A Comparative Study of Logistic Regression and Random Forest for Predicting Student Adaptability in Embedded Systems Entrepreneurship Education

Quba Siddique[1,*]

[1] Institute of Banking and Finance, Bahauddin Zakariya University Multan, Pakistan

ABSTRACT

The increasing demand for innovation in technology sectors necessitates a deeper understanding of the factors that foster adaptability among students in specialized fields like embedded systems entrepreneurship. This study provides a comparative performance analysis of two prominent machine learning algorithms, Logistic Regression and Random Forest, for predicting student adaptability levels (Low, Medium, and High). Utilizing a dataset comprising academic, experiential, and psychometric features, this research addresses the critical challenge of identifying student potential in a data-driven manner. The methodology employed a stratified data split and the Synthetic Minority Over-sampling Technique (SMOTE) to mitigate the severe class imbalance inherent in the dataset, followed by a rigorous hyperparameter tuning process for both models. The results revealed a nuanced outcome. While the linear Logistic Regression model achieved a superior overall accuracy (98.4%) compared to the more complex Random Forest model (87.2%), both algorithms completely failed to identify any instances of the High adaptability class. This critical failure underscores the limitations of standard classification techniques when faced with extremely rare positive instances. Furthermore, a feature importance analysis conducted with the Random Forest model indicated that practical skills, such as innovation and model deployment scores, were the most significant predictors of adaptability, whereas traditional academic metrics like GPA had negligible influence. This study concludes that while AI-driven models show significant promise as an early-warning system to identify students who may require additional support, they are currently unsuitable for talent identification due to data limitations. The findings strongly advocate for a pedagogical shift in technical entrepreneurship education, emphasizing the need to prioritize experiential learning and practical skill development over conventional academic measures to cultivate the next generation of adaptable innovators.

Keywords Adaptability, AI in Education, Embedded Systems, Predictive Modeling, Random Forest

Submitted 14 July 2025
Accepted 10 August 2025
Published 1 September 2025

Corresponding author
Quba Siddique,
qubassindhu@gmail.com

Additional Information and Declarations can be found on page 256

DOI: 10.63913/ail.v1i3.12

Introduction

The growing importance of entrepreneurship education in technology, particularly in embedded systems, cannot be understated. In an increasingly competitive and rapidly evolving technological landscape, the adaptability of students is crucial for innovation, technology development, and the success of entrepreneurial ventures. Within this context, a significant challenge arises: the ability to identify and nurture adaptability in students pursuing embedded systems entrepreneurship, as existing frameworks and tools often fall short in providing robust predictive capabilities that support proactive student

How to cite this article: Q. Siddique, "A Comparative Study of Logistic Regression and Random Forest for Predicting Student Adaptability in Embedded Systems Entrepreneurship Education," *Artif. Intell. Learn.*, vol. 1, no. 3, pp. 245-257, 2025.

engagement and development.

Recent literature highlights the application of artificial intelligence (AI) as a transformative tool in educational settings. AI-driven models can facilitate personalized learning experiences, enabling curricula that respond to individual student needs and potentials. Chan and Zary discuss how AI can enhance adaptive assessments, allowing the selection of subsequent questions based on previous student responses, which underscores the technologys capability to inform and optimize learning paths [1]. This technological leverage is particularly crucial in embedded systems education, where diverse learning styles and paces can create disparities in understanding complex concepts [2]. Additionally, predictive analytics can assist in developing early warning systems for academic performance, allowing educators to intervene proactively and guide students effectively [3].

The integration of AI in educational frameworks could support curriculum development specifically tailored for embedded systems and entrepreneurship. By harnessing AIs predictive capabilities, educators can create a responsive learning environment that accommodates different learning trajectories, leading to improved student outcomes [4]. For instance, Jiao et al demonstrate how AI models in online engineering education can accurately predict student performance, identifying areas in need of attention or improvement [3]. This predictive modeling is essential for fostering entrepreneurial readiness, allowing institutions to align educational outcomes with industry requirements.

