

Article

Explainable Artificial Intelligence (XAI) for Identifying the Integration of International Students in the Host Country and Its Culture

James Vakilian ^{1,*}, Fareed Ud Din ¹, Edmund J. Sadgrove ¹, Mohammadreza Haghighat ² and Niusha Shafiabady ^{3,4}

¹ School of Science and Technology, University of New England, Armidale, NSW, 2350, Australia; fuddin@une.edu.au (F.U.D.); esadgro2@une.edu.au (E.J.S.)

² Independent Researcher, Townsville, QLD, 4814, Australia; mr.haghighat1998@gmail.com

³ Women in AI for Social Good Lab, Australian Catholic University, North Sydney, NSW, 2060, Australia; niusha.shafiabady@acu.edu.au

⁴ Faculty of Science and Technology, Charles Darwin University, Sydney, NSW, 2000, Australia

* Correspondence: james.vakilian@gmail.com

Abstract

The integration of international students into host countries and their cultures is a multifaceted challenge that significantly impacts their academic success and well-being. This study leverages Explainable Artificial Intelligence (XAI) to model and interpret variables associated with the self-rated integration of 175 international students at Charles Darwin University (CDU) in Australia, using data from a 42-question survey. Employing machine learning models, including Decision Tree (DT) and Gradient Boosting Machine (GBM), we use XAI techniques to identify variables most strongly associated with students' self-rated integration, including career confidence, perceived future happiness, and perceived career obstacles. SHAP analyses and partial dependence plots provide global and instance-level insights, revealing both the magnitude and directional effects of these features. The findings highlight the predictive relevance of psychological and social variables in students' self-rated integration, offering exploratory insights that inform targeted support programs. By enhancing model transparency through XAI, this research fosters trust in AI-driven educational interventions, addressing ethical considerations and promoting equitable outcomes for diverse student populations.

Keywords: international students; social integration; XAI; SHAP; LIME

Academic Editor: Miguel Angel Cazorla

Received: 6 May 2026

Revised: 9 June 2026

Accepted: 18 June 2026

Published: 25 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The increasing globalization of higher education has led to a significant rise in the number of international students pursuing academic opportunities abroad [1–4]. While these students contribute to the cultural and intellectual diversity of host countries, their successful integration into the new environment remains a complex and multifaceted challenge [5–8]. Traditional methods of assessing integration, such as surveys and interviews, often suffer from limitations in scale, objectivity [9–11], and the ability to capture nuanced patterns within large datasets [12–14].

Artificial Intelligence (AI), particularly machine learning techniques, offers a promising avenue for analyzing the integration process by leveraging the wealth of data generated through students' academic performance [15], social interactions [16,17], and

engagement with cultural activities [18,19]. However, the “black box” nature of many AI models poses a significant barrier to understanding ‘why’ certain factors are predictive of successful integration. This lack of transparency can hinder the development of effective support systems and interventions designed to facilitate a smoother transition for international students [20,21].

Explainable Artificial Intelligence (XAI) emerges as a critical solution to address this challenge [22,23]. By providing insights into the decision-making processes of AI models, XAI enables us to identify variables associated with international students’ self-rated integration, such as language proficiency, social network composition, participation in cultural events, and academic support systems [24,25]. Furthermore, XAI can reveal potential biases in the data or algorithms that may disproportionately affect certain student populations, promoting fairness and equity in integration initiatives [26,27].

This paper explores the application of XAI techniques to model the integration of international students into the host country and its culture. We focus on developing transparent and interpretable AI models that can identify variables associated with students’ self-rated integration, uncover hidden patterns in student behavior, and provide actionable insights for universities and policymakers [28]. By leveraging XAI, we aim to move beyond simply predicting integration outcomes to understanding the patterns contributing to model predictions, ultimately fostering a more inclusive and supportive environment for international students [29].

In this study, we define integration as the process by which international students balance the maintenance of their cultural identity with active participation in the social, academic, and cultural life of the host country, as conceptualized in Berry’s acculturation model [16]. This model posits integration as one of four acculturation strategies, where individuals adopt elements of the host culture while retaining aspects of their original culture, leading to positive adaptation outcomes. This definition underpins our focus on self-rated integration (survey question Q36), which captures students’ perceptions of their cultural and social adaptation in Australia.

This research contributes to the growing body of literature on XAI by demonstrating its potential to address complex social challenges within international education. Using data from a 42-question survey completed by 175 international students in Australia, we apply machine learning and explainability techniques to identify the social, psychological, and economic factors most strongly associated with self-rated integration [30–33].

The novelty of this study is not limited to the application of SHAP and LIME to survey data. Conceptually, the study links interpretable machine learning with Berry’s acculturation framework to examine international student integration as a multidimensional educational and sociocultural process, rather than as a purely academic performance or retention outcome. Methodologically, the study combines predictive modelling with global and instance-level XAI techniques to identify how social, psychological, academic, financial, and career-related survey variables contribute to students’ self-rated integration. This provides a transparent framework for interpreting model outputs in a small-scale educational dataset and demonstrates how XAI can be used to generate institutionally relevant hypotheses for student support, rather than only to maximise predictive accuracy.

2. Research Methodology

Various explainability methods are used to understand how machine learning models make decisions [34,35]. They aid in our understanding of the model’s process and the traits or variables that influence a prediction the most [36,37]. This is especially important when working with complicated models like ensemble models or deep neural networks, which are frequently referred to as “black boxes” because of their level of complexity [38,39].

This study provides a comprehensive analysis of XAI techniques for predicting the integration of international students into the host country and its culture, with a particular emphasis on model performance and interpretability. The questions are combinations of dependent, independent, and intervening variables, as informed by Berry's acculturation model. For instance, the dependent variable, self-rated integration (Q36: "How do you rate your integration in Australia and its culture?"), aligns with Berry's model in terms of capturing the extent to which students perceive themselves as balancing their cultural identity with participation in Australian society. For model training and interpretation, the three Q36 response categories were encoded consistently as Class 0 = Excellent integration, Class 1 = Fair integration, and Class 2 = Poor integration. These class labels correspond to the original survey responses 1 = Excellent, 2 = Fair, and 3 = Poor, respectively. To avoid ambiguity, the Section 3 reports both the model class label and the corresponding survey response category where relevant.

Model performance was evaluated using multiple indicators, including accuracy, precision, recall/sensitivity, specificity, and F1 score, rather than accuracy alone. This was important because the Q36 outcome contains three ordered response categories and performance may be affected by the distribution of responses across classes [40–42]. SHAP (SHapley Additive exPlanations) [43] and LIME (Local Interpretable Model-agnostic Explanations) [44] are employed to ensure transparent predictions for the two best performing methods in terms of accuracy (GBM Accuracy: 0.787234, DT Accuracy: 0.765957), critical for educational applications.

Prior to model training, survey responses were converted into machine-readable categorical or ordinal variables according to the response options listed in Table 1. Binary yes/no responses were encoded as categorical binary variables, while ordered response categories, such as levels of confidence, happiness, stress, study habits, and self-rated integration, were encoded according to their ordinal survey structure. The outcome variable was Q36, self-rated integration in Australia and its culture, which was encoded as Class 0 = Excellent integration, Class 1 = Fair integration, and Class 2 = Poor integration. Complete survey responses were used for the analysis; therefore, no imputation procedure was applied. Because the dataset was relatively small, model evaluation did not rely on accuracy alone. Sensitivity, specificity, F1 score, precision/recall, and accuracy were considered together to assess performance and to reduce the risk of misleading interpretation due to possible class imbalance.

Table 1. The questionnaire which is used as the input features.

No.	Questions	Answer Sheet
Q1	Where are you originally from?	India, Bangladesh, Nepal, Pakistan, Vietnam, Philippines, China
Q2	Are you an Australian citizen or permanent resident?	YES or NO
Q3	Are you here on student visa?	YES or NO
Q4	Do you have to work to support yourself?	YES or NO
Q5	Do you work in an area relevant to your professional expertise?	YES or NO
Q6	Does your family support you while you are studying in Australia?	YES or NO
Q7	Do you have a circle among your friends/family to mentally/emotionally support you in Australia?	YES or NO
Q8	Can you score at least 7 in an IELTS test or a similar English test?	YES or NO
Q9	Is English language a barrier for you?	YES or NO

Q10	Is your first degree related to your post-graduate studies?	YES or NO
Q11	Do you have a partner?	YES or NO
Q12	Are you living with your partner in Australia?	YES or NO
Q13	Do your parents have a university degree?	YES or NO
Q14	Do you consider your family to be well off financially?	YES or NO
Q15	Do your parents support you mentally?	YES or NO
Q16	Do your parents support you financially?	YES or NO
Q17	Are you here alone?	YES or NO
Q18	How many hours do you typically study per week?	(1) Less than 5 h
		(2) 5–10 h
		(3) 10–15 h
		(4) 15–20 h
Q19	How often do you complete your assignments on time?	(1) Always
		(2) Sometimes
		(3) Rarely
Q20	How do you rate your study habits and time management skills?	(1) Excellent
		(2) Good
		(3) Fair
		(4) Poor
Q21	Can you focus on the lesson in your classes?	YES or NO
Q22	How well do you engage in your studies in your opinion?	(1) Very much
		(2) Moderately
		(3) Somewhat
		(4) Not at all
Q23	How frequently do you feel depressed or hopeless?	(1) Never
		(2) Rarely
		(3) Sometimes
		(4) Often
		(5) Always
Q24	Can you eat well while living and studying here?	YES or NO
Q25	Do you do regular exercises?	YES or NO
Q26	How often do you feel anxious or stressed?	(1) Never
		(2) Rarely
		(3) Sometimes
		(4) Often
		(5) Always
Q27	Did you have a happy childhood?	YES or NO
Q28	Do you think of yourself as a successful person?	(1) Yes
		(2) Somehow
		(3) No
Q29	Do you experience any types of hardships that affect your focus on your studies (e.g., financial, trauma)?	(1) I do but I can handle it on my own.
		(2) I do and I can't manage alone and need help
		(3) I don't have any issues.
Q30	Do you like the major you are studying?	(1) Yes
		(2) Somehow
		(3) No
Q31	Are you on the right track to achieve your career aspirations?	(1) Yes
		(2) Somehow
		(3) No
Q32	How confident do you feel about achieving your career goals?	(1) Very confident
		(2) Somewhat confident

		(3) Not confident at all
Q33	Have you talked to anyone about your career aspirations?	YES or NO
Q34	What obstacles do you think may stand in the way of achieving your career aspirations?	(1) Lack of finance (2) Lack of access to educational resources (3) Discrimination or bias (4) Family responsibilities (5) Other
Q35	Do you think your studies are making you ready for your career?	(1) Yes (2) Somehow (3) No
Q36	How do you rate your integration in Australia and its culture?	(1) Excellent (2) Fair (3) Poor
Q37	Are you planning to stay in Australia and become an Australian citizen?	YES or NO
Q38	Do you want to integrate with Australian culture or be exactly the way you were before coming to Australia?	(1) I want to stay exactly like before and make no changes (2) I want to keep some components of my own culture and integrate with some parts of the Australian culture (3) I want to change everything and become exactly like an Australian born and bred in Australia
Q39	Do you like to stay and live in Australia all your life (note: mark 'No' if you would like to go to another country)?	YES or NO
Q40	How confident do you feel about being able to pay the bills in the future independently?	(1) Very confident (2) Somewhat confident (3) Not confident at all
Q41	Based on your performance and your aspirations, do you perceive yourself as a successful person in the future?	YES or NO
Q42	How do you rate your perceived future happiness in Australia?	(1) Very happy (2) Somewhat happy (3) Not happy at all

To evaluate model performance and reduce the risk of overfitting, the dataset of 175 complete responses was divided into training and testing subsets using an 80:20 split. The training set was used for model development, while the held-out test set was reserved for final model evaluation. Where feasible, stratified sampling was applied to preserve the distribution of the three Q36 self-rated integration classes across the training and test subsets. A fixed random seed was used to support reproducibility of the data split and model training process. In addition, 5-fold cross-validation was applied within the training data to assess model stability across different folds.

