

# ***Interpretable Machine Learning for Predicting Innovation and Entrepreneurship Education Performance in Universities: Evidence from 1,389 Chinese Students***

**Jinming Liu**

*Department of Student Affairs (School of Innovation and Entrepreneurship), Northwest Normal University, Lanzhou, Gansu, China  
415365042@qq.com*

**Abstract.** Evaluation of innovation and entrepreneurship education still rests on satisfaction surveys and aggregate counts, which cannot say which part of the provision is associated with student-level outcomes. Using open survey data from 1,389 Chinese university students, this study operationalises education performance as a proximal outcome indicator, a mean entrepreneurial intention of "agree" or above on a five-point scale, groups fifty variables into seven predictor blocks, and selects the model by training-set cross-validation alone. The selected random forest reaches an area under the ROC curve of 0.838 (95% confidence interval 0.799 to 0.876); its difference from an L2-regularised logistic regression is 0.021 (interval  $-0.003$  to  $0.047$ ), which includes zero, so the ensemble gains no distinguishable accuracy advantage. The value lies in the interpretation layer: Shapley additive explanations attribute 30.1% of global importance to entrepreneurship education provision, the highest of seven blocks, and in a leave-one-block-out ablation with every reduced model re-tuned it is the only block whose interval lies entirely above zero, losing 0.020 on removal (interval 0.005 to 0.036), while no independent contribution is detected for entrepreneurial spirit (27.9%) or perceived access to finance (20.9%). Course-to-practice linkage and lecturers' experience carry the largest attributions among the perceived provision items and are the candidates for feature-level follow-up.

**Keywords:** innovation and entrepreneurship education, performance prediction, interpretable machine learning, SHAP, higher education

## **1. Introduction**

Innovation and entrepreneurship education is now a core component of talent development in Chinese universities, supported by courses, competitions and dedicated funding. Evaluation has not kept pace: aggregate indicators such as course satisfaction, participant counts and funded projects show how much has been done, not which part of the provision is associated with differences in the student-level outcome, and so offer no basis for dividing resources among curriculum reform, faculty development, policy communication and financing support.

Predictive modelling offers another route. Educational data mining and learning analytics have established a mature paradigm for academic early warning [1, 2], and machine learning has been applied to entrepreneurial intention [3, 4]. Such work shows prediction is feasible but stops at comparing accuracy: a model returns a score without explaining how it is composed, so administrators cannot tell what to adjust. Interpretable machine learning addresses this gap [5]: Shapley additive explanations decompose a prediction into additive feature contributions [6] and now serve as an interface between models and decision-makers [7]. One question nonetheless remains open, because global importance answers which variables are most strongly associated with the outcome whereas allocation asks how much predictive power is lost when a provision is removed; where blocks overlap in information a block can rank high and still be fully substitutable.

This study therefore compares four models under one protocol, uses Shapley additive explanations for global attribution and effect direction, and quantifies each block's irreplaceable contribution through leave-one-block-out ablation, separating "important" from "not substitutable", with stability examined through leave-one-major-group-out validation, alternative thresholds and a continuous-outcome regression.

## 2. Literature review

Entrepreneurial intention is widely used as a proximal outcome indicator of innovation and entrepreneurship education, since intention is the immediate antecedent of behaviour [8]. Cross-national evidence links course provision and quality to intention through self-efficacy and motivation [9]; conditions outside the classroom matter too, as support policies and ecosystems operate through resource availability and perceived legitimacy [10], and curriculum, practice platforms and institutional guarantees must work together [11]. This literature relies mainly on structural equation modelling, whose path coefficients capture average association strength: they cannot be converted into the discriminative ability lost by removing one block of provision, and yield no student-level discrimination.

Prediction-oriented modelling is well established in education, most maturely in dropout early warning, where ensemble trees generally outperform linear baselines [1], and in learning analytics, which links early identification to feedback intervention [2]. Work in entrepreneurship is scattered: a kernel extreme learning machine has predicted entrepreneurial intention [3], and several algorithms have been compared on a sample of academics [4].

