RESEARCH

From Class to Career: Predicting Student Employability and Personalizing Advice Through Machine Learning Using Mock Interview Results

Ian Joseph CoKehyeng† *, Rex GregorLaylo†, John Cris Orenday†, Miguel Lorenzo Singian† and Gregory Patrick Uy†
Aboitiz School of Innovation, Technology, and Entrepreneurship, Asian Institute of Management, Makati City, Philippines

*Correspondence: icokehyeng.MSDS2024@aim.edu
†Equal contributor

Abstract

This study leverages machine learning techniques to predict the employability of students based on mock interview results from the Technological Institute of the Philippines (TIP). The dataset comprises 2,982 entries, assessing students on various attributes like General Appearance, Communication Skill, and Mental Alertness. It also contains their On-the-Job Training (OJT) employability assessments of either "Employable" or "Less Employable". Our objective was to identify key factors influencing employability to provide individualized advice for improving professional qualities.

Multiple machine learning (ML) models were evaluated across different resampling techniques. The XGBClassifier, with specific hyperparameters (`learning_rate=0.2`, `max_depth=6`, and `n_estimators=200`), and utilizing Random Undersampling, emerged as the optimal model, achieving 85% accuracy and an AUPRC of 0.96.

Further analysis employed Diverse Counterfactual Explanations (DiCE) and SHapley Additive exPlanations (SHAP) to dissect local feature importance and generate counterfactual scenarios, offering individualized pathways for improving student employability. SHAP global feature analysis revealed that Mental Alertness stood out as the most significant factor, accounting for an average impact of 12% on a TIP student's likelihood of success, either enhancing or diminishing their prospects.

This study also highlights the relatively low barrier to entry of an ML-powered mock interview employability screening, achieving 80% accuracy with as few as 500 mock interview results. This suggests that educational institutions can swiftly adopt such a predictive framework to enhance their mock interview programs, significantly boosting students' job prospects.

Keywords: student employability; mock interview; machine learning; shap; dice

Highlights

- Mental Alertness was the top global feature predicting employability of TIP students based on SHAP value.
- DiCE counterfactuals provide individualized insights and recommendations for students' areas of improvement.
- 80% accuracy was achieved with a random sample of only 500 data points, indicating good deployment feasibility.

1 Introduction

The Commission on Higher Education (CHED) in the Philippines mandates Internship or On-the-Job Training (OJT) programs in most undergraduate programs to bridge the gap between theoretical knowledge and practical experience [1]. Complementing these OJT programs, many universities taken to holding mock interviews [2]. These mock interviews allow institutions to assess student employability, identify strengths and weaknesses, and tailor training programs accordingly through targeted interventions. Through analyzing mock interview data and highlighting students' strengths and weaknesses, universities better bridge the gap between education and employment.

Informed by this, our study aimed to use mock interview results to train an ML model to predict student employability, and provide better individualized advice by leveraging Explainable AI in SHAP explanations [3] and DiCE counterfactuals [4].

Related Literature and Contribution

Casuat et al. [5] from the Technological Institute of the Philippines (TIP) put together the seminal student employability dataset in 2020 from 2,982 mock job interview results and other various assessments between 2015 and 2018. Preprocssing the data through SMOTE and training a support vector machine, they yielded an test accuracy of 92.2%.

Vo et al. [6] in 2023 used the same dataset to demonstrate the novel OPT-BAG (OPTimisation of BAGging classifiers) algorithm, achieving a 91.15% test accuracy.

Our approach differs from that of both Casuat et. al [5] and Vo et al. [6] was to exclude the feature *student performance rating* in order to train a model that robustly predicts student employability using the subjective mock interview result features only, without any other assessments. In addition to this, we avoid using SMOTE oversampling or any interpolation-based oversampling technique, since the dataset itself consists of ordinal values, and putting values between them may alter the distribution. Our method also leverages interpretability techniques to create a proof-of-concept report that may be used by university counselors to provide individualized employablity interventions.

2 Data and Methods

2.1 Data Source

The dataset consisted of mock job interview results and OJT employability assessments of 2,982 students from the TIP between 2015 and 2018, compiled and used in

Casuat et al. [5]. The version obtained on Kaggle [7] did not include General Point Grade, Students' Program, and Student Number.