No automated or manual hyperparameter optimisation procedure was conducted. Specifically, GridSearchCV, RandomizedSearchCV, Bayesian optimisation, auto-sklearn, Optuna, Hyperopt, Hyperband, or other AutoML-based tuning procedures were not used. Instead, the Decision Tree and Gradient Boosting Machine models were trained using the standard scikit-learn classifier configurations. The main scikit-learn parameter settings used in the analysis are reported here. For the Decision Tree classifier, the key parameters were criterion = 'gini', splitter = 'best', max_depth = None, min_samples_split = 2, min_samples_leaf = 1, max_features = None, class_weight = None, and ccp_alpha = 0.0. For the Gradient Boosting classifier, the key parameters were loss = 'log_loss', learning_rate = 0.1, n_estimators = 100,

subsample = 1.0, criterion = 'friedman_mse', min_samples_split = 2, min_samples_leaf = 1, max_depth = 3, max_features = None, validation_fraction = 0.1, n_iter_no_change = None, tol = 0.0001, and ccp_alpha = 0.0. This decision was made because the study was exploratory and the dataset was relatively small. Extensive hyperparameter tuning on a small dataset could increase the risk of overfitting and produce model settings that are too specific to this sample. Therefore, the reported model performance should be interpreted within the exploratory scope of the study and in light of the modelling choices described above.

Given that the outcome variable was self-rated integration (Q36), we also acknowledge the possibility of conceptual overlap between Q36 and some survey variables that capture students' future intentions, perceived happiness in Australia, or attitudes toward integration. These variables were retained in the exploratory analysis because they represent theoretically relevant dimensions of students' perceived adaptation; however, the results should be interpreted as predictive associations rather than evidence of independent determinants. Future studies should examine reduced models that exclude conceptually proximate variables to assess the robustness of the identified predictors and reduce the risk of target leakage. The study employs a survey-based data collection approach to gather relevant information. Survey questions were developed through a literature review and consultations with education experts and success coaches at Charles Darwin University (CDU), ensuring alignment with constructs of student integration (e.g., cultural adaptation, academic engagement).

The selection and grouping of survey variables were informed by prior literature on international student integration [15,16,33] and consultations with CDU experts. Variables were categorized into dependent (Q36, self-rated integration), independent (e.g., country of origin [Q1], socioeconomic status (Q14), visa status (Q2)), and intervening variables (e.g., language proficiency (Q8), social support (Q7), psychological stressors (Q29)) based on Berry's acculturation model [23] and studies identifying key acculturation stressors [25,29]. This categorization ensures a comprehensive assessment of social, psychological, and economic factors influencing integration, aligning with theoretical frameworks and practical insights from educational practitioners.

Ethics approval was obtained from the Human Ethics Committee at CDU, as the survey targeted international students. Data was collected from 175 international students at CDU's Sydney campus, using a 42-question questionnaire, with complete responses obtained anonymously to protect respondents' identities. We further discuss the implications of XAI-driven analysis for student integration, highlighting its potential to inform educational decision-making and optimize resource allocation (e.g., targeted support programs) [45,46]. It underscores the importance of transparent and interpretable models in educational settings, where understanding the reasons behind predictions fosters trust as well as ethical considerations in AI systems [23,47].

The survey questions were adapted from existing literature and allocated to relevant constructs of the conceptual framework (e.g., social integration, language proficiency). The questionnaire comprises 42 questions with responses from 175 participants, and further details are explored in subsequent sections.

3. Results

We used an investigation questionnaire detailed in Table 1 for our analysis. To further develop the questionnaire with regard to the literature, the questions to include in the questionnaire were considered in consultation with education experts and success coaches dealing with international students at CDU. The questions are combinations of dependent, independent and intervening variables. It focuses on:

- Dependent Variable: Integration into Australian society and culture.

- Independent Variables: Country of origin, host country policies, previous education, socioeconomic status, and scholarship/work arrangements.
- Intervening Variables: Language proficiency, cultural adaptation, social support, access to resources, pre-arrival preparation, attitude and openness, discrimination/prejudice, and financial stability.

To investigate variables associated with the self-rated integration of international students at CDU, as captured by the survey item Q36 (*“How do you rate your integration in Australia and its culture?”*), machine learning models including Decision Tree, GBM were employed. SHAP (SHapley Additive exPlanations) analyses, partial dependence plots (PDPs), and feature importance metrics were utilized to interpret model predictions, providing both global and instance-level insights into feature contributions. Feature importance, directional effects, and their implications for understanding student integration will be further discussed in detail. As shown in Figure 1, the superiority of tree-based methodologies, most notably GBM and DT, suggests that the feature space may contain non-linear relationships and hierarchical interactions that linear or distance-based models (such as Logistic Regression and KNN) fail to fully capture. Conversely, the underperformance of Naive Bayes (NB) implies that the dataset violates the algorithm’s core assumption of feature independence, pointing to multicollinearity or dependencies among the predictor variables. Furthermore, the tight internal consistency across all five evaluation metrics (Accuracy, Precision, Recall, F1-score, and Specificity) within each respective model indicates a balanced dataset although small; had there been severe class imbalances, a distinct divergence between accuracy and precision/recall would be evident. These results suggest that non-linear, recursive partitioning algorithms may be better suited to modelling the decision boundaries in this dataset to adequately map the decision boundaries inherent to this specific classification problem.

The SHAP, PDP, and feature-importance analyses were used to interpret the behaviour of the trained models rather than to infer causal relationships. SHAP values indicate how individual variables contributed to model predictions, while PDPs show marginal modelled associations between feature values and predicted outcomes. Therefore, terms such as “importance,” “contribution,” and “association” are used in relation to model predictions, not as evidence that these variables causally determine students’ integration.

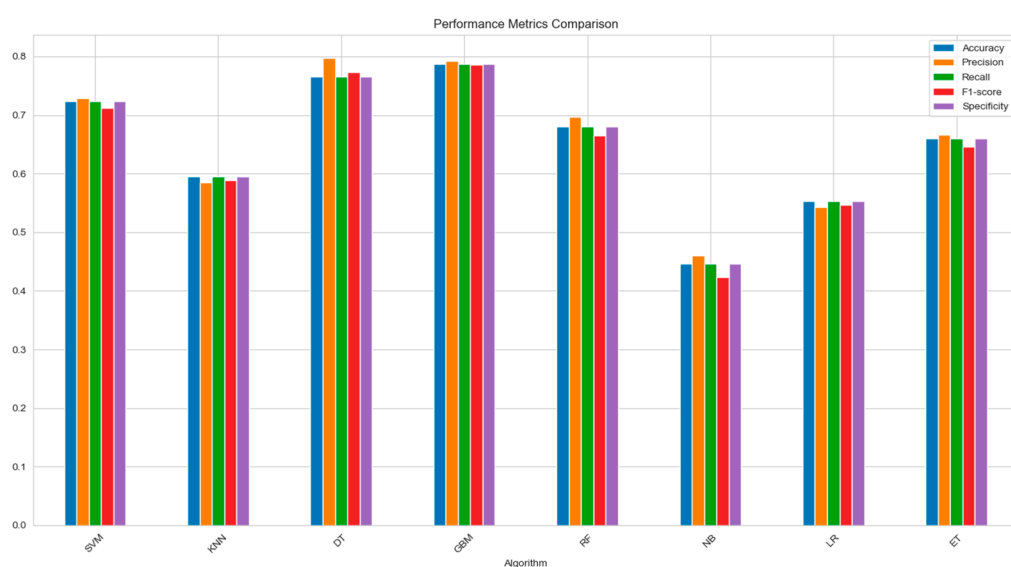


Figure 1. Comparative analysis of five performance metrics (Accuracy, Precision, Recall, F1-score, and Specificity) across eight machine learning classification algorithms.

3.1. Global Feature Importance

The SHAP plot shown in Figure 2 for the Decision Tree model revealed the top five predictors of international students' self-rated integration into Australian culture and society. These were: perceived future happiness in Australia, anticipated barriers to achieving career aspirations, confidence in attaining career goals, self-assessed study habits and time management, and perception of personal success. These features exhibited the highest mean absolute SHAP values, ranging from approximately 0.7 to 0.3, indicating their strong influence on model predictions. Among them, perceived future happiness showed the largest contribution to the model prediction (SHAP ≈ 0.7), followed by career-related obstacles (≈ 0.6) and confidence in goal attainment (≈ 0.5). Because perceived future happiness may be conceptually related to self-rated integration, this result should be interpreted as evidence of predictive association rather than as an independent explanatory factor.

Moderate contributions were observed from variables such as the alignment between prior and current academic fields, the presence of emotional support networks, and employment relevance to professional background (SHAP values ≈ 0.1). Meanwhile, factors including the necessity of working for financial support, family support during studies, and perceived future personal success yielded lower SHAP values (<0.05), suggesting a relatively limited role in predicting integration outcomes within this model.