Moreover, the challenges associated with operationalizing AI in education, particularly regarding data identification and bias in algorithms, present significant considerations that must be addressed. Fahimirad and Kotamjani explore the complexities of integrating AI within teaching and learning contexts, emphasizing the need for careful implementation to avoid exacerbating inequalities [5]. Therefore, a focused approach to faculty training and resource allocation becomes imperative to ensure that both educators and students benefit from these technological advancements.

The potential of AI-driven predictive modeling to enhance student adaptability in embedded systems entrepreneurship presents a compelling avenue for research and curriculum reform. By integrating AI technologies into educational practices, institutions can better prepare students for the entrepreneurial challenges they will face in their professional careers. The successful deployment of these models not only stands to benefit individual student learning experiences but also contributes to scalable improvements in educational effectiveness across the board.

The primary objective of this study is to evaluate and compare the performance of two supervised machine learning models, Logistic Regression and Random Forest, in predicting student adaptability levels within the context of embedded systems entrepreneurship education. This research seeks to determine which model offers a more robust framework for identifying students who may require additional support to cultivate their adaptive capabilities. To achieve this, the study will first investigate how accurately a Logistic Regression model can predict adaptability levels, followed by a similar evaluation of a Random Forest model. Subsequently, the research will conduct a comparative analysis of the two models based on their predictive performance, specifically concerning accuracy, precision, recall, F1-score, and Area Under the Curve (AUC) metrics.

This study is specifically delimited to the analysis of the Students Industrial Engagement and Readiness Dataset (SIERD), and its findings are therefore contingent on the variables and population represented therein. The research focuses exclusively on Logistic Regression and Random Forest, with other machine learning algorithms falling outside the scope of this investigation, and the definition of "adaptability" is based on the constructs provided within the dataset.

Literature Review

Student Adaptability in Education

Student adaptability is increasingly recognized as a crucial competency in educational settings, particularly within the realms of STEM (Science, Technology, Engineering, and Mathematics) and entrepreneurship. Conceptually, adaptability encompasses cognitive, behavioral, and affective dimensions that influence how students respond to changing environments and challenges. According to Dewani and Nuzulia, students exposed to structured career education demonstrate higher career adaptability, suggesting that institutional support can enhance students ability to navigate fluctuating academic and professional landscapes [6]. Influencing factors such as social support, emotional well-being, and prior experiences significantly shape adaptability, impacting students capacity to engage and excel in demanding fields [7].

The importance of adaptability extends to entrepreneurship education, where students must integrate theoretical knowledge with practical applications, particularly in complex fields like embedded systems entrepreneurship. Embedded systems require a convergence of hardware and software knowledge and the ability to prototype rapidly and validate market needs [8]. This presents unique challenges that necessitate a strong foundation in adaptability, enabling students to maneuver through uncertainties effectively [9]. By fostering project-based learning environments and encouraging industry engagement, educational frameworks can enhance the technical skills and entrepreneurial capabilities of students within these domains, such as skills related to Raspberry Pi and deep learning model deployment [10].

Entrepreneurship Education in Technical Domains

Entrepreneurship education in technical fields like embedded systems comprises various pedagogical approaches that prioritize hands-on learning and interdisciplinary collaboration. Teaching strategies must integrate not just technical competencies but also nurture soft skills, which are increasingly acknowledged as critical for entrepreneurial success. Kormakova et al emphasize that the STEM approach, which focuses on real-world problem-solving, can effectively enhance students critical thinking and creativity [11], which are vital in entrepreneurial ventures [12]. Furthermore, Wu et al explore how teacher readiness significantly influences the effectiveness of STEM implementation, positing that teacher self-efficacy can catalyze student engagement and interest in STEM careers [13].

However, embedded systems entrepreneurship often faces specific challenges such as effective hardware-software integration, market validation, and rapid prototyping. According to Rakićević et al, providing students with robust entrepreneurial education can significantly impact their preparedness to tackle these challenges, emphasizing the interplay between technical knowledge and entrepreneurial mindset [8]. Engaging students with real-world projects not only solidifies their technical skills but also enhances their ability to innovate within rapidly changing environments.

