Identifying factors affecting placement status of engineering students using explainable machine learning

[1,*]Pawan Kumar , [2]Manmohan Sharma, [3]Sandeep Kaur, [4]Sanjay Sood

[1,2]School of Computer Science and Engineering, Lovely Professional University, India

[3]Guru Nanak College for Women, Banga, Punjab, India

[4]Division of Student Career Services, Lovely Professional University, India

Abstract:
Machine learning (ML) models are increasing their penetration in different problem domains due to their increasing accuracy. Majority of the advanced ML-based stand high in predictive accuracy but are wanting in terms of explaining their outcomes. This research study has two objectives. First, to develop predictive models that can predict the placement status of undergraduate engineering students studying in India. Second, to understand the behavior of these ML-based models in terms of most important features affecting the prediction outcome. The outcomes of such study are of importance to any university as ability to anticipate placement status of students can help in planning timely interventions. It is concluded that 'CurrentCGPA' and 'AttendancePercentage' of a student plays an important role towards getting placed. Moreover, understanding of the behavior of a model help building trust in the model and provide multiple insights of importance to different stake holders.

Keywords: machine learning, binary classification, placement status, random forests, gradient boosting, human interpretability

1. Introduction

ML is a sub-field of Artificial Intelligence where objective is to learn from experiences and use this learning in future unseen situations [1]. In the recent years, ML has consistently been among the top buzzwords in the computer science field. Researchers have proposed a plethora of ML algorithms in literature and new additions are being made every year to the kitty of ML algorithms. These algorithms aim to learn from experiences (in the form of dataset) and the outcome is a formal mathematical model. This model is then used to make predictions for the target variable in new records. Accuracy of predictions is the most commonly used metric to perform evaluation of a ML-based model. Over the years, ML models have improved in terms of their prediction accuracy.

Some of these models like linear classifiers and decision trees are easily interpretable. Still, most of the advanced ML algorithms are complex. It is difficult for layman human users to understand the reasoning employed by the model while making predictions [2]. This lack of ease in interpreting the prediction outcomes act as a hurdle towards trusting a model and its subsequent deployment [3]. Moreover, ability to interpret outcomes of a ML-based model brings advantages like model debugging, trust, fairness and discovering new knowledge [4]. Due to the growing interest in making ML-based models interpretable to their human stake holders, different tools are coming up with in-built capabilities for human interpretability. Recently, several packages like 'lime', 'imp' and 'DALEX' have been developed for understanding behavior of your ML models.

This study has two primary objectives. First, to explore the applicability of machine learning algorithms in predicting placement status of students. Second, to investigate the behavior of these ML models using 'DALEX' package [5]. The results of such study can help academic institutions (i) Developing a predictive model to predict placement status (ii) Extracting important factors that most affect chances of getting placed. Performance of students is also an important criterion when educational institutes go for accreditation and rankings [6].

Organization: Section 2 gives a brief summary of the literature review; Section 3 gives a brief introduction to the methods used; Section 4 gives details of the experimental setup and experiments planned; Section 5 compiles the outcomes of the experiments; Section 6 gives conclusion and future lines of research for this work.

2. Literature Review

Table 1 gives a brief summary of a few approaches proposed in this problem domain covering objective, dataset used, methods employed and possible classes of the target variable. Most of the approaches are formulating it as a binary classification problem.

Table 1. Summary of related work

Objective	Dataset	Methods	Target Variable
Predicting placement status [7]	Data of 65 MCA students at VBS Purvanchal Univ	NB, C4.5, Multi-layer Perceptron	Placed = {Yes, No}
Predicting placement status [8]	261engineering students	ID3, RF	Placed = {Yes, No}
Predicting placement status; Comparing performance on different tools [99]	Final year B. Tech students	J48, NB, RF in Weka; LR and RPART using R	Placed = {Placed, Not Placed}
A system that recommend suitable type of placement for a student [10]	289 UG engineering students	DT, LR	Placement Status = {Dream Company, Core Company, Mass Recruiters, Not Eligible, Not Interested}
Predicting placement status [11]	1400 MCA students of different colleges	J48, RF, SMO, NB, Multi-layer Perceptron	Placement Status = {Yes, No}

3. Methods

3.1 Classification algorithms

The following three ML algorithms are explored for developing models to predict academic performance of international students.

(i) Regularized logistic regression: Logistic regression is a derivation from linear regression. Linear regression is used for predicting a continuous variable whereas logistic regression is used for predicting a categorical variable. Regularization is applied to a ML model to improve its generalization. It helps avoid overfitting of the model [12].

(ii) Random Forests: Random forest are a member of tree-based ML algorithms. The basic idea is to construct a number of small decision trees using different subset of data. To predict for a given instance, each tree is used and outcome is decided using majority voting [13].

(iii) Gradient boosting: It is an ensemble approach used in classification as well as regression problems [14]. The idea is to employ a number of decision trees and build a model in a sequential manner. At every stage, misclassified examples are given more weight so that model sequentially keep learning from mistakes.

3.2 Model interpretability techniques

To understand the model behavior, the following interpretability techniques are used:

(i) Residual diagnostics: This technique refers to measure residuals of predicted vs actuals to find out where models deviate. By plotting the residual quantiles using box plots, absolute residual values across different models can be visualized.

(ii) Variable Importance: The most intuitive question in ML model interpretability is to identify which features are contributing most towards the prediction outcome. In order to compare variable importance across different ML models, a model-agnostic approach is needed. Model-agnostic approach refers to methods that are applicable to any ML model. The idea is to measure drop in model performance when values of features are permuted.

(iii) Predictor-Response relationships: After the identification of most influential predictors, the next objective is to check how these important variables are related to response variable individually. It helps in exploring whether each model is capturing similar type of relationship or not.