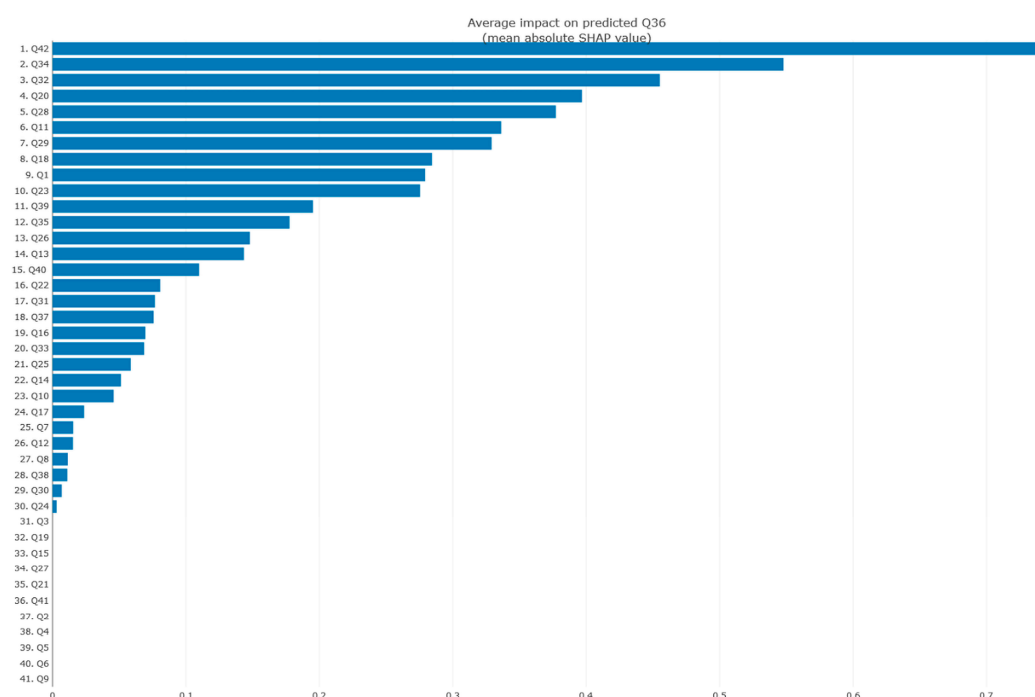


Figure 2. Ranking Feature Importance: SHAP Values for Q36 Predictions (Mean Absolute Impact).

A complementary feature importance analysis using the “drop loss” metric from a GBM classifier identified perceived future success, study habits and time management, and confidence in achieving career goals as the most influential factors in predicting students who rated their integration in the lowest category (Class 2, corresponding to the original Q36 response category “Poor integration”) as illustrated in Figure 3. These variables exhibited drop loss values of approximately 0.15, 0.05, and 0.03, respectively. Notably, this ranking diverged from the SHAP-based analysis by prioritizing perceived future success over perceived future happiness, illustrating how different importance metrics can yield varying insights.

At the lower end of the spectrum, variables such as emotional support from friends or family, mental support from parents, and the ability to maintain a healthy diet while studying had negligible importance (drop loss ≤ 0.001). Additionally, features such as family's financial standing and studying in Australia alone registered small negative drop loss values (≈ -0.005), which may suggest the presence of model noise or potential overfitting.

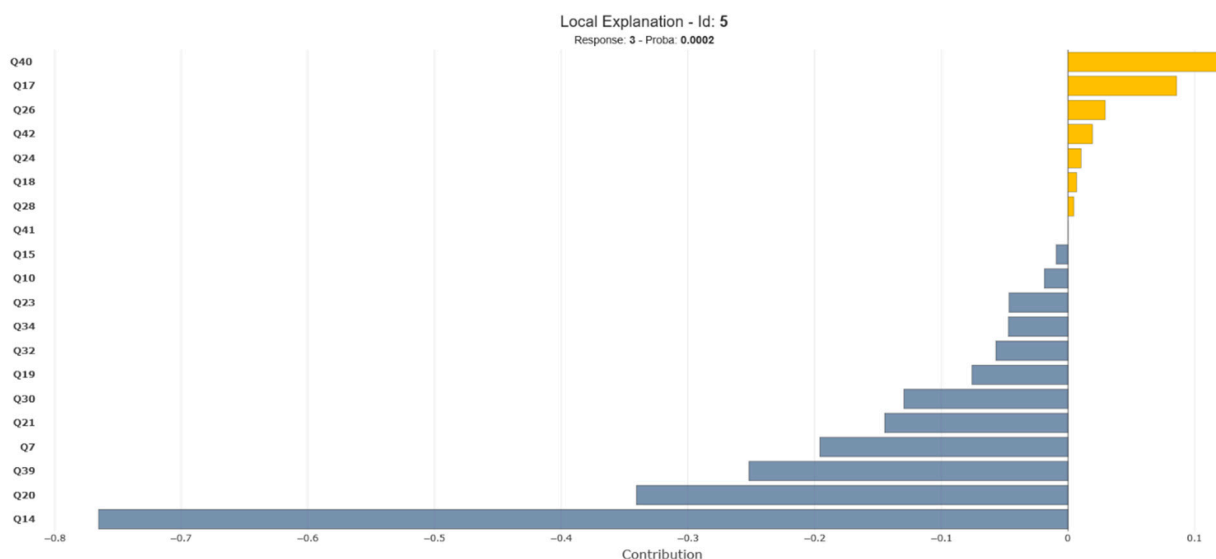


Figure 3. Local Explanation: Feature Contribution to Response with Probability 0.0002 (Id: 5).

In the multi-class classification task conducted using GBM, the SHAP results were used to examine class-specific feature contributions across the three outcome categories (Class 0, Class 1, and Class 2) according to Figure 4. The features with the highest cumulative SHAP values reaching up to 1.4 were confidence in achieving career goals, perceived future happiness in Australia, and anticipated obstacles to career aspirations. Confidence in career goals was found to primarily influence predictions for Class 0, while perceived future happiness significantly contributed to both Class 0 and Class 1 classifications.

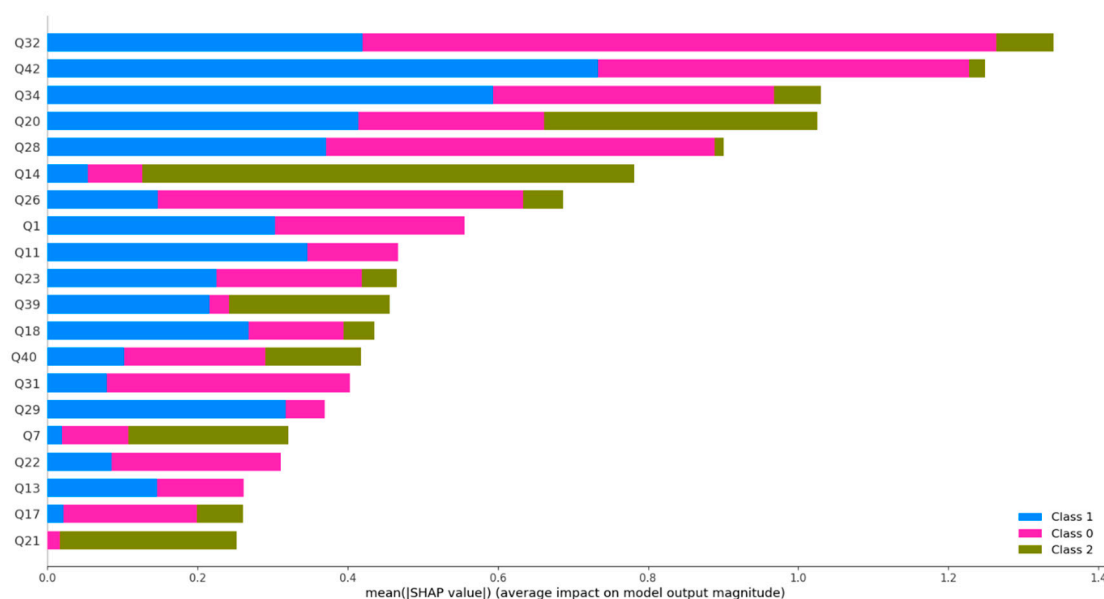


Figure 4. Feature Importance Across Classes: Average SHAP Values for Model Output Magnitude.

In contrast, family's financial standing and frequency of anxiety or stress were more influential in predicting Class 2, suggesting these factors are more relevant among students with lower levels of integration. Additionally, variables such as cohabitation with a partner and frequency of depressive symptoms demonstrated relatively balanced influence across all classes, indicating their generalized relevance to student integration. On the other hand, features like emotional support networks and parents' educational background had minimal predictive power, with SHAP values close to zero, reflecting limited impact in this context.

3.2. Directional Effects of Features

Figure 5 examines the log-odds of predicted integration levels (Q36) offered nuanced insights into both the direction and magnitude of individual feature effects. The feature representing perceived future happiness in Australia displayed the widest range of SHAP values (approximately -1 to $+1$). Higher response levels for this feature (color-coded red) were associated with increased SHAP values up to $+1$ indicating a strong positive influence on the predicted log-odds of higher integration. Conversely, lower values (in blue) corresponded to negative SHAP contributions, reducing the likelihood of a positive integration rating.

In contrast, the feature reflecting the need to work for financial support exhibited minimal variance, with SHAP values concentrated around zero, underscoring its negligible effect on the model's predictive confidence. Collectively, these findings emphasize the central role of students' anticipated future well-being in shaping their cultural and social integration outcomes.

The feature contribution for Q14, family financial standing, illustrates its context-specific influence on the model prediction for Class 2, corresponding to the original Q36 response category "Poor integration" of self-rated integration in Australia and its culture (Q36), as shown in Figure 6. When students reported lower financial well-being (corresponding to model input values near -0.5), SHAP values ranged from approximately -0.1 to $+0.5$ primarily concentrated between 0 and $+0.2$ correlating with higher predicted probabilities of integration (up to 0.16). In contrast, when students perceived their families to be more financially secure (input values near $+0.5$), SHAP values ranged from -0.5 to $+0.1$, predominantly falling between -0.3 and 0 , which aligned with significantly lower predicted probabilities (≈ 0.02).

This pattern should be interpreted cautiously. Although the model suggested that lower perceived family financial standing was associated with a higher predicted probability for this integration class, this finding is counterintuitive and may reflect sample-specific variation, interactions with other variables, feature encoding, or model instability arising from the small dataset. Therefore, this result should not be interpreted as evidence that lower family financial standing improves integration. Rather, it should be viewed as an exploratory model-generated pattern that requires validation in larger datasets and further qualitative investigation.