Model opacity remains the obstacle. One position holds that high-stakes settings should use inherently interpretable models rather than post-hoc explanations of black boxes [5]; another makes post-hoc explanation reliable, the exact Shapley value algorithm for tree models delivering attribution at acceptable cost [6], now standard in decision support systems [7]. Both lines leave the same gap: entrepreneurship education research offers associational evidence but no student-level discriminative tool, while educational prediction rarely treats provision as a distinct block whose contribution is quantified against individual traits and family background.

## 3. Methodology

### 3.1. Data and variables

The data come from a CC BY 4.0 survey of university students in Shandong Province, China, released with its parent paper [12]: 1,389 valid respondents, 628 men and 761 women, 882 undergraduate and 507 vocational students, with no missing values in any of the 74 fields.

The outcome is the mean of four five-point entrepreneurial intention items. Because this group's column names do not match the accompanying documentation, its identity was confirmed twice over: the confirmatory factor analysis in the data file gives standardised loadings of 0.849, 0.817, 0.813 and 0.789 on the intention latent variable, and Cronbach's alpha recomputed from the raw data is 0.887, consistent with the 0.87 reported in the parent paper. Sample mean intention is 3.538 (standard deviation 0.861). Under a pre-specified cut-point of "agree or above", respondents with a mean of at least 4.0 meet the performance standard, giving 544 positives, or 39.2%. The threshold was fixed before modelling; robustness checks use 3.5 and 4.25.

The predictors are 50 variables in seven blocks: individual and family background (12 items), policy awareness (2), entrepreneurial spirit (9), perceived policy support (7), perceived access to finance (8), entrepreneurship education provision (10) and entrepreneurial role models (2). The provision items cover course-to-practice linkage, lecturers' experience, interactive teaching, practical activities and campus climate. Nominal variables are one-hot encoded and binary variables coded so that 1 denotes the state named in the label, yielding 62 columns. Three groups are removed beforehand: the respondent identifier; the item on whether a student already has a venture idea, so proximal that it would make prediction near-tautological; and two scales totalling 18 items absent from the documentation, whose meaning cannot be verified, whereas an interpretability study requires every ranked feature to carry a verifiable definition.

### 3.2. Models and evaluation protocol

The sample is split by outcome stratification into 972 training and 417 held-out cases, and all four models are tuned by five-fold stratified cross-validation within the training set: logistic regression over the L2 regularisation strength; the random forest [13] over tree depth, minimum leaf size and features per split; XGBoost [14] over tree depth, learning rate and L2 penalty; and LightGBM, a leaf-wise variant of the same gradient-boosting framework [14], over leaves, learning rate and minimum samples per leaf. Model choice rests on the training set's cross-validated area under the curve alone, although the resulting held-out predictions are then reused for the model contrast and for the interpretation below; the selection margin is narrow, the random forest leading XGBoost by 0.0005. Equation (1) defines the label,

$$y_i = \mathbb{I}\left(\frac{1}{4} \sum_{j=1}^4 e_{ij} \geq \tau\right), \quad \tau = 4.0 \quad (1)$$

where  $e_{ij}$  is the score of student  $i$  on intention item  $j$ ,  $\tau$  the pre-set threshold and  $\mathbb{I}(\cdot)$  the indicator function. Intervals come from 2,000 bootstrap resamples of the held-out set and are therefore conditional on the fitted models; comparisons use a paired bootstrap whose point estimate is the observed difference in area under the curve over identical resampling indices.

