populations. Third, the classification target was derived from ORD, which is based on survey responses, not observed startup outcomes such as business registration, prototype development, sales, investment, or incubation completion. Fourth, the supervised models used a limited predictor set to avoid data leakage. This made the classification problem more conservative and methodologically cleaner, but it also restricted predictive power.

The clustering results should also be treated cautiously. The silhouette scores were low, indicating that the profiles were not strongly separated in geometric terms. The retained four cluster solution is useful because it supports intervention planning, but it should not be interpreted as a definitive typology of students. Future studies should test cluster stability across larger datasets and across institutions. They should also compare alternative clustering methods, such as hierarchical clustering, Gaussian mixture models, or density-based clustering, if larger and more varied data become available.

Future work should include longitudinal tracking to determine whether readiness profiles predict startup attempts, incubation participation, prototype completion, pitch competition performance, or venture survival. Additional predictors should also be collected, including entrepreneurial self-efficacy, risk tolerance, prior work experience, family business background, digital skills, access to capital, mentor availability, and participation in experiential entrepreneurship programs (Shukla & Kumar, 2024). Larger datasets would support more robust model validation, explainable artificial intelligence techniques, and fairness checks across demographic and academic groups.

4.3 RECOMMENDATIONS FOR HEI ENTREPRENEURSHIP SUPPORT:

Based on the observed knowledge-action gap, the moderate classification results, and the cluster-based profiles, three practical recommendations are proposed for HEI entrepreneurship support. First, entrepreneurship education should be connected to measurable readiness development. Courses should not end with conceptual assessment alone. They should require students to conduct customer discovery, prepare simple financial assumptions, test value propositions, and receive feedback from mentors or industry partners (Astuty et al., 2022; Hagg & Gabrielsson, 2019; Lackeus, 2020; Nabi et al., 2016). These activities directly target the skill and resource dimensions that were weaker than knowledge.

Second, policy awareness should be converted into guided access. HEIs may offer startup policy clinics, grant orientation sessions, eligibility mapping, and referral support so that students can move from knowing that support exists to actually using it (Armas & Jose, 2024; Astuty et al., 2022; Quimba et al., 2024).

Third, HEIs should use differentiated support rather than a uniform entrepreneurship intervention. The clusters suggest that some students need confidence and practice, some need policy linkage, some need resources, and others need translation from awareness to concrete action. An analytics-based support system can assign these routes early and update them after each major learning activity. In this way, the institution can use data not only to describe students, but also to organize the support environment around their specific readiness gaps.

5. CONCLUSIONS

This study developed a machine learning decision support framework for predicting and profiling startup readiness among higher education institution students in Pangasinan. The results confirm a relative knowledge action gap: students reported very high startup knowledge and high policy awareness, but practical readiness dimensions were lower. Statistical tests showed that knowledge did not significantly predict practical readiness, while policy awareness had weak but significant associations with skills, resource readiness, and overall readiness. The Friedman test confirmed significant differences across constructs, with cognitive dimensions higher than practical dimensions.

The machine learning layer extended these findings into a more operational form. Logistic Regression produced the best overall classification performance among the tested models, while Random Forest feature importance identified policy awareness and startup knowledge as the strongest predictors. K-Means clustering identified four student profiles that can guide differentiated interventions. These outputs convert the readiness survey into a decision support workflow for screening, profiling, and planning entrepreneurship support.

The study shows how a structured analytics workflow can help institutions see the gap more clearly and respond more deliberately; it does not claim that machine learning alone can solve the knowledge action gap. For provincial higher education institutions, this approach offers a practical way to connect entrepreneurship education, student readiness assessment, and ecosystem-based support. The result is a more targeted path from learning about entrepreneurship to preparing students to practice it.

Declarations

A. Data Availability

The dataset and analysis files are available from the corresponding author upon reasonable request, subject to restrictions necessary to protect respondent confidentiality. No personally identifiable respondent information is included in the reported results.

B. Ethical Considerations

Participation was voluntary, and informed consent was obtained prior to survey completion. Respondents were informed that their answers would be used for research purposes only. Data were stored securely and reported in aggregate form.

C. Declaration of Generative AI and AI-Assisted Technologies

Generative AI tools were used to support language refinement only. The study design, dataset, analysis, interpretation of results, and final scholarly claims remain the responsibility of the authors. The authors reviewed and edited the final manuscript and take full responsibility for its content.