3.3. Partial Dependence Plots

The PDP generated for the GBM classifier provides detailed insights into the marginal effects of survey variables on students' self-rated integration (Q36). According to Figure 7, features such as student visa status and family support while studying exhibited monotonic trends, indicating consistent directional influences on integration predictions. Conversely, variables including country of origin, need to work for financial support, mental support from parents, weekly study hours, perceived career readiness, and social support network size demonstrated sharp inflection points in their PDPs, suggesting the presence of nonlinear relationships or threshold effects.

Meanwhile, features like the alignment of undergraduate and postgraduate studies and frequency of depressive symptoms showed flat PDP curves, implying limited marginal influence on integration outcomes—potentially due to complex higher-order interactions with other variables.

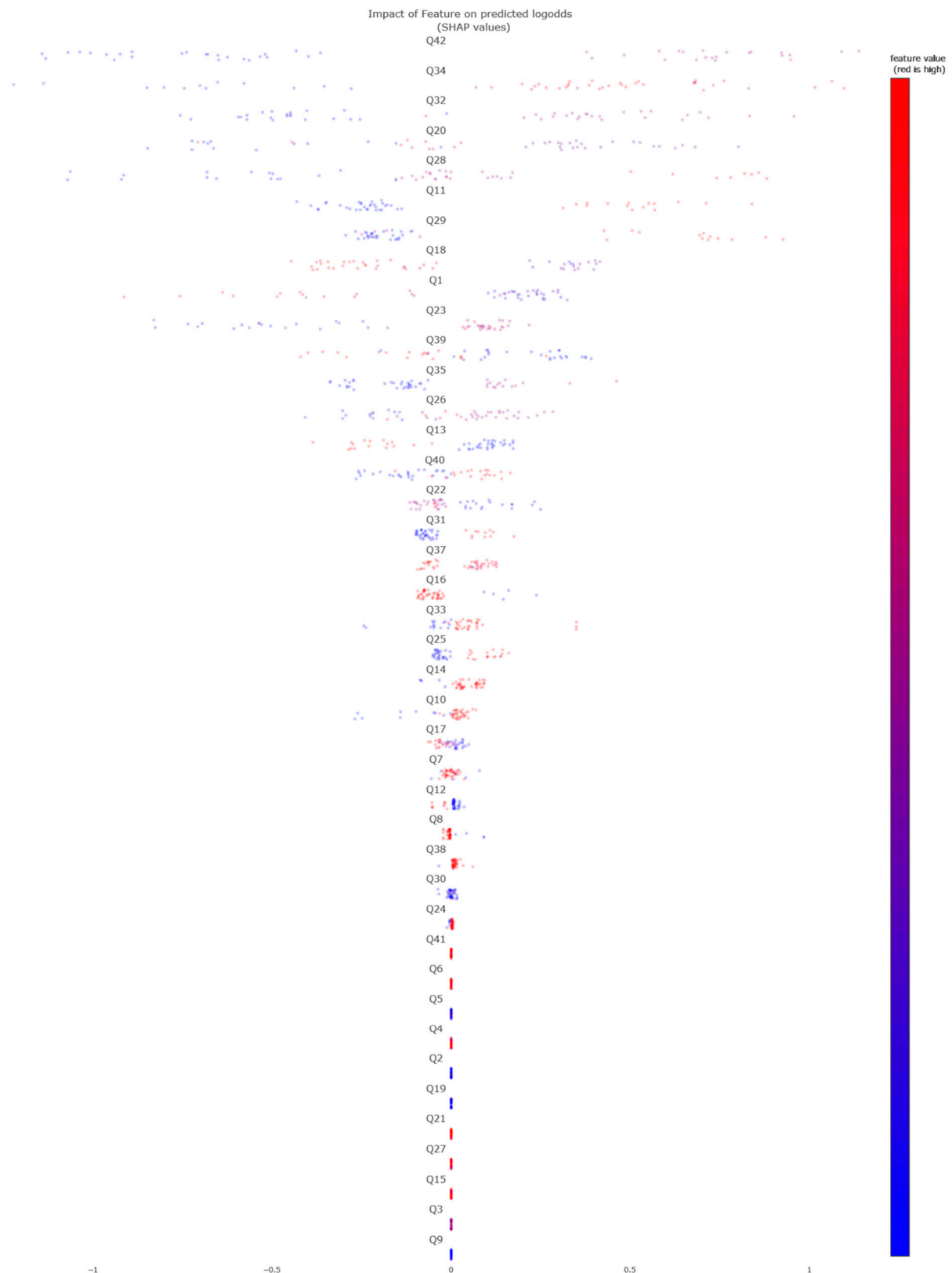


Figure 5. Impact of Features on Predicted Log-Odds: SHAP Values.

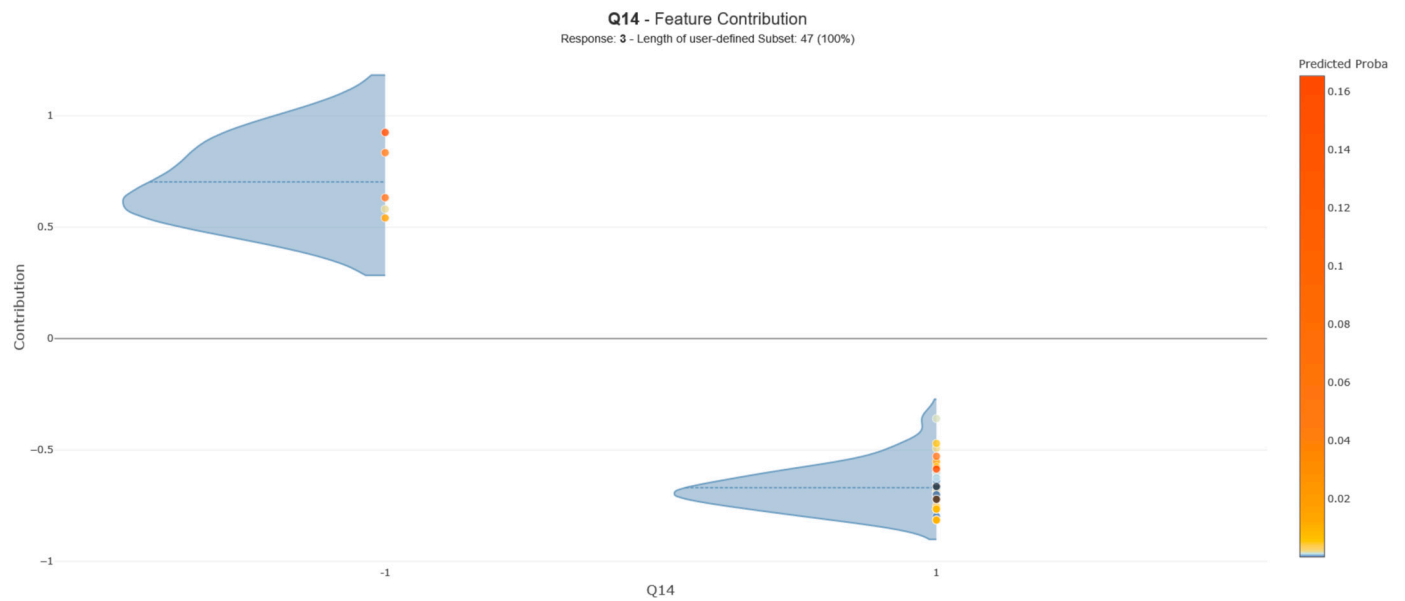
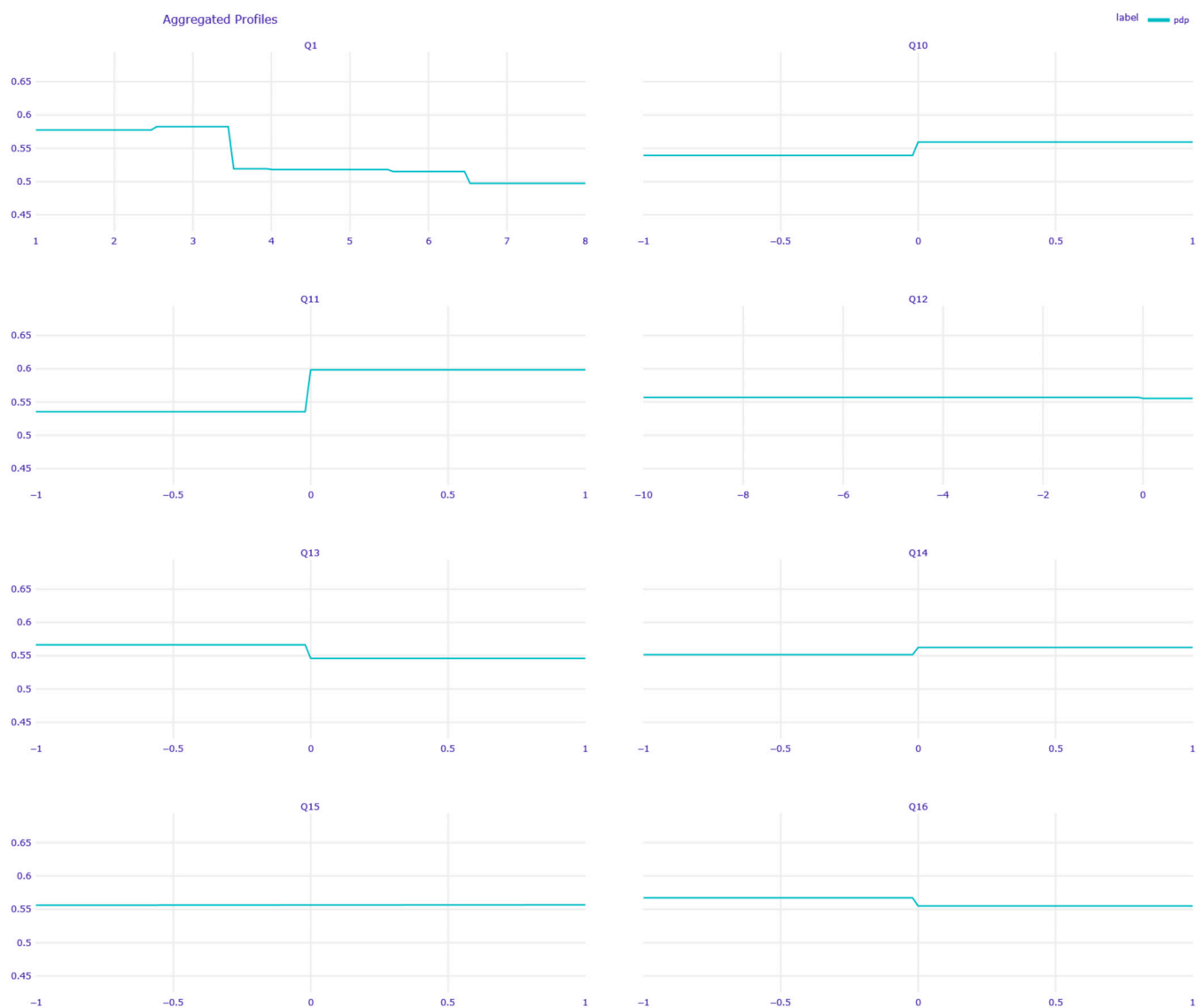
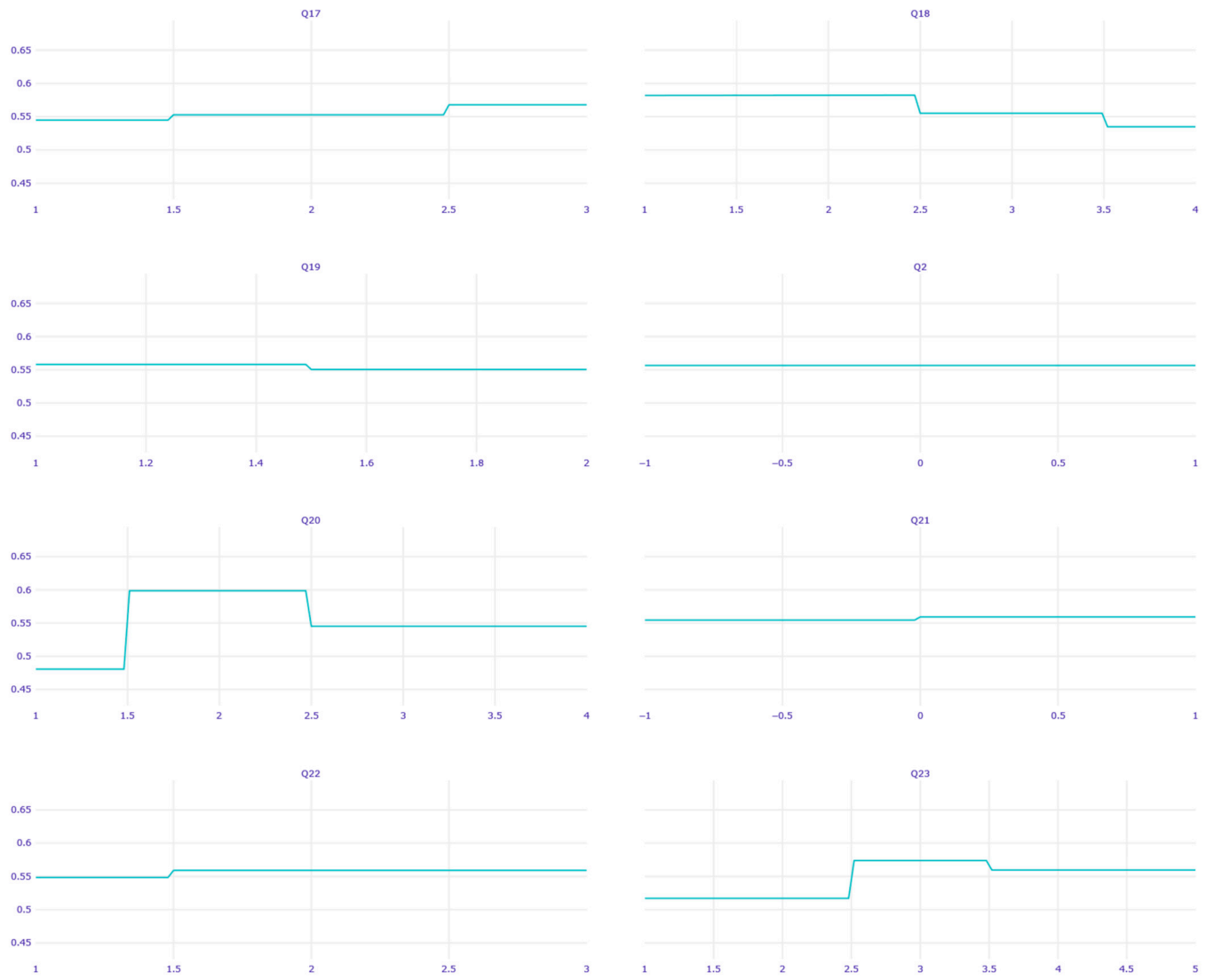