Machine Learning in Educational Data Mining (EDM) and AI in Learning

Machine learning (ML) applications in education are increasingly being recognized as powerful tools for enhancing educational outcomes, particularly in the areas of predicting student performance, dropout rates, and learning styles. These applications allow educational institutions to harness the potential of data-driven insights to improve student success. By integrating ML, educators can analyze vast datasets to identify patterns and trends that can inform intervention strategies, create tailored learning experiences, and ultimately enhance the overall educational experience for students [14]. For instance, prior research has highlighted the effectiveness of methods such as Logistic Regression and Random Forest in predicting academic outcomes across various educational contexts [14][15]. These methods have been shown to provide valuable insights that can guide educational institutions in making informed decisions to support student learning and success.

As more institutions adopt these technologies, they can leverage the insights gained from predictive modeling to provide timely support that is tailored to the specific needs of each student. This personalized approach not only enhances adaptability but also improves performance in higher education settings [16]. By utilizing ML, educators can better understand the unique needs and challenges faced by each student, allowing them to provide more effective and targeted support. However, despite these advancements, there are still challenges that need to be addressed. One significant challenge is the accurate quantification of soft skills and behavioral traits, such as adaptability and innovation, which are crucial in entrepreneurship. These skills are often difficult to measure and quantify, making it challenging to incorporate them into predictive models.

Existing studies do not sufficiently address these complexities, as highlighted in the gap analysis conducted by Chen et al [17]. This analysis points to the need for more comparative research that utilizes robust datasets to predict adaptability in embedded systems entrepreneurship. By conducting such research, educators and policymakers can gain a better understanding of how to foster these essential skills in students, ultimately preparing them for success in both academic and entrepreneurial endeavors. In conclusion, while ML offers significant potential for improving educational outcomes, there is still a need for further research and development to fully realize its benefits, particularly in the area of predicting and enhancing adaptability and innovation.

Predictive Modeling for Soft Skills and Behavioral Traits

The endeavor to quantify and predict soft skills like adaptability presents significant challenges in educational research. While methodologies exist that have successfully utilized ML for quantifying various academic traits, the specific contextual application to traits such as adaptability remains underexplored [18]. Identifying suitable dataset features that encapsulate the multifaceted nature of adaptability is critical for developing effective predictive models [17]. Therefore, developing robust predictive techniques for soft skills is essential for enhancing students readiness for entrepreneurial roles, especially in technical domains like embedded systems.

In sum, the literature indicates that fostering student adaptability in tech-centric entrepreneurship education requires a multifaceted approach. This includes establishing supportive educational frameworks, deploying effective pedagogical strategies, and integrating advanced predictive technologies. The existing gaps in predictive modeling for psychological traits such as adaptability underline the need for further research, suggesting that future studies could provide critical insights into enhancing student preparedness for the complexities of modern entrepreneurial landscapes.

Method

This section provides a detailed exposition of the research methodology, encompassing the dataset characteristics, the multi-stage data preprocessing pipeline, the architecture and theoretical underpinnings of the selected machine learning models, and the rigorous framework for hyperparameter optimization and performance evaluation. The entire experimental workflow was programmatically executed in a Python environment, utilizing high-level libraries including Scikit-learn for modeling, Pandas for data manipulation, and Imbalanced-learn for handling class distribution anomalies.

Dataset and Initial Preparation

The foundation of this research is Student Entrepreneurship dataset, a specialized collection of student-centric attributes designed to capture the multifaceted nature of entrepreneurial potential in a technical domain. The primary analytical goal is the prediction of the Adaptability_Label, a multi-class ordinal target variable with three logically sequenced levels: Low (encoded as 0), Medium (1), and High (2). While the models in this study treat the variable as categorical, its ordinal nature implies an inherent order, a characteristic that adds valuable context to the interpretation of classification errors. The feature space is heterogeneous, comprising a mix of academic performance indicators (Academic_GPA), experiential metrics (Project_Count, Industry_Collaboration), and psychometric constructs (Innovation_Score, Self_Efficacy_Score).

The initial data preparation phase involved the programmatic exclusion of the Student_ID feature. This column, being a high-cardinality, arbitrary identifier, possesses no intrinsic or generalizable predictive information. Its inclusion would introduce severe methodological flaws; a model could simply memorize the outcome for each unique ID, leading to near-perfect performance on the training data. This phenomenon, a form of extreme overfitting, results from the model learning spurious correlations that are not present in the underlying population, rendering it completely ineffective at making predictions on new, unseen data. Removing such identifier columns is a fundamental and non-negotiable step in building a valid and generalizable predictive model.