### 3.3. Interpretation and robustness design

Interpretation uses the exact Shapley value algorithm for tree models [6] on the 417 held-out students, decomposing each predicted probability into feature contributions,

$$f(\mathbf{x}_i) = \phi_0 + \sum_{j=1}^p \phi_{ij} \quad (2)$$

where  $\phi_0$  is the base value, 0.391 here, and  $\phi_{ij}$  the contribution of feature  $j$  for student  $i$ . Global importance is the mean absolute value,

$$\Phi_j = \frac{1}{n} \sum_{i=1}^n |\phi_{ij}| \quad (3)$$

Direction is read from the Spearman correlation between a feature's value and its Shapley value; the mean signed Shapley value is reported alongside it as a descriptive quantity only, since it is centred on the base value and carries no directional guarantee. Equations (2) and (3) define the attribution, and the irreplaceable contribution of block  $b$  is the loss after removing all its features,

$$\Delta_b = \text{AUC}(\mathcal{F}) - \text{AUC}(\mathcal{F} \setminus b) \quad (4)$$

where  $\mathcal{F}$  is the full feature set, so Equation (4) reports a block-level loss. Every reduced model is re-tuned on the training set, so the loss does not arise from reusing the full model's hyperparameters; intervals come from the paired bootstrap and the seven are not corrected for multiple comparisons. Robustness is examined by leave-one-major-group-out validation, by label thresholds of 3.5 and 4.25, and by a regression on the continuous mean.

## 4. Results

### 4.1. Predictive performance

On training-set cross-validation the random forest ranks first at 0.8331, ahead of XGBoost at 0.8325, LightGBM at 0.8263 and logistic regression at 0.8188, and is selected. On the held-out set it attains an area under the curve of 0.838 (95% interval 0.799 to 0.876), accuracy 0.767, F1 0.692 and a Brier score of 0.157, with 211 true negatives, 43 false positives, 54 false negatives and 109 true positives. Its observed difference from logistic regression is 0.021, with a paired bootstrap interval of  $-0.003$  to  $0.047$  that includes zero: the ensemble gains no distinguishable advantage over the logistic specification, whose baseline already reaches 0.818.

### 4.2. Attribution share and irreplaceable contribution

Table 1 places each block's attribution share beside the loss incurred when it is removed. Entrepreneurship education provision leads with 30.1%, ahead of entrepreneurial spirit at 27.9%, perceived access to finance at 20.9% and individual and family background at 10.2%. The ablation supplies stronger evidence: of the seven blocks only provision has an interval entirely above zero, its removal lowering the area under the curve from 0.838 to 0.818, an observed loss of 0.020 (interval 0.005 to 0.036). No independent contribution is detected for any other block, entrepreneurial spirit included despite its 27.9% share, though the interval widths do not license calling them equivalent to no contribution. Removing policy awareness or role models slightly raises the area under the curve, and the interval for the former,  $-0.011$  to  $-0.001$ , lies entirely below zero; across seven uncorrected comparisons this reads as no useful signal, not as harm.

Table 1. Attributed importance and unique predictive contribution of each predictor block

| Predictor block     | SHAP share (%) | AUC without | $\Delta\text{AUC}$ [95% CI] | Above zero |
|---------------------|----------------|-------------|-----------------------------|------------|
| Education provision | 30.1           | 0.818       | +0.020 [+0.005, +0.036]     | yes        |

Table 1. (continued)

|                        |      |       |                         |    |
|------------------------|------|-------|-------------------------|----|
| Entrepreneurial spirit | 27.9 | 0.826 | +0.012 [-0.005, +0.030] | no |
| Access to finance      | 20.9 | 0.837 | +0.001 [-0.010, +0.012] | no |
| Individual/family      | 10.2 | 0.840 | -0.002 [-0.009, +0.004] | no |
| Policy support         | 7.1  | 0.843 | -0.004 [-0.011, +0.002] | no |
| Policy awareness       | 2.1  | 0.844 | -0.006 [-0.011, -0.001] | no |
| Role models            | 1.8  | 0.843 | -0.004 [-0.009, +0.000] | no |

Shares are of total mean absolute SHAP value; intervals are paired bootstraps, uncorrected for seven comparisons.