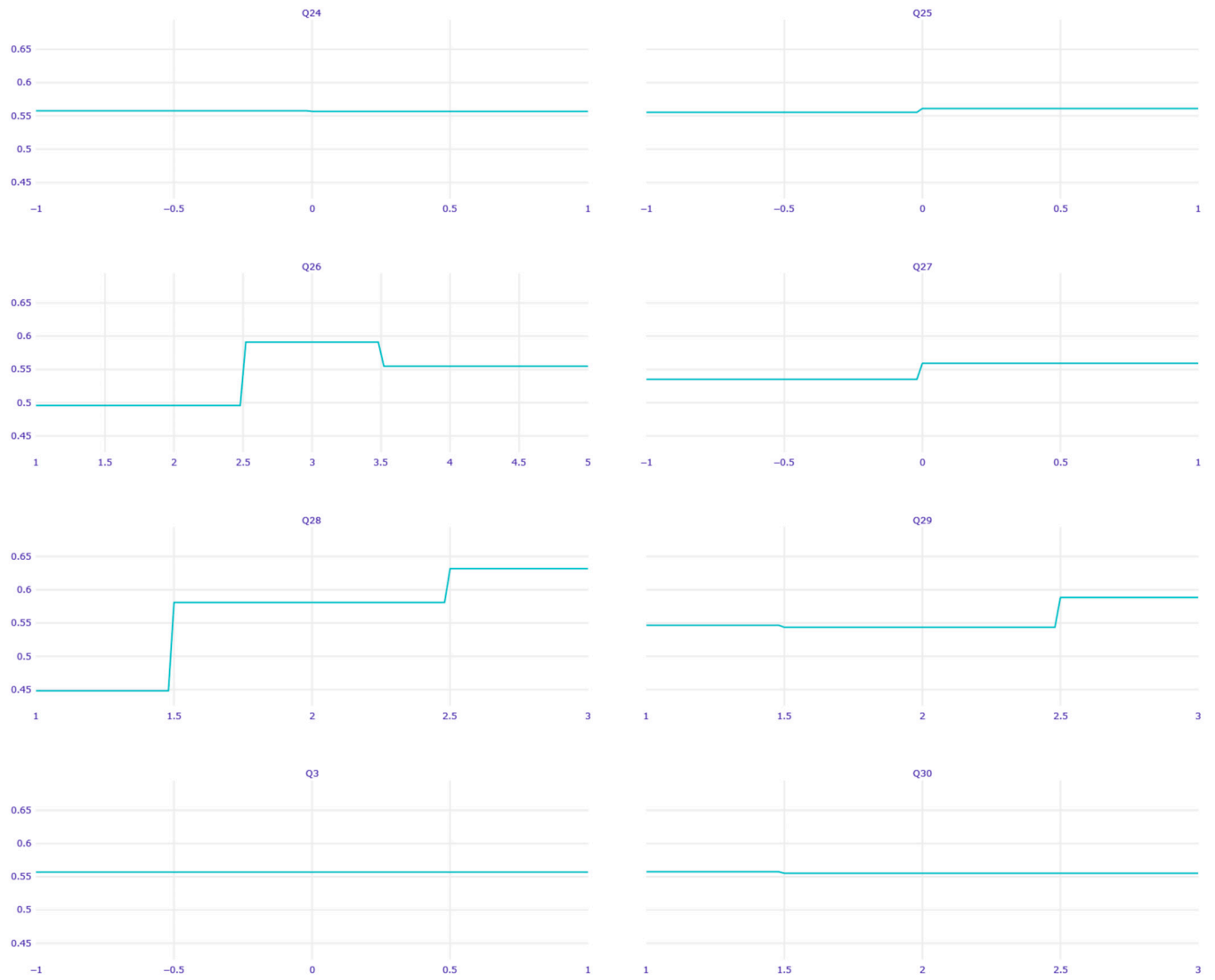
Figure 6. Effect of Q14, family financial standing, on the model prediction for Class 2, corresponding to the original Q36 response category “Poor integration” of self-rated integration in Australia and its culture (Q36).



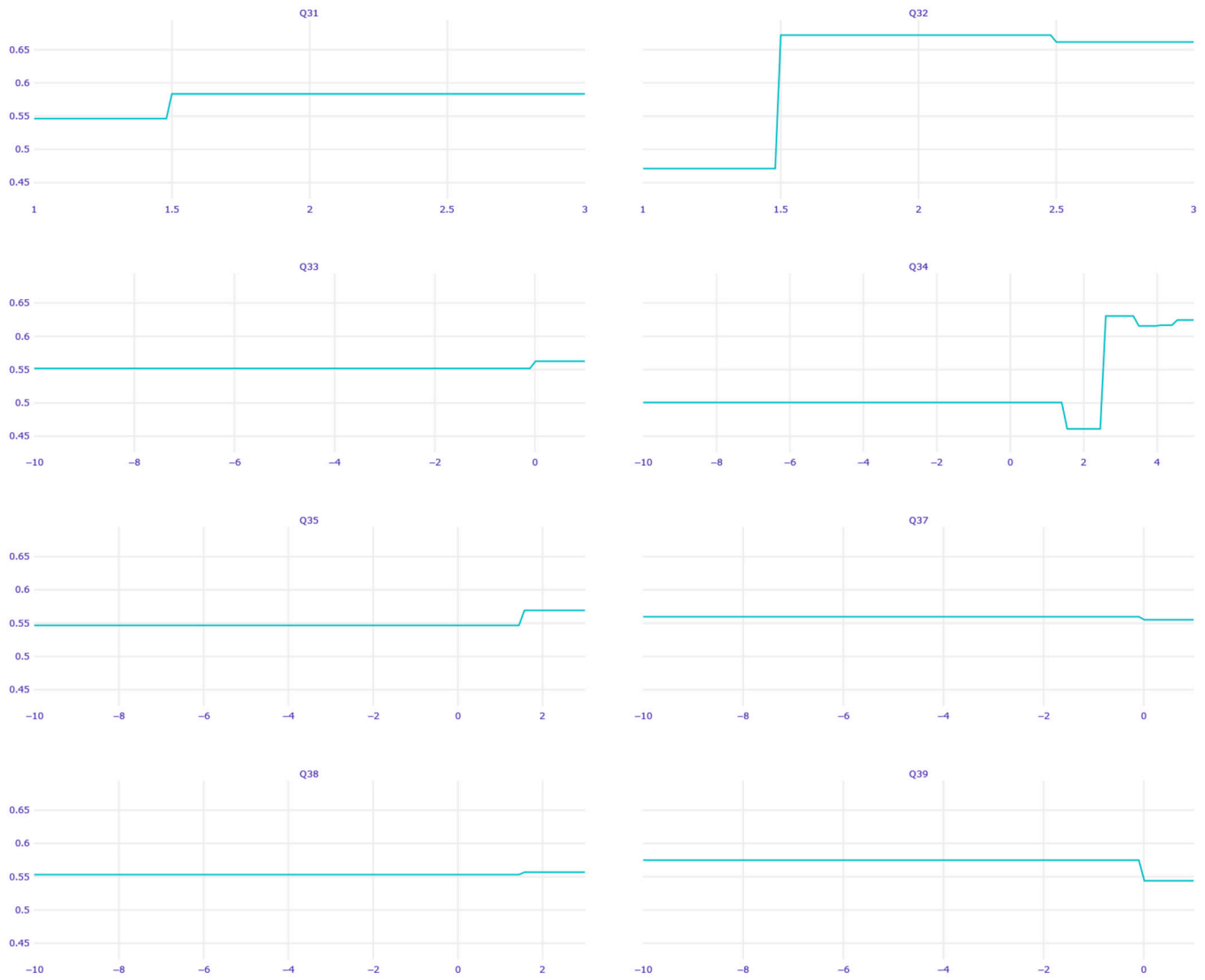
(a)