Data Preprocessing and Feature Engineering

Following initial preparation, the preprocessed dataset was partitioned into a training set and a hold-out test set using a 75/25 ratio. This is a conventional and empirically validated split that balances the competing needs of providing the models with a substantial majority of the data for robust parameter learning, while reserving a sufficiently large, independent set for an unbiased final evaluation of the models generalization capabilities. To preserve the a priori class probabilities of the Adaptability_Label in both subsets, a stratified sampling strategy was implemented. This technique ensures that the relative proportions of the low, medium, and high adaptability classes are maintained across both the training and testing partitions, a critical step that prevents sampling bias and guarantees that the test set provides a faithful representation of the original datasets distribution. For the purpose of experimental reproducibility, the pseudo-random number generator was seeded with a constant value of 42, ensuring that the exact same data split can be recreated in future analyses.

A key challenge identified during exploratory data analysis was a notable class imbalance within the training data, a common issue in real-world datasets where certain outcomes are naturally rarer than others. To rectify this and prevent the learning algorithms from developing a predictive bias towards the majority class, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. SMOTE operates in the feature space by creating synthetic instances of the minority classes. For each minority class sample, it identifies its k-nearest minority-class neighbors and generates new samples along the line segments joining the sample and its chosen neighbors. This procedure, unlike simple random over-sampling, creates more varied and robust decision boundaries. Crucially, this over-sampling was applied only to the training partition after the train-test split to avoid data leakage, a critical methodological error where information from the test set contaminates the training process, leading to artificially inflated and invalid performance estimates.

Subsequent to resampling, all predictor variables underwent feature scaling via the StandardScaler. This transformation standardizes each feature by applying the z-score normalization formula, $\frac{(x-\mu)}{\sigma}$, where μ is the mean and σ is the standard deviation of the feature in the training data. This process results in a distribution with a mean of zero and a standard deviation of one. This step is essential for algorithms like Logistic Regression, whose gradient-based optimization process can converge much faster and more reliably on standardized features. It also prevents features with larger magnitudes and variances from disproportionately influencing the models parameter optimization. The scaler was fitted exclusively on the resampled training data, and the identical, learned transformation (using the same μ and σ values) was subsequently applied to the test data to ensure methodological consistency.

Machine Learning Models and Hyperparameter Tuning

Two distinct, yet powerful, classification algorithms were selected for this comparative analysis: Multinomial Logistic Regression and the Random Forest Classifier, representing linear and non-linear modeling paradigms, respectively.

Logistic Regression was implemented as a robust, interpretable linear baseline. Given the three-level target variable, a multinomial (or Softmax) configuration was used. This approach generalizes binary logistic regression by employing the Softmax function to calculate a vector of probabilities, one for each class, which collectively sum to one. The model was optimized using the lbfgs solver, an efficient quasi-Newton method well-suited for this problem. To mitigate overfitting and improve generalization, L2 regularization (Ridge) was incorporated. This technique adds a penalty term to the cost function that is proportional to the square of the magnitude of the models coefficients, effectively discouraging overly complex models with large coefficients that might be fitting to noise in the training data.

The Random Forest Classifier was selected as a more complex, non-linear model renowned for its high accuracy and robustness. It is an ensemble learning method that operates by constructing a large number of individual decision trees at training time. Its predictive power stems from two key principles that combat the high variance and overfitting tendencies of single decision trees: bootstrap aggregating (bagging), where each tree is trained on a different random subsample of the data drawn with replacement, and feature randomness, where each split in a tree is determined from a random subset of the total features. This dual-randomization strategy effectively decorrelates the individual trees. The final prediction is made by aggregating the votes from all trees in the forest (majority vote), which significantly reduces the variance of the final model compared to its individual components.