### 4.3. Feature-level attribution and robustness

Figure 1 shows the twelve features with the largest absolute attribution. The link between courses and real settings leads at 0.060, ahead of lecturers' entrepreneurial experience at 0.036, the perceived surmountability of the financing threshold at 0.032 and the ability to use financing instruments at 0.031. Three provision items enter the top ten, the third being "the university is the best place for systematic entrepreneurship training" at 0.020; all three have positive Spearman correlations of 0.894, 0.839 and 0.834, with mean signed Shapley values of 0.006, 0.004 and 0.004. Campus climate items rank lower.

Gender ranks twelfth at 0.013, with a Spearman correlation of  $-0.865$ . The group comparison is the more direct evidence: on the held-out set the mean predicted probability is 0.459 for men with an observed positive rate of 0.455, against 0.376 and 0.333 for women, so the model reproduces a group gap already present in the sample.

Under leave-one-major-group-out validation the six discipline groups return 0.843, 0.842, 0.831, 0.811, 0.787 and 0.737 (mean 0.808). The weakest point estimate is 0.737 (0.634 to 0.825), while the smallest group carries the widest interval, 0.695 to 0.915; two lower bounds therefore fall below 0.70, the spread is 0.106, and between-group differences were not tested. Discrimination holds under the alternative thresholds, 0.855 at 3.5 and 0.893 at 4.25, and a random-forest regressor reusing the classifier's selected hyper-parameters gives a coefficient of determination of 0.306 and a root mean square error of 0.724 on the continuous mean. These checks re-run predictive performance only; the attribution ranking and ablation were not recomputed under the alternative outcomes.

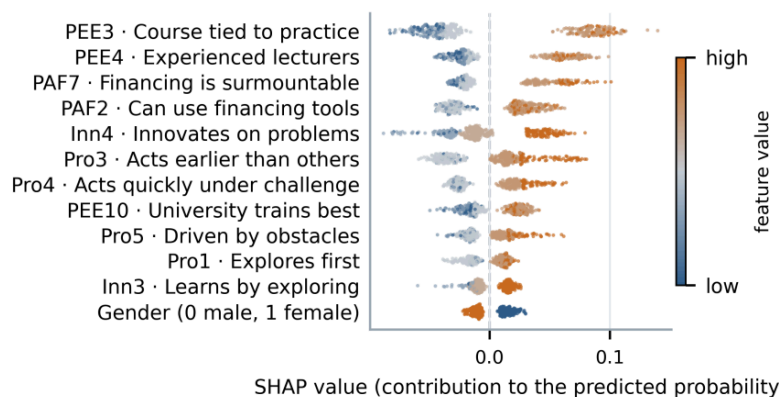


Figure 1. Global SHAP attribution, twelve most influential features

## 5. Discussion

Entrepreneurship education provision holds both the highest attribution share and the only ablation interval entirely above zero. Entrepreneurial spirit accounts for 27.9% of attribution yet shows no detected independent contribution, and individual and family background shows neither a contribution nor any handle for teaching practice. Provision alone combines a high share with a detectable predictive loss on removal, which makes it the first candidate for an intervention study.

The divergence between attribution share and ablation is itself a methodological point. Perceived access to finance holds 20.9% of attribution and two of the top four items, so ranking by importance alone would put financing support first, yet removing the block barely changes the prediction. Global importance measures association strength while ablation measures the unique information a block carries, so an evaluation intended to generate allocation hypotheses should report both. That the ensemble's accuracy gain is not distinguishable here points the same way: progress lies in the interpretation layer, not in more complex models.

The provision block also differentiates internally: its three top-ten items point to course-to-practice linkage, lecturers' experience, and recognition of the university as the site of systematic training, whereas campus climate items attract markedly lower attribution. This agrees with research stressing that curriculum, practice platforms and institutional guarantees must work together [11], and locates the weight within it: these perceived provision items received larger mean absolute SHAP values than the climate items.

Three boundaries apply. First, the data are a single cross-sectional self-report survey in which perceived provision and intention were measured at the same time by the same instrument, so reverse causality and unobserved confounding cannot be ruled out and the results are predictive associations, not causal effects of investment. Second, the sample comes from one province; cross-discipline validation puts every group estimate above 0.70, but extrapolation across regions requires new samples. Third, the outcome is intention as a proximal indicator, and whether it predicts actual venture behaviour can only be answered with longitudinal data tracking that behaviour.