(b)



(c)



(d)

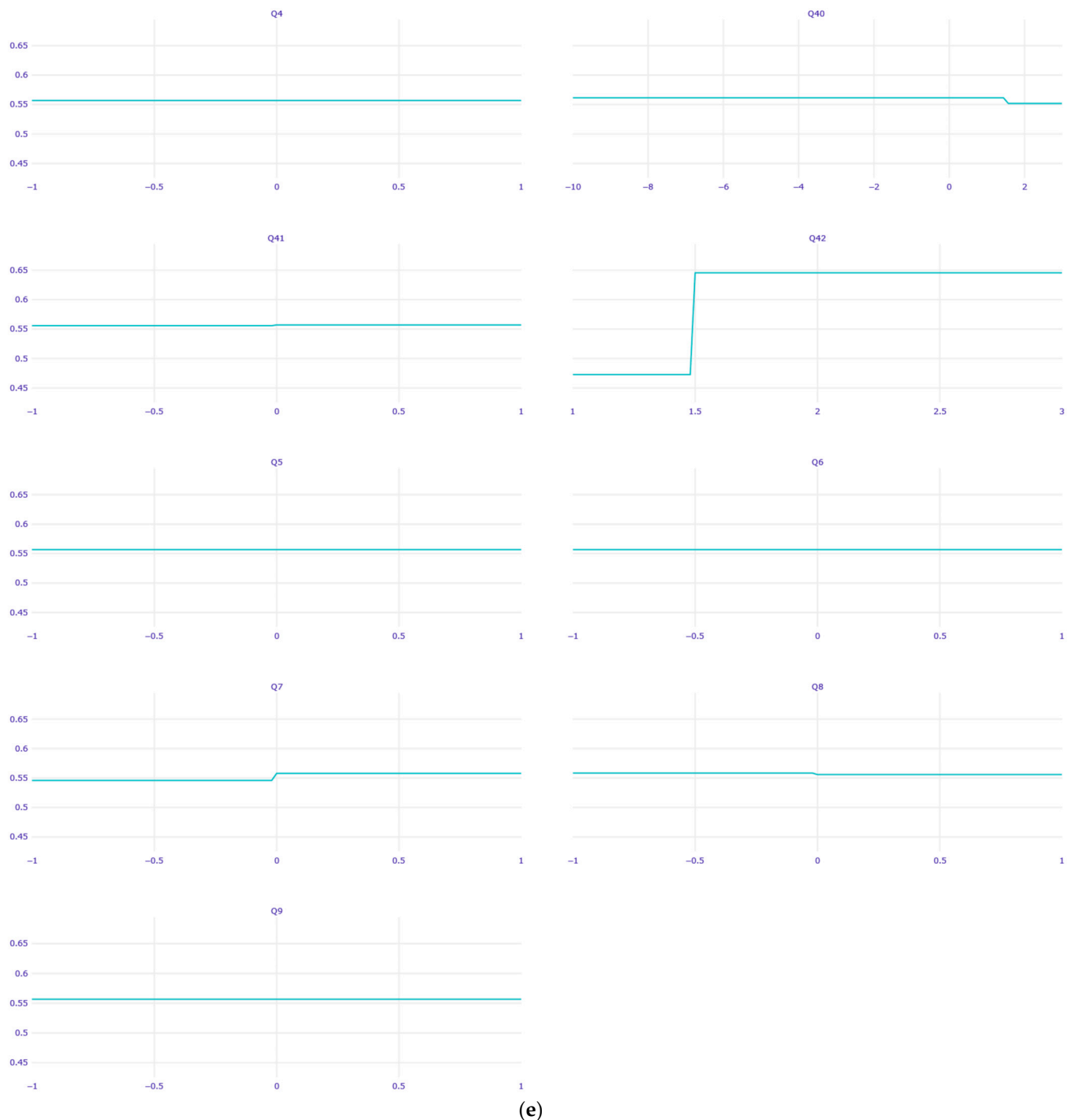


Figure 7. Partial Dependence Plots: Marginal Effects of Q1 to Q40 on Self-Rated Integration (Q36). (a) Background, family and support (Q1, Q10–Q16): Country of origin, prior/current study alignment, partner and living status, parents’ education, family finances, and parental support. (b) Visa/living, study and wellbeing (Q2, Q17–Q23): Citizenship/PR status, living alone, study habits (hours, completion, time management, focus, engagement), and depressive/hopeless feelings. (c) Health, stress and academics (Q3, Q24–Q30): Student visa status, diet, exercise, anxiety/stress, childhood happiness, self-perceived success, study hardships, and affinity for their major. (d) Career and settlement (Q31–Q35, Q37–Q39): Career aspirations, confidence, readiness, perceived obstacles, intention to stay/live long-term in Australia, and cultural integration preferences. (e) Work, language, finances and future (Q4–Q9, Q40–Q42): Work status and relevance, English proficiency/barriers, social support, ability to pay bills, and perceived future success and happiness in Australia.

3.4. Instance-Level Interpretability

Two instance-level SHAP analyses provide detailed insights into individual predictions of students' self-rated integration (Q36). In one SHAP plot for a single student instance yielding a predicted probability of 0.907 confidence in achieving career goals (value = 2.0) emerged as the strongest positive contributor (+0.557), followed by perceived future happiness in Australia (+0.115) and self-assessed study habits and time management (+0.072) as detailed in Figure 8. Minor negative influences from family's financial standing and experiencing hardships affecting studies (−0.04 and −0.003, respectively) were insufficient to offset these positive effects, resulting in a high likelihood of positive integration.

Similarly, Shapley values plot corroborated the key roles of career confidence (+0.211) and self-perception of success (+0.094) as major positive contributors, while variables such as having a partner and living in Australia alone exerted minor negative impacts (−0.021 and −0.015).

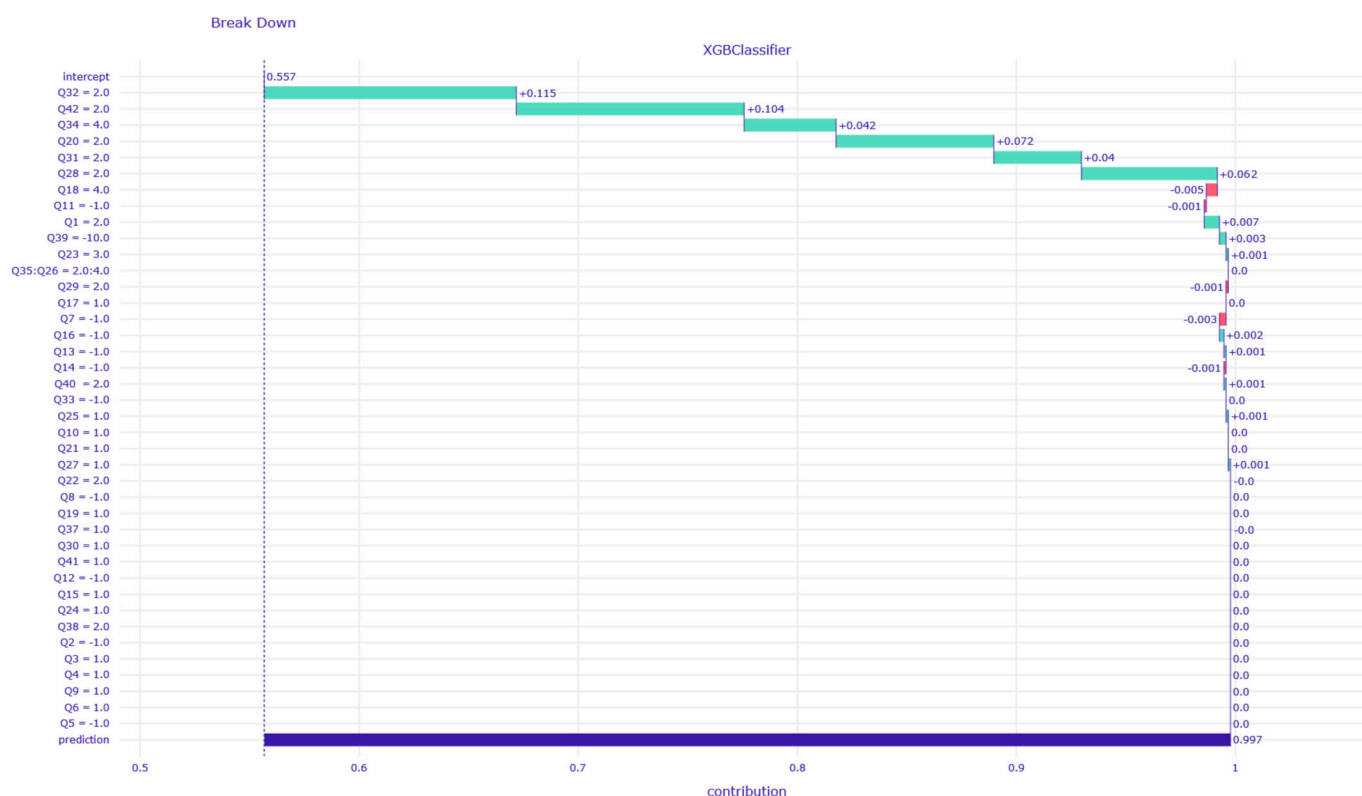


Figure 8. GBM Classifier Breakdown: Feature Contributions and Predicted Probabilities for Q1 to Q40.