Student Performance Rating was dropped from the dataset since we were most interested in predicting the OJT employability assessment based on the qualitative mock interview results, rather than other assessments.

The features in the raw dataset are shown below, in Table 1:

Table 1 Data Description

Feature Name	Data Type	Remarks
Name of Student	String	Anonymized, ID Feature
General Appearance	Integer	Ordinal 1-5; Training Feature
Manner of Speaking	Integer	Ordinal 1-5; Training Feature
Physical Condition	Integer	Ordinal 1-5; Training Feature
Mental Alertness	Integer	Ordinal 1-5; Training Feature
Self-Confidence	Integer	Ordinal 1-5; Training Feature
Ability to Present Ideas	Integer	Ordinal 1-5; Training Feature
Communication Skills	Integer	Training Feature
Student Performance Rating	Integer	Ordinal 1-5; Dropped
Employability Class	String	Binary "Employable, Less Employable"; Target Feature

2.2 Models and Resampling Selection

Several resampling techniques and machine learning models were tested together in a grid search format, to find the optimal pair with the best model accuracy.

Resampling Methods

Random Oversampling, Random Undersampling, and TomekLinks were each tested as preprocessing steps for their effect on their final model accuracy. We intentionally did not include interpolation-based oversampling techniques such as SMOTE in our search, since, as mentioned, the dataset consists of ordinal data and interpolating values between them would change the nature of the data.

Models

K-Nearest Neighbors, Logistic Regression (L2), XGBoost, Random Forest, Gradient Boosting Machine, and Decision Tree were each hyper-parameter tuned using GridSearchCV and included in our search for best overall model accuracy.

3 Results

The final model was determined to be `XGBClassifier` with the following hyperparameters: `learning_rate=0.2`, `max_depth=6`, and `n_estimators=200`. It was fitted on randomly undersampled training data. This choice was made due to its high model accuracy of 86% and Area Under Precision-Recall Curve (AUPRC) of 96%. This model will be used in the subsequent Interpretability section to understand both local and global feature importances.

Table 2 Model Performance with Different Resampling Techniques

Model	Train Acc. (%)	Test Acc. (%)	AUC	PR-AUC	F1	Precision	Recall
GBM-ROS	85.51	85.39	94.80	96.14	87.60	86.32	88.91
RF-RUS	85.76	85.39	94.74	96.09	87.49	86.99	87.99
GBM-RUS	85.76	85.39	94.73	96.06	87.49	86.99	87.99
XGBClassifier-RUS	85.75	85.39	94.65	96.06	87.49	86.99	87.99
XGBClassifier-ROS	85.29	85.39	94.36	95.80	87.60	86.32	88.91
RF-ROS	85.53	84.85	94.96	96.23	87.32	84.93	89.84
XGBClassifier-Original	86.13	84.72	94.21	95.69	87.61	82.75	93.07
XGBClassifier-TOMEK	86.13	84.72	94.21	95.69	87.61	82.75	93.07
RF-Original	86.22	84.45	94.98	96.26	87.14	83.80	90.76
RF-TOMEK	86.22	84.45	94.94	96.24	87.14	83.80	90.76
GBM-Original	86.14	84.45	94.23	95.66	87.25	83.23	91.69
GBM-TOMEK	86.14	84.45	94.23	95.66	87.25	83.23	91.69
KNN-Original	85.32	83.91	92.34	92.11	87.04	81.74	93.07
KNN-TOMEK	85.32	83.91	92.34	92.11	87.04	81.74	93.07
KNN-ROS	83.50	83.11	91.54	90.81	86.39	81.14	92.38

As shown in Table 2, the optimal model and its hyperparameter-tuned settings were identified for the final model, which was then fit on resampled training data. This final model will be passed to the DiCE model and SHAP explainer for interpretability.

3.1 Global Interpretability using SHAP

SHAP [3] is recommended for global interpretability because it offers a unique blend of individual prediction explanations and aggregate insights. By calculating the contribution of each feature to every prediction, SHAP allows for a nuanced understanding of model behavior on a global scale. It distinguishes itself by ensuring that even in complex models, where interactions and correlations among features can obscure their effects, the impact of each feature is quantified and understood. This makes SHAP a powerful tool for unveiling the underlying patterns and influences within a model, facilitating more informed decision-making and model improvement efforts.