## 6. Conclusion

The random forest, selected solely on training-set cross-validation, attains 0.838 on the held-out set, with an interval for its difference from logistic regression that includes zero, and every discipline-specific estimate exceeds 0.70, though they vary and are imprecise in the smaller groups. The methodological contribution is to place global attribution beside leave-one-block-out ablation, separating importance as association from importance as irreplaceable information; perceived access to finance shows why an attribution ranking alone does not establish incremental predictive value.

In practice, provision is the only block whose removal produces a detectable predictive loss, and within it course-to-practice linkage and lecturers' entrepreneurial experience discriminate most sharply, making them priority candidates for intervention studies with longitudinal, cross-regional data.

## Declaration of interest

The author declares no competing interests and confirms sole responsibility for the design, analysis, interpretation and preparation of this work.

## References

- [1] Guanin-Fajardo, J. H., Guaña-Moya, J., & Casillas, J. (2024). Predicting academic success of college students using machine learning techniques. *Data*, 9(4), 60. <https://doi.org/10.3390/data9040060>
- [2] Dai, W., Lin, J., Jin, F. J. Y., Tsai, Y. S., Srivastava, N., Le Bodic, P., Gašević, D., & Chen, G. (2025). Learning analytics for early identification of at-risk students and feedback intervention. *Journal of Learning Analytics*, 12(3), 102-125. <https://doi.org/10.18608/jla.2025.8735>
- [3] Zhang, L., Fu, Y., Wei, Y., Chen, H., Xia, C., & Cai, Z. (2022). Predicting entrepreneurial intention of students: Kernel extreme learning machine with boosted crow search algorithm. *Applied Sciences*, 12(14), 6907. <https://doi.org/10.3390/app12146907>
- [4] Aristizábal, J. M., Tarapuez, E., & Astudillo, C. A. (2025). Predicting entrepreneurial intention in Colombian academics: A machine learning approach. *Journal of Entrepreneurship in Emerging Economies*, 17(2), 260-288. <https://doi.org/10.1108/jeee-04-2023-0141>
- [5] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- [6] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56-67. <https://doi.org/10.1038/s42256-019-0138-9>
- [7] Kostopoulos, G., Davrazos, G., & Kotsiantis, S. (2024). Explainable artificial intelligence-based decision support systems: A recent review. *Electronics*, 13(14), 2842. <https://doi.org/10.3390/electronics13142842>
- [8] Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- [9] Silesky-Gonzalez, E., Lezcano-Calderon, Y., & Mora-Cruz, A. (2025). Effects of education for entrepreneurship and entrepreneurial intention in university students. *International Entrepreneurship and Management Journal*, 21(1), 26. <https://doi.org/10.1007/s11365-024-01039-4>
- [10] Murad, M., Othman, S. B., & Kamarudin, M. A. I. B. (2024). Entrepreneurial university support and entrepreneurial career: The directions for university policy to influence students' entrepreneurial intention and behavior. *Journal of Entrepreneurship and Public Policy*, 13(3), 441-467. <https://doi.org/10.1108/jep-08-2023-0082>
- [11] Lv, M., Zhang, H., Georgescu, P., Li, T., & Zhang, B. (2022). Improving education for innovation and entrepreneurship in Chinese technical universities: A quest for building a sustainable framework. *Sustainability*, 14(2), 595. <https://doi.org/10.3390/su14020595>
- [12] Xiao, B., Wang, X., Liu, Q., & Zhang, P. (2026). Linking entrepreneurial spirit and entrepreneurial environment perception to entrepreneurial intention: Moderation by entrepreneurial role models. *PLOS ONE*, 21(7), e0352807. <https://doi.org/10.1371/journal.pone.0352807>
- [13] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- [14] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM. <https://doi.org/10.1145/2939672.2939785>