In contrast, for instance ID 5 associated with a very low predicted probability (0.0002) for the lowest integration category (Class 2, corresponding to the original Q36 response category “Poor integration”) highlighted low confidence in financial independence (−0.7) and studying in Australia alone (−0.2) as the primary negative contributors driving this outcome as illustrated in Figure 9. Although perceived future happiness in Australia provided a minor positive influence (+0.1), it was insufficient to overcome stronger negative effects from maintaining a healthy diet, family's financial standing, and weekly study hours, which collectively reduced the predicted probability.

Across the global and instance-level analyses, the same group of career- and future-oriented variables appeared repeatedly among the most important contributors to model predictions. However, the ranking of individual variables varied across SHAP, drop-loss,

and instance-level explanations, indicating that these interpretability tools provide complementary rather than identical perspectives. Variables with consistently low predictive contribution, such as the need to work for financial support, emotional support networks, and parental mental support, may warrant further investigation before being considered for exclusion from future streamlined models.

The directional effects of family's financial standing highlighted the nuanced role of feature values in shaping integration outcomes, with partial dependence plots revealing both linear and nonlinear relationships. Instance-level analyses further demonstrated variability, with confidence in future financial independence emerging as a dominant factor in certain individual cases despite having lower global importance—emphasizing the necessity of context-specific interpretation.

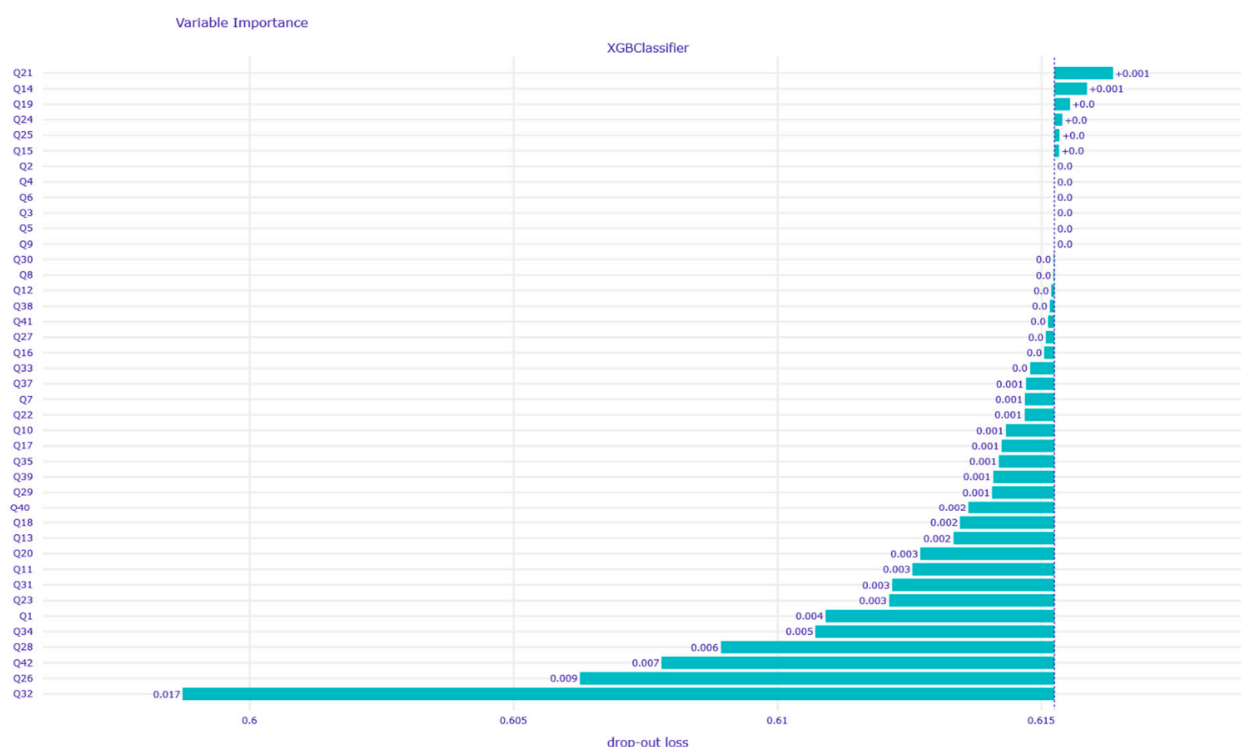


Figure 9. GBM Classifier Variable Importance: Drop Loss Analysis for Q1 to Q40 Features.

Collectively, these findings offer actionable insights for prioritizing interventions aimed at enhancing student integration. High-impact features linked to cultural adaptation and social support warrant deeper investigation to inform targeted support strategies. Moreover, the integration of global and local interpretability methods enhances model transparency, fostering greater stakeholder trust and a clearer understanding of predictive outcomes.

4. Discussion

The findings indicate that students' future-oriented perceptions, particularly those related to career confidence, perceived future happiness, and perceived career barriers, were among the strongest variables associated with model predictions of self-rated integration. Rather than treating these variables as causal drivers, the results suggest that students' expectations about future success and well-being may be important markers of perceived integration in this dataset. The variation between SHAP, drop-loss, and instance-level explanations also shows the value of using multiple interpretability methods, as each method highlights different aspects of the model's behaviour.

The directional pattern observed for family financial standing illustrates how specific feature values contributed to model predictions; however, this counterintuitive association should be interpreted cautiously, as it may reflect sample-specific variation, feature interactions, coding effects, or model instability rather than a substantive relationship. Instance-level analyses further revealed considerable variability; for example, confidence in future financial independence was a dominant factor in some individual cases despite showing lower importance at the global level, underscoring the importance of context-specific interpretations.

These findings offer actionable insights for prioritizing interventions aimed at improving student integration. High-impact features tied to cultural adaptation and social support warrant deeper exploration to inform targeted support strategies. Additionally, the combined use of global and local interpretability tools enhances model transparency, fostering greater trust among stakeholders and a clearer understanding of the variables contributing to model predictions.

There are several limitations that should be considered when interpreting the results. First, the dataset consists of only 175 survey responses while incorporating 42 questionnaire variables. Although 5-fold cross-validation and hold-out validation were employed to reduce model instability, the relatively small sample size compared with the number of predictors increases the risk of overfitting, whereby models may capture patterns specific to this dataset rather than relationships that generalize to broader international student populations. Consequently, the reported predictive performance should be interpreted cautiously and viewed as exploratory rather than definitive. Future studies should validate the findings using substantially larger and more diverse datasets collected across multiple institutions and geographical contexts.

Second, the sample was collected from a single Australian university campus, which may further limit external validity and the transferability of the findings to other educational environments.

Third, the reliance on self-rated integration (Q36) as the dependent variable introduces potential response bias, as students' perceptions may be influenced by subjective factors such as mood, personality characteristics, or cultural differences in self-reporting. Future research could incorporate objective indicators such as academic performance, retention, participation in university activities, and measurable social engagement metrics to complement self-reported assessments. Finally, because the study is cross-sectional and survey-based, the findings should not be interpreted as evidence of causal relationships. The XAI results indicate variables associated with model predictions of self-rated integration, rather than determinants or causal drivers of integration outcomes. In addition, some predictors may be conceptually close to the outcome variable, particularly variables relating to students' future happiness in Australia, intention to remain in Australia, or attitudes toward cultural integration. This overlap may inflate the apparent predictive importance of such variables and should be considered when interpreting the XAI results. Future work should compare the present model with reduced models that exclude conceptually overlapping predictors to evaluate whether the main findings remain stable.

5. Conclusions

This study leverages SHAP and advanced visualization techniques to elucidate the predictive dynamics of machine learning models targeting Q36, utilizing Decision Tree, Gradient Boosting Machine. Through SHAP summary plots, partial dependence plots, and instance-level breakdowns, the analysis uncovers a nuanced hierarchy of feature influences, with Q42 (perceived future happiness), Q34 (career obstacles), and Q32 (career confidence) consistently identified as the strongest predictive variables associated with

self-rated integration outcomes, while features like Q04 and Q06 show minimal impact. Multi-class SHAP analysis for GBM highlights class-specific effects, with Q32 dominating Class 0 predictions and Q14 showing strong predictive relevance for Class 2, enabling tailored intervention strategies. Instance-level insights, such as Q40's pronounced negative contribution to model predictions in specific cases, emphasize the importance of localized explanations in contextualizing global trends. PDPs reveal both monotonic and nonlinear feature relationships, informing model refinement and validation. By enhancing transparency with XAI tools, this research empowers stakeholders to prioritize high-impact factors for resource allocation, optimize model performance, and build trust in predictive systems. The findings should be interpreted in light of the relatively small sample size and the potential for model overfitting despite the use of cross-validation procedures. Future work could incorporate temporal data or additional contextual features to further enrich these insights, ensuring robust and adaptable models for diverse predictive applications. Moreover, further studies could incorporate objective measures, such as academic performance metrics or quantifiable social engagement indicators, to complement self-reported data and enhance the reliability of integration assessments.

Author Contributions: Conceptualization, J.V.; methodology, J.V., F.U.D., E.J.S. and N.S.; software, J.V.; validation, J.V. and M.H.; formal analysis, J.V. and M.H.; investigation, J.V.; resources, N.S.; data curation, J.V.; writing—original draft preparation, M.H., J.V., F.U.D., E.J.S. and N.S.; writing—review and editing, M.H., J.V., F.U.D., E.J.S. and N.S.; visualization, J.V. and M.H. All authors have read and agreed to the published version of the manuscript.

Funding: No funding was received for this study.

Institutional Review Board Statement: The study was conducted in accordance with the guidelines of the Declaration of Helsinki, and approved by the CDU Human Research Ethics Committee, Research integrity and Ethics, Research and Innovation (H24020 and 18 March 2024L) and the UNIVERSITY OF NEW ENGLAND HUMAN RESEARCH ETHICS COMMITTEE (HE-2025-2253-2784, 18/03/2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: The datasets presented in this article are not readily available due to the conditions of the Ethics Approval.

Conflicts of Interest: The authors have no relevant financial or non-financial interests to disclose.