To determine the optimal architecture for each model, an exhaustive Grid Search with 5-fold Stratified Cross-Validation was performed. This procedure systematically trains and evaluates a model for every combination of hyperparameters specified in a predefined grid. The training data is split into five "folds," and the process iterates five times. In each iteration, one fold is held out as a validation set, while the model is trained on the remaining four. For Logistic Regression, the grid search focused on the inverse regularization strength parameter C. For Random Forest, the search space was more extensive, exploring combinations of n_estimators, max_depth, min_samples_split, min_samples_leaf, and class_weight. The guiding metric for selecting the superior hyperparameter set was the weighted F1-score, which is particularly well-suited for imbalanced datasets as it computes the F1-score for each class and combines them using a weight proportional to the number of true instances for each class, providing a more balanced performance measure than raw accuracy.

Evaluation Metrics

The predictive efficacy of the final, optimized models was rigorously quantified on the unseen test set using a suite of standard metrics designed to provide a comprehensive and multi-faceted view of performance. The confusion matrix was generated as the foundational tool for this analysis, providing a granular view of classification performance by tabulating the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each of the three adaptability classes.

From this matrix, several key metrics were derived. Overall accuracy, the ratio of all correct predictions to the total number of instances, served as a general performance indicator. However, to gain a more nuanced understanding, precision $\left(\frac{TP}{TP+FP}\right)$, recall $(\frac{TP}{TP+FN})$, and the F1-score (the harmonic mean of precision and recall) were calculated. These metrics were computed for each class individually, providing insight into the models ability to correctly classify Low, Medium, and High adaptability, and as a weighted average, which accounts for the class distribution in the test set to provide a holistic assessment.

Finally, to evaluate the models discriminative ability across all possible classification thresholds, the Area Under the Receiver Operating Characteristic (AUC-ROC) score was computed. Since ROC curves are inherently binary, this metric was adapted for the multiclass context using two standard strategies: One-vs-Rest (OVR), which computes the AUC for each class against all others, and One-vs-One (OVO), which computes the AUC for every pair of classes. Calculating and reporting both provides a more complete and robust picture of the models ability to distinguish between the different levels of student adaptability.

Result and Discussion

Dataset Overview and Preprocessing Outcome

The initial dataset consisted of 500 unique student records, each described by 10 predictive features and one target variable, Adaptability_Label. Exploratory Data Analysis immediately revealed a severe class imbalance in this target variable, a critical challenge that can significantly bias model training. The Medium adaptability class (Label 1) constituted the vast majority of samples (75.2%), followed by the Low class (Label 0) at a much smaller 24.0%. Most critically, the High adaptability class (Label 2) was extremely rare, representing only 0.8% of the entire dataset, which translates to a mere four instances. This distribution poses a significant risk that a standard classifier might achieve high accuracy simply by defaulting to the majority class and ignoring the minority classes entirely. Following a 75/25 stratified split to preserve this distribution in the test set, the SMOTE procedure was applied to the training data. This process successfully rebalanced the training set by generating synthetic instances of the minority classes (Low and High), resulting in a new, larger training set of 846 instances where all three classes had an equal 33.3% representation, providing a theoretically unbiased dataset for model development.

Logistic Regression Model Performance

The optimized Multinomial Logistic Regression model, configured with a strong regularization parameter (C of 100), demonstrated exceptionally high performance on the test set in aggregate terms. It achieved an overall accuracy of 98.40% and a weighted F1-score of 0.9800. The F1-score, being the harmonic mean of precision and recall, indicates that the model was highly effective at both correctly identifying instances of the majority classes and avoiding false alarms. The models discriminative ability was also robust, reflected in a weighted AUC-ROC score of 0.9926 (OVR), suggesting it can reliably distinguish between classes across various thresholds.

However, a detailed look at the per-class metrics and the confusion matrix reveals a more nuanced and critical picture. The model performed almost perfectly for the Low (F1-score: 0.9831) and Medium (F1-score: 0.9895) adaptability classes, correctly classifying 29 out of 30 Low instances and all 94 Medium instances. Despite this, it completely failed to identify the single instance of the High adaptability class present in the test set. The confusion matrix shows this instance was misclassified as Medium. This total failure resulted in precision, recall, and F1-scores of 0.0000 for Class 2, indicating that, despite its high overall accuracy, the model has no predictive power for the most desirable student outcome.