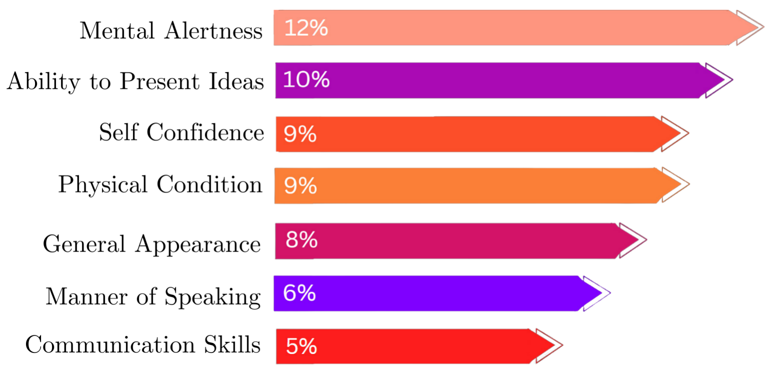

Figure 1 SHAP Global Importance

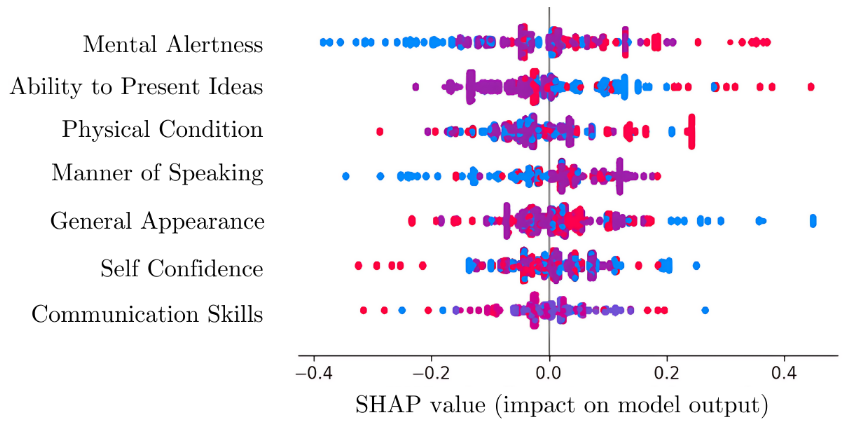

Figure 2 SHAP Beeswarm Plot

Using this approach, Mental Alertness was identified as the top feature, contributing 12% on average to each student's predicted probability of success as seen in Figure 1.

3.2 Beeswarm Plot

The Beeswarm Plot in Figure 2 reveals a minimal correlation between feature values and their corresponding SHAP values, which might initially seem perplexing. In simpler terms, this indicates that the values of all other features in the dataset influence the SHAP value assigned to a specific feature. For instance, a rating of 4 in Mental Alertness could result in a negative SHAP value if all other features are rated higher, at 5. On the other hand, the same rating of 4 in Mental Alertness could lead to a positive SHAP value if the ratings for all other features are lower at 3. This illustrates how SHAP values reflect the contextual importance of a feature within a particular data entry.

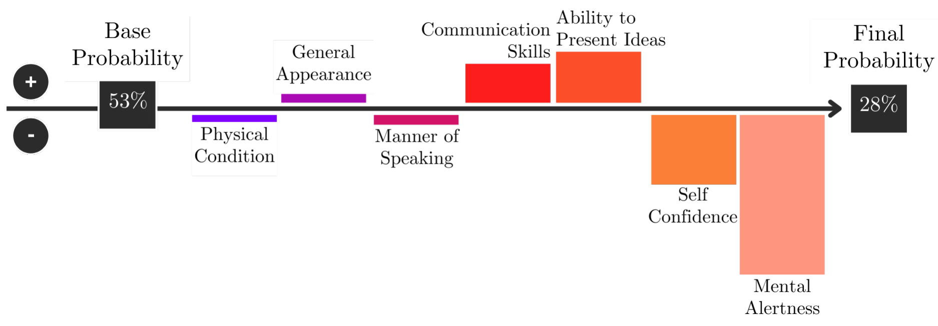

Figure 3 Student 15 SHAP Local Importance

3.3 Understanding the Prediction

The SHAP Waterfall plot in Figure 3 illustrates how Student 15's feature values contribute positively or negatively to the student's probability of success. There are three main key takeaways:

1 The baseline probability for all students is 53%.
2 However, Student 15's feature values reduce this to a final probability of 28%.
3 Although the student scored 3s in all features, some features, like Mental Alertness, have a more significant impact than others.