D. Author Contribution

Conceptualization, R.J.C.A.-C. and R.R.F.; methodology, R.J.C.A.-C. and R.R.F.; validation, R.J.C.A.-C., G.M.D., and R.R.F.; formal analysis, R.R.F.; investigation and data collection, R.J.C.A.-C. and G.M.D.; data curation, R.J.C.A.-C. and R.R.F.; machine learning implementation and visualization, R.R.F.; writing—original draft preparation, R.J.C.A.-C. and R.R.F.; writing—review and editing, R.J.C.A.-C., G.M.D., and R.R.F.; project administration, R.J.C.A.-C.; supervision, R.J.C.A.-C. and R.R.F.

E. Conflict of Interest and Funding

The authors declare no conflict of interest. Funding was solely provided by Pangasinan State University.

References

1. World Bank Group. (2021). The World Bank annual report 2021: From crisis to green, resilient, and inclusive recovery. World Bank.
2. Ilas-Panganiban, D., & Mitra-Ventanilla, R. (2020). The new Philippine innovation laws: A response to the call of MSMEs and startups. GRUR International, 69(7), 693–705. https://doi.org/10.1093/grurint/ikaa082
3. Carpenter, A., & Wilson, R. (2022). A systematic review looking at the effect of entrepreneurship education on higher education student. The International Journal of Management Education, 20(2), Article 100541. https://doi.org/10.1016/j.ijme.2021.100541
4. Lyu, J., Shepherd, D., & Lee, K. (2023). The impact of entrepreneurship pedagogy on nascent student entrepreneurship: An entrepreneurial process perspective. Studies in Higher Education, 49(1), 62–83. https://doi.org/10.1080/03075079.2023.2220722
5. Widati, S. T., Muawiyah, S. N., & Suparno. (2024). Entrepreneurship education and self-efficacy: How do they affect the entrepreneurial readiness of vocational students? International Student Conference on Business, Education, Economics, Accounting, and Management, 2(1), 1733–1744. https://doi.org/10.21009/isc-beam.012.116
6. Armas, K. L., & Jose, K. R. Y. (2024). Assessing higher education institutions readiness for startup development in the Philippines: Policies, challenges, and recommendations. Journal of Lifestyle and SDGs Review, 5(1), Article e01788. https://doi.org/10.47172/2965-730x.sdgsreview.v5.n01.pe01788
7. Commission on Higher Education Regional Office 1. (2026b). List of authorized HEIs in Region 1. Retrieved June 10, 2026, from https://region1.ched.gov.ph/list-of-authorized-heis-in-region-1
8. Center for Pangasinan Studies. (2023). About Center for Pangasinan Studies. Retrieved June 10, 2026, from https://centerforpangasinanstudies.com/about-us
9. Commission on Higher Education Regional Office 1. (2026a). CHED RO1, DOT R1 conduct Tourism Start-Up Challenge 2025 provincial vetting. https://region1.ched.gov.ph/CHED-RO1-DOT-R1-cond
10. Ndofirepi, T. M. (2022). Entrepreneurship goal and implementation intentions formation: The role of higher education institutions and contexts. Journal of Innovation and Entrepreneurship, 11(1), Article 69. https://doi.org/10.1186/s13731-022-00263-y
11. Iwu, C. G., Maziriri, E. T., Yu, H., & Ayandibu, A. O. (2024). The dynamic interplay of societal values, entrepreneurial abilities and skill sets in shaping and propelling entrepreneurial activity among South African higher learning students. International Review of Management and Marketing, 14(4), 169–179. https://doi.org/10.32479/irmm.16273
12. Ng, S. I., Ismail, I. A., Rahman, R. A., & Thangiah, G. (2023). Investigating traits towards readiness in entrepreneurship among university students. In Advances in Economics, Business and Management Research (pp. 101–111). https://doi.org/10.2991/978-94-6463-162-3_10
13. Panakaje, N., Lasrado, F., Parvin, S., Maurya, A. K., Sheikh, S., Irfana, S., Madhura, M., & Bazigha, N. (2025). Assessing students readiness for startups from higher education: Scale development and validation. The International Journal of Management Education, 24(2), Article 101354. https://doi.org/10.1016/j.ijme.2025.101354