Random Forest Model Performance

The tuned Random Forest Classifier, configured with 100 estimators and a max_depth of 10, yielded a considerably lower overall performance compared to the Logistic Regression model. It achieved a more modest accuracy of 87.20% and a weighted F1-score of 0.8662. The weighted AUC-ROC score was also lower at 0.9202 (OVR). The confusion matrix for the Random Forest shows more classification errors between the two main classes than the Logistic Regression model, misclassifying 9 Low adaptability students as Medium and 6 Medium students as Low.

The Random Forest performed reasonably well on the majority and primary minority classes, achieving an F1-score of 0.7368 for Low adaptability and 0.9167 for Medium adaptability. However, mirroring the Logistic Regression models critical weakness, it also failed entirely to predict the High adaptability class. The single true instance of Class 2 was again misclassified as Medium, leading to identical null scores (0.0000) for precision, recall, and F1-score for this class. This result underscores that the complexity of the Random Forest model offered no advantage in overcoming the core challenge presented by the extreme minority class.

Feature Importance Analysis

The Random Forest model, by virtue of its tree-based structure, provided valuable insights into the relative importance of the predictor variables in its decision-making process. The analysis revealed a clear hierarchy of influence. Innovation_Score (20.3%) emerged as the most influential feature, followed closely by DL_Model_Deployment_Score (20.0%) and Industry_Collaboration (15.5%). These three features, which represent a blend of mindset, advanced practical skills, and real-world experience, collectively accounted for over 55% of the models predictive power. In stark contrast, traditional academic metrics held significantly less sway. Most notably, Academic_GPA (2.8%) was found to be the least important feature, suggesting that a students grades have a negligible relationship with their predicted adaptability in this context.

Interpretation of Model Performance

The primary objective was to compare Logistic Regression and Random Forest for predicting student adaptability. The results present a compelling, albeit complex, conclusion: the simpler, linear Logistic Regression model significantly outperformed the more complex, non-linear Random Forest model on nearly all aggregate metrics. This counterintuitive finding strongly suggests that the underlying relationships between the features and the Low and Medium adaptability classes are predominantly linear. The additional complexity of the Random Forest, designed to capture intricate, non-linear patterns, appears to have been detrimental, likely leading to a degree of overfitting on the nuances of the training data (including the synthetic SMOTE samples) that did not generalize well to the unseen test set.

However, the most critical and revealing finding is the uniform failure of both models to predict the High Adaptability class. This is a direct and unambiguous consequence of the extreme class imbalance in the original dataset. Despite the application of SMOTE to balance the training data, the synthetic samples generated for Class 2 were likely derived from a very small and homogenous set of just three initial training instances. This lack of diversity meant the models could not learn a robust, generalizable pattern for this class. Instead, their optimization algorithms, driven by the goal of maximizing overall accuracy, found it mathematically optimal to effectively ignore the High adaptability category. This is a classic pitfall in imbalanced classification problems. This outcome directly answers the research questions by demonstrating that while both models can reliably distinguish between Low and Medium adaptability, neither is currently a viable tool for identifying students with high adaptability.

Insights from Feature Importance

The feature importance analysis provides actionable pedagogical insights that challenge traditional educational paradigms. The prominence of Innovation_Score, DL_Model_Deployment_Score, and Industry_Collaboration strongly suggests that adaptability in the demanding context of embedded systems entrepreneurship is more closely tied to a students innovative mindset, their proficiency in advanced, practical technical skills, and their engagement with real-world industry challenges than to their academic record.

The fact that Academic_GPA was the least important predictor is a particularly disruptive finding. It challenges the long-held institutional wisdom of relying heavily on grades as the primary indicator of a students potential for success in dynamic, applied fields. This implies that educational programs aiming to foster adaptability should strategically shift their focus. Curricula could be redesigned to prioritize hands-on, project-based learning, creative problem-solving through hackathons or design challenges, and mandatory industry partnerships or internships. These activities directly cultivate the skills and experiences that the model found to be most predictive, offering a clear roadmap for curriculum reform.