3.4 Identifying Paths to Improvement

DiCE [4] generates counterfactual explanations by optimizing for minimal changes to the input features that would change the model's prediction to a desired outcome. These counterfactuals are as similar to the current state as possible while still achieving the opposite predicted class.

Table 3 Student 15 Counterfactuals

	Current	**Counterfactuals**		
		1	**2**	**3**
Manner of Speaking	3	-	-	-
Ability to Present Ideas	3	-	-	-
General Appearance	3	4	-	-
Communication Skills	3	-	-	-
Self Confidence	3	-	-	5
Physical Condition	3	-	-	-
Mental Alertness	3	-	4	-
Outcome	Less Employable		Employable	

In the context of Student 15 in Figure 3, counterfactual explanations were generated to provide possible recommended paths for improvement to become predicted as Employable. In many scenarios, just improving one feature by 1 was enough

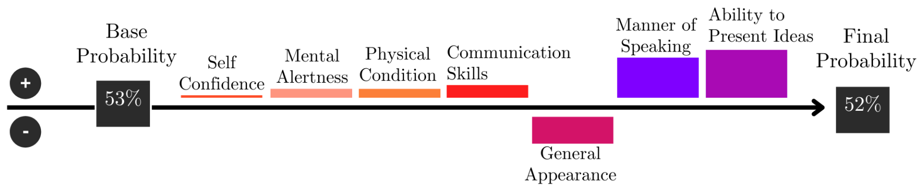

Figure 4 Student 8 SHAP Local Importance

to reach the opposite predicted class (Table 3). This offers Student 15 options if Student 15 struggles to improve in some areas over others.

As counterfactuals are from Less Employable to Employable, permitted ranges were restricted to only increase per feature.

3.5 Understanding the Prediction

The SHAP Waterfall plot in Figure 4 illustrates how Student 8's feature values contribute positively or negatively to the student's probability of success. There are two main key takeaways:

1 Despite an optimistic prediction, the predicted probability is only 53%, which puts Student 8 at high risk for failure.

2 The Ability to Present Ideas, while equal to all other feature values, drags down Student 8's probability, indicating that this feature is relative and may require a higher standard by employers than other features.

Table 4 Student 8 Counterfactuals

	Current	**Counterfactuals**		
		1	**2**	**3**
Manner of Speaking	4	-	-	-
Ability to Present Ideas	4	-	-	3
General Appearance	4	-	-	2
Communication Skills	4	-	3	-
Self Confidence	4	-	-	-
Physical Condition	4	3	-	-
Mental Alertness	4	-	-	2
Outcome	Employable	Less Employable		

3.6 Identifying at Risk Areas

Counterfactual explanations indicate the areas that can alter the predicted class if minimally changed. In the context of Student 8, it was identified that if even one of "Physical Condition" or "Communication Skills" fell slightly, this would be enough to become predicted as "Less Employable" as seen in Table 4. Further counterfac-