14. Otache, I., Edopkolor, J. E., Sani, I. M., & Umar, A. (2023). Entrepreneurship education and entrepreneurial intentions: Do entrepreneurial self-efficacy, alertness and opportunity recognition matter? The International Journal of Management Education, 22(1), Article 100917. https://doi.org/10.1016/j.ijme.2023.100917
15. Kim, M. S., Huruta, A. D., & Lee, S. H. (2022). Predictors of entrepreneurial intention among high school students in South Korea. Sustainability, 14(21), Article 14168. https://doi.org/10.3390/su142114168
16. Nabi, G., Linan, F., Fayolle, A., Krueger, N., & Walmsley, A. (2016). The impact of entrepreneurship education in higher education: A systematic review and research agenda. Academy of Management Learning & Education, 16(2), 277–299. https://doi.org/10.5465/amle.2015.0026
17. Hagg, G., & Gabrielsson, J. (2019). A systematic literature review of the evolution of pedagogy in entrepreneurial education research. International Journal of Entrepreneurial Behavior & Research, 26(5), 829–861. https://doi.org/10.1108/ijebr-04-2018-0272
18. Lackeus, M. (2020). Comparing the impact of three different experiential approaches to entrepreneurship in education. International Journal of Entrepreneurial Behavior & Research, 26(5), 937–971. https://doi.org/10.1108/ijebr-04-2018-0236
19. Donbesuur, F., Boso, N., & Hultman, M. (2020). The effect of entrepreneurial orientation on new venture performance: Contingency roles of entrepreneurial actions. Journal of Business Research, 118, 150–161. https://doi.org/10.1016/j.jbusres.2020.06.042
20. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), Article e1355. https://doi.org/10.1002/widm.1355
21. Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, 46(5), 30–40
22. Gibson, D. (2017). Big data in higher education: Research methods and analytics supporting the learning journey. Technology, Knowledge and Learning, 22(3), 237–241. https://doi.org/10.1007/s10758-017-9331-2
23. Uddin, S., Ong, S., & Lu, H. (2022). Machine learning in project analytics: A data-driven framework and case study. Scientific Reports, 12(1), Article 15252. https://doi.org/10.1038/s41598-022-19728-x
24. Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1, 81–106. https://doi.org/10.1007/BF00116251
25. Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324
26. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 281–297).
27. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
28. Salleh, K. M., Sulaiman, N., Gloeckner, G., Ramli, R., & Al-Asfour, A. (2026). Analyzing Likert scale type data and interpretation in TVET research: A guideline for the novice researcher. Architecture Image Studies, 7(1), 550–560. https://doi.org/10.62754/ais.v7i1.871
29. Santos, J. R. A. (1999). Cronbach alpha: A tool for assessing the reliability of scales. Journal of Extension, 37(2).
30. Franzese, M., & Iuliano, A. (2019). Correlation analysis. In Encyclopedia of Bioinformatics and Computational Biology (Vol. 1, pp. 706–721). Academic Press. https://doi.org/10.1016/B978-0-12-809633-8.20358-0
31. Sullivan, G. M., & Artino, A. R., Jr. (2013). Analyzing and interpreting data from Likert type scales. Journal of Graduate Medical Education, 5(4), 541–542. https://doi.org/10.4300/JGME-5-4-18
32. Hancock, G. R., Stapleton, L. M., & Mueller, R. O. (Eds.). (2018). The reviewer guide to quantitative methods in the social sciences (2nd ed.). Routledge.
33. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
34. Imani, M., Joudaki, M., Bagheri, A., Arabnia, H. R., & Oberski, C. (2026). Why ROC-AUC is misleading for highly imbalanced data: In-depth evaluation of MCC, F2-score, H-measure, and AUC-based metrics across diverse classifiers. Technologies, 14(1), Article 54. https://doi.org/10.3390/technologies14010054
35. Khan, A. A., Chaudhari, O., & Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. Expert Systems with Applications, 244, Article 122778. https://doi.org/10.1016/j.eswa.2023.122778
36. Quimba, F. M., Andrada, A., & Barral, M. A. (2024). HEIGHTS-IE: Higher education institutions generating holistic and transferable solutions innovation ecosystems, Philippines (Discussion paper). Philippine Institute for Development Studies. https://doi.org/10.62986/dp2024.25
37. Astuty, E., Yustian, O. R., & Ratnapuri, C. I. (2022). Building student entrepreneurship activities through the synergy of the university entrepreneurship ecosystem. Frontiers in Education, 7, Article 757012. https://doi.org/10.3389/feduc.2022.757012
38. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), Article 115. https://doi.org/10.1145/3457607
39. Shukla, S., & Kumar, R. (2024). Venturing into a new business: Do self-efficacy and risk-taking propensity help? Vikalpa: The Journal for Decision Makers, 49(1), 25–44. https://doi.org/10.1177/02560909241234226