References

1. Abdul-Rahaman, N.; Arkorful, V.E.; Okereke, T. Academic integration in higher education: A review of effective institutional strategies and personal factors. *Front. Educ.* **2022**, *7*, 856967. <https://doi.org/10.3389/educ.2022.856967>.
2. Arkoudis, S.; Dollinger, M.; Baik, C.; Patience, A. International students' experience in Australian higher education: Can we do better? *High. Educ.* **2018**, *77*, 799–813.
3. Carroll, J.; Ryan, J. *Teaching International Students: Improving Learning for All*; Routledge: London, UK, 2007. <https://doi.org/10.4324/9780203696132>.
4. Gaitán-Aguilar, L.; Hofhuis, J.; Bierwaczzonek, K.; Carmona, C. Social media use, social identification and cross-cultural adaptation of international students: A longitudinal examination. *Front. Psychol.* **2022**, *13*, 1013375. <https://doi.org/10.3389/fpsyg.2022.1013375>.
5. Montgomery, C.; McDowell, L. Social networks and the international student experience: An international community of practice? *J. Stud. Int. Educ.* **2009**, *13*, 455–466. <https://doi.org/10.1177/1028315308321994>.
6. Singh, J.K.N. Enhancing international student experience: Pre-support support services provided to postgraduate international students in a Malaysian research university. *Int. J. Incl. Educ.* **2021**, *27*, 972–986. <https://doi.org/10.1080/13603116.2021.1879953>.

7. Tang, L.; Zhang, C. Global Research on International Students' Intercultural Adaptation in a Foreign Context: A Visualized Bibliometric Analysis of the Scientific Landscape; SAGE Open: Thousand Oaks, CA, USA, 2023; pp. 1–26. <https://doi.org/10.1177/21582440231218849>.
8. UNESCO. Why Technology in Education Must Be on Our Terms. 2023. Available online: <https://www.unesco.org/en/articles/why-technology-education-must-be-our-terms> (accessed on 15 January 2025).
9. Udah, H.; Gatwir, K.; Francis, A. Resilience of International Students during a Global Pandemic: An Australian Context. *J. Int. Stud.* **2024**, *14*, 821–840.
10. Wang, T.; Lund, B.D.; Marengo, A.; Pagano, A.; Mannuru, N.R.; Teel, Z.A.; Pange, J. Exploring the potential impact of artificial intelligence (AI) on international students in higher education: Generative AI, chatbots, analytics, and international student success. *Appl. Sci.* **2023**, *13*, 6716. <https://doi.org/10.3390/app13116716>.
11. Ward, C.; Kennedy, A. The measurement of sociocultural adaptation. *Int. J. Intercult. Relat.* **1999**, *23*, 659–677. [https://doi.org/10.1016/S0147-1767\(99\)00014-0](https://doi.org/10.1016/S0147-1767(99)00014-0).
12. Sherry, M.; Thomas, P.; Chui, W.H. International students: A vulnerable student population. *High. Educ.* **2010**, *60*, 33–46.
13. Bianchi, C. Satisfiers and dissatisfiers for international students of higher education: An exploratory study in Australia. *J. High. Educ. Policy Manag.* **2013**, *35*, 396–409.
14. Longo, L.; Brcic, M.; Cabitza, F.; Choi, J.; Confalonieri, R.; Ser, J.D.; Guidotti, R.; Hayashi, Y.; Herrera, F.; Holzinger, A.; et al. Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Inf. Fusion* **2024**, *106*, 102301. <https://doi.org/10.1016/j.inffus.2024.102301>.
15. Smith, R.A.; Khawaja, N.G. A review of the acculturation experiences of international students. *Int. J. Intercult. Relat.* **2011**, *35*, 699–713.
16. Berry, J.W. Immigration, Acculturation, and Adaptation. *Appl. Psychol.* **1997**, *46*, 5–34. <https://doi.org/10.1111/j.1464-0597.1997.tb01087.x>.
17. Yeh, C.J.; Inose, M. International students' reported English fluency, social support satisfaction, and social connectedness as predictors of acculturative stress. *Couns. Psychol. Q.* **2003**, *16*, 15–28.
18. Khalil, M.; McGough, A.S.; Pourmirza, Z.; Pazhoohesh, M.; Walker, S. Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption—A systematic review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105287. <https://doi.org/10.1016/j.engappai.2022.105287>.
19. Bidwai, P.; Gite, S.; Pahuja, K.; Kotecha, K. A Systematic Literature Review on Diabetic Retinopathy Using an Artificial Intelligence Approach. *Big Data Cogn. Comput.* **2022**, *6*, 152. <https://doi.org/10.3390/bdcc6040152>.
20. Yang, G.; Ye, Q.; Xia, J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Inf. Fusion* **2022**, *77*, 29–52. <https://doi.org/10.1016/j.inffus.2021.07.016>.
21. van der Velden, B.H.M.; Kuijff, H.J.; Gilhuijs, K.G.A.; Viergeever, M.A. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med. Image Anal.* **2022**, *79*, 102470. <https://doi.org/10.1016/j.media.2022.102470>.
22. Dwivedi, R.; Dave, D.; Naik, H.; Singhal, S.; Omer, R.; Patel, P.; Qian, B.; Wen, Z.; Shah, T.; Morgan, G.; et al. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Comput. Surv.* **2023**, *55*, 1–33. <https://doi.org/10.1145/3561048>.
23. Das, A.; Rad, P. Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey (Version 2). *arXiv* **2020**. <https://doi.org/10.48550/ARXIV.2006.11371>.
24. Sado, F.; Loo, C.K.; Liew, W.S.; Kerzel, M.; Wermter, S. Explainable Goal-driven Agents and Robots—A Comprehensive Review. *ACM Comput. Surv.* **2023**, *55*, 1–41. <https://doi.org/10.1145/3564240>.
25. Vainio-Pekka, H.; Agbese, M.O.-O.; Jantunen, M.; Vakkuri, V.; Mikkonen, T.; Rousi, R.; Abrahamsson, P. The Role of Explainable AI in the Research Field of AI Ethics. *ACM Trans. Interact. Intell. Syst.* **2023**, *13*, 1–39. <https://doi.org/10.1145/3599974>.
26. Bokadia, H.; Yang, S.C.; Li, Z.; Folke, T.; Shafto, P. Evaluating perceptual and semantic interpretability of saliency methods: A case study of melanoma. *Appl. AI Lett.* **2022**, *3*, e77.
27. Gerlings, J.; Shollo, A.; Constantiou, I. Reviewing the Need for Explainable Artificial Intelligence (xAI) (Version 2). *arXiv* **2020**. <https://doi.org/10.48550/ARXIV.2012.01007>.
28. Meske, C.; Bunde, E. Transparency and Trust in Human-AI-Interaction: The Role of Model-Agnostic Explanations in Computer Vision-Based Decision Support. *arXiv* **2020**. <https://doi.org/10.48550/ARXIV.2002.01543>.
29. Hamida, S.U.; Chowdhury, M.J.M.; Chakraborty, N.R.; Biswas, K.; Sami, S.K. Exploring the Landscape of Explainable Artificial Intelligence (XAI): A Systematic Review of Techniques and Applications. *Big Data Cogn. Comput.* **2024**, *8*, 149. <https://doi.org/10.3390/bdcc8110149>.

30. Mandinach, E.B.; Honey, M. *Data-Driven School Improvement: Linking Data and Learning*; Teachers College Press: New York, NY, USA, 2008.
31. Khosravi, H.; Shum, S.B.; Chen, G.; Conati, C.; Tsai, Y.S.; Kay, J.; Knight, S.; Martinez-Maldonado, R.; Sadiq, S.; Gašević, D. Explainable artificial intelligence in education. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100074.
32. Sullivan, S.E.; Al Ariss, A. Making sense of different perspectives on career transitions: A review and agenda for future research. *Hum. Resour. Manag. Rev.* **2021**, *31*, 100727.
33. Sawir, E.; Marginson, S.; Deumert, A.; Nyland, C.; Ramia, G. Loneliness and international students: An Australian study. *J. Stud. Int. Educ.* **2008**, *12*, 148–180.
34. Buhrmester, V.; Münch, D.; Arens, M. Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey. *arXiv* **2019**. <https://doi.org/10.48550/ARXIV.1911.12116>.
35. Islam, S.R.; Eberle, W.; Bundy, S.; Ghafoor, S.K. Infusing Domain Knowledge in AI-Based “Black Box” Models for Better Explainability with Application in Bankruptcy Prediction. *arXiv* **2019**. <https://doi.org/10.48550/ARXIV.1905.11474>.
36. Dengel, A. Some Shades of Grey! Interpretability and Explainability of Deep Neural Networks. In Proceedings of the ACM Workshop on Crossmodal Learning and Application, Ottawa, ON, Canada, 5 June 2019; pp. 1–1.
37. Ali, A.M.O. Explainability in AI: Interpretable Models for Data Science. *Int. J. Res. Appl. Sci. Eng. Technol.* **2025**, *13*, 766–771. <https://doi.org/10.22214/ijraset.2025.66968>.
38. Swathi, Y.; Challa, M. A Comparative Analysis of Explainable AI Techniques for Enhanced Model Interpretability. In Proceedings of the 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 19–20 June 2023; pp. 229–234. <https://doi.org/10.1109/icpcsn58827.2023.00043>.
39. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. <https://doi.org/10.1038/s42256-019-0048-x>.
40. ŞAHİN, E.; Arslan, N.N.; Özdemir, D. Unlocking the Black Box: An in-Depth Review on Interpretability, Explainability, and Reliability in Deep Learning. *Neural Comput. Appl.* **2024**, *37*, 859–965. <https://doi.org/10.1007/s00521-024-10437-2>.
41. Lin, Z.Q.; Shafiee, M.J.; Bochkarev, S.; Jules, M.S.; Wang, X.Y.; Wong, A. Do Explanations Reflect Decisions? A Machine-Centric Strategy to Quantify the Performance of Explainability Algorithms. *arXiv* **2019**. <https://doi.org/10.48550/ARXIV.1910.07387>.
42. Velmurugan, M.; Ouyang, C.; Sindhgatta, R.; Moreira, C. Through the Looking Glass: Evaluating Post Hoc Explanations Using Transparent Models. *Int. J. Data Sci. Anal.* **2023**, *20*, 615–635. <https://doi.org/10.1007/s41060-023-00445-1>.
43. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
44. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should i trust you?” Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
45. Özturgut, O.; Murphy, C. Literature vs. practice: Challenges for international students in the US. *Int. J. Teach. Learn. High. Educ.* **2010**, *22*, 374–385.
46. Misra, R.; Castillo, L.G. Academic stress among college students: Comparison of American and international students. *Int. J. Stress Manag.* **2004**, *11*, 132.
47. Minh, D.; Wang, H.X.; Li, Y.F.; Nguyen, T.N. Explainable artificial intelligence: A comprehensive review. *Artif. Intell. Rev.* **2022**, *55*, 3503–3568.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.