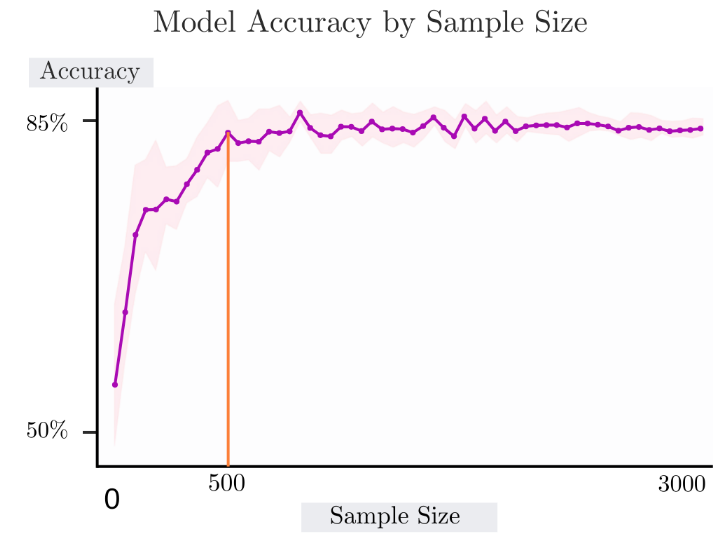

Figure 5 Module Accuracy by Sample Size

tuals would require more significant declines in features, indicating that Physical Condition and Communication should be the key focus areas to maintain predicted Employability. As counterfactuals are from Employable to Less Employable, permitted ranges were restricted to only decrease per feature.

3.7 Deployment Feasibility

Feasibility of deployment was considered an essential aspect of this model. Ideally, a model should achieve high accuracy without requiring a large amount of data, as data collection is often arduous, lengthy, and costly. This is especially true of collecting mock interviews. Fortunately, we found that as few as 500 mock interviews could already achieve an 80% test accuracy (See Figure 5). It means that this model can be deployed and made functional within one or two years provided 250-500 students per year are interviewed, making this accessible for many medium-large sized universities to pursue and implement.

4 Conclusion

This study successfully developed a machine learning model to predict student employability based on mock interview results. The XGBoost classifier with random undersampling emerged as the optimal model, achieving 85% accuracy and an impressive 0.96 AUPRC score. The model's performance showcases the potential of leveraging machine learning techniques to assess and enhance students' professional readiness.

Through SHAP analysis, this study identified Mental Alertness as the most significant factor influencing employability predictions among TIP students, contributing an average of 12% to each student's likelihood of success. This finding underscores the importance of developing students' critical thinking ability and responding effectively in interview situations.

Moreover, the application of DiCE generated individualized recommendations for improving specific areas and providing tailored guidance to students based on their strengths and weaknesses.

Remarkably, the model demonstrated robustness and ease of deployment, achieving 80% accuracy with as few as 500 mock interview results. It highlights educational institutions' potential to adopt such a predictive framework, fostering a more efficient and practical approach to bridging the gap between academia and industry.

5 Recommendations

We recommended that universities explore integrating mock interview simulations and assessments into their curricula based on their demonstrated predictive power, adapting and tailoring our methodology to their own institutions, programs, and preferred interview questions. Using our approach, they would leverage interpretability methods to offer tailored advice to students. Doing so, they may streamline the process of identifying areas for improvement and provide targeted interventions, ultimately ensuring that graduates are better equipped to navigate the competitive job market successfully.

References

1. Commission on Higher Education: Revised Guidelines for Student Internship Program in the Philippines (SIPP) for All Programs (CMO No. 104, S. 2017). Retrieved March 10, 2024. https://ched.gov.ph/wp-content/uploads/2018/03/CMO-NO.-104-S.-2017.pdf
2. Huss, R., Jhileek, T., Butler, J.: Mock Interviews in the Workplace: Giving Interns the Skills They Need for Success. The University of West Georgia and Academy for Advanced Studies (2017)
3. Lipovetsky, S., Conklin, M.: Analysis of regression in game theory approach. Applied Stochastic Models in Business and Industry **17**(4) (2001)
4. Mothlilal, R., Sharma, A., Tan, C.: Explaining machine learning classifiers through diverse counterfactual explanations. Conference on Fairness, Accountability, and Transparency, 607–617 (2020)
5. Casuat, C.D., Festijo, E.D., Alon, A.S.: Predicting Students' Employability using Support Vector Machine: A SMOTE-Optimized Machine Learning System. International Journal of Emerging Technologies in Engineering Research (IJETER) **10**(2) (2020)
6. Vo, M.-T., Nguyen, T., Le, T.: OPT-BAG Model for Predicting Student Employability. Computers, Materials & Continua **76**(2) (2023)
7. Kaggle - Students' Employability Dataset - Philippines. Retrieved February 23, 2024. https://www.kaggle.com/datasets/anashamoutni/students-employability-dataset