Theoretical and Practical Implications

Theoretically, this study contributes to the educational data mining field by serving as a potent case study on the limitations of standard classification algorithms and corrective techniques like SMOTE when faced with extreme minority classes. It demonstrates that while over-sampling can balance a dataset numerically, it cannot create new, meaningful information if the initial variance in the minority class is insufficient.

Practically, the implications for the field of educational AI are twofold and carry significant ethical weight. On one hand, the high accuracy of the Logistic Regression model in distinguishing between Low and Medium adaptability students presents a valuable tool for proactive student support. It could be deployed as an early-warning system to flag students who may require additional support, mentorship, or targeted interventions to improve their adaptability. On the other hand, the models complete inability to identify high-potential students means they are dangerously unsuited for talent identification for specialized programs, scholarships, or startup incubators. Relying on these models for such a purpose would not only be ineffective but would create a system that systematically excludes the very students it is designed to find, potentially reinforcing existing biases and overlooking unconventional talent.

Limitations and Future Work

The primary and most significant limitation of this study is the severe underrepresentation of the High Adaptability class within the dataset. The findings are therefore constrained by the synthetic nature of the data, and future work must prioritize the collection of larger, real-world student datasets with a more balanced class distribution to validate these models. Further research should also explore more advanced imbalance-handling techniques. For instance, cost-sensitive learning, which assigns a much higher misclassification penalty to the minority class during training, could force the model to pay more attention to it. Alternatively, framing the problem as an anomaly detection task, where High Adaptability students are treated as rare and desirable outliers, might yield better results. Finally, a longitudinal study that tracks student outcomes (e.g., startup creation, career progression) over several years would be invaluable in confirming whether the features identified here are truly predictive of long-term entrepreneurial success and adaptability.

Conclusion

This study conducted a comparative analysis of Logistic Regression and Random Forest models to predict student adaptability in embedded systems entrepreneurship education. The findings revealed that while the simpler Logistic Regression model demonstrated superior accuracy in distinguishing between students with Low and Medium levels of adaptability, both it and the more complex Random Forest model completely failed to identify the High adaptability class due to its severe underrepresentation in the dataset. Furthermore, the research highlighted that practical skills, innovative mindset, and industry engagement were substantially more influential predictors of adaptability than traditional academic metrics like GPA. This underscores a critical disconnect between conventional measures of student success and the attributes required for entrepreneurial readiness in technical fields. Ultimately, this research contributes a dual-sided perspective to the application of AI in learning. It presents a functional, albeit limited, predictive tool that can serve as an early-warning system for educators to support students who may struggle with adaptability, while simultaneously offering a stark, data-driven caution against using such models for talent identification without sufficient and balanced data. The path forward involves not only the collection of more robust, real-world datasets and the exploration of advanced modeling techniques but also a pedagogical shift. By embracing the insights from the feature importance analysis, educational institutions can better align their curricula with the demands of the modern tech landscape, fostering the adaptable, innovative, and resilient entrepreneurs of the future through a greater emphasis on experiential, project-based learning.

Declarations

Author Contributions

Conceptualization: Q.S.; Methodology: Q.S.; Software: Q.S.; Validation: Q.S.; Formal Analysis: Q.S.; Investigation: Q.S.; Resources: Q.S.; Data Curation: Q.S.; Writing Original Draft Preparation: Q.S.; Writing Review and Editing: Q.S.; Visualization: Q.S.; The author have read and agreed to the published version of the manuscript.

Data Availability Statement

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

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The author 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] K. S. Chan and N. Zary, "Applications and Challenges of Implementing Artificial Intelligence in Medical Education: Integrative Review," *Jmir Med. Educ.*, 2019, doi: 10.2196/13930.

[2] M. Roshanaei, H. Olivares, and R. R. Lopez, "Harnessing AI to Foster Equity in Education: Opportunities, Challenges, and Emerging Strategies," *J. Intell. Learn. Syst. Appl.*, 2023, doi: 10.4236/jilsa.2023.154009.

[3] P. Jiao, F. Ouyang, Q. Zhang, and A. H. Alavi, "Artificial Intelligence-Enabled Prediction Model of Student Academic Performance in Online Engineering Education," *Artif. Intell. Rev.*, 2022, doi: 10.1007/s10462-022-10155-y.

[4] Prof. Dr. Nirvikar Katiyar *et al.*, "Ai-Driven Personalized Learning Systems: Enhancing Educational Effectiveness," *Eatp*, 2024, doi: 10.53555/kuey.v30i5.4961.

[5] M. Fahimirad and S. S. Kotamjani, "A Review on Application of Artificial Intelligence in Teaching and Learning in Educational Contexts," *Int. J. Learn. Dev.*, 2018, doi: 10.5296/ijld.v8i4.14057.

[6] Y. R. Dewani and S. Nuzulia, "Does Social Support Contribute to Career Adaptability? Study at the Final Year Student," *J. Indones. Sos. Teknol.*, 2024, doi: 10.59141/jist.v5i3.946.

[7] O. Starynska, Л. Співак, A. Osmanova, and O. Revutska, "Social Intelligence as a Factor of Socio-Psychological Adaptation of University Students With Special Educational Needs During Distance Learning Due to the COVID-19," *Rev. Romaneasca Pentru Educ. Multidimens.*, 2023, doi: 10.18662/rrem/15.1/705.

[8] Z. Rakićević, J. Rakičević, J. A. Labrović, and B. Ljamić-Ivanović, "How Entrepreneurial Education and Environment Affect Entrepreneurial Readiness of STEM and Business Students? A Longitudinal Study," *Eng. Econ.*, 2022, doi: 10.5755/j01.ee.33.4.30244.

[9] C. Lopez and S. Jones, "Examination of Factors That Predict Academic Adjustment and Success of Community College Transfer Students in STEM at 4-Year Institutions," *Community Coll. J. Res. Pract.*, 2016, doi: 10.1080/10668926.2016.1168328.

[10] T. R. Kelley and J. G. Knowles, "A Conceptual Framework for Integrated STEM Education," *Int. J. Stem Educ.*, 2016, doi: 10.1186/s40594-016-0046-z.

[11] V. N. Kormakova, S. D. Chernyavskikh, O. N. Satler, and L. N. Trikula, "Digitalization in STEM Education: Experience of Empirical Research," *Res. Result Pedagogy Psychol. Educ.*, 2023, doi: 10.18413/2313-8971-2023-9-1-0-01.

[12] M. N. Asilevi, S. Havu-Nuutinen, and J. Kang, "Secondary School Teachers Interest and Self-Efficacy in Implementing STEM Education in the Science Curriculum," *Eur. J. Sci. Math. Educ.*, 2024, doi: 10.30935/scimath/14383.

[13] P. Wu *et al.*, "How K12 Teachers Readiness Influences Their Intention to Implement STEM Education: Exploratory Study Based on Decomposed Theory of Planned Behavior," *Appl. Sci.*, 2022, doi: 10.3390/app122311989.

[14] Y. Li, "Analysis and Prediction of Students Adaptation to Online Education Systems Based on Data Analysis and Decision Tree Machine Learning Algorithms," *Adv. Soc. Behav. Res.*, 2024, doi: 10.54254/2753-7102/7/2024053.

[15] Z. Wang, J. Wang, and M. Jia, "Case Study on the Development and

Implementation of STEM Projects: DIY Traffic Lights," *Res. Educ. Assess. Learn.*, 2022, doi: 10.37906//real.2022.4.

[16] L. T. Bich Le, T. T. Tran, and N. H. Trần, "Challenges to STEM Education in Vietnamese High School Contexts," *Heliyon*, 2021, doi: 10.1016/j.heliyon.2021.e08649.

[17] G. Ping-qian, "Research on the Strategy of Improving College Students Career Adaptability in Application-Oriented Colleges and Universities," *J. Educ. Teach. Soc. Stud.*, 2021, doi: 10.22158/jetss.v3n3p54.

[18] M. Teychenne, K. Parker, D. Teychenne, S. Sahlqvist, S. Macfarlane, and S. A. Costigan, "A Pre-Post Evaluation of an Online Career Planning Module on University Students Career Adaptability," *J. Teach. Learn. Grad. Employab.*, 2019, doi: 10.21153/jtlge2019vol10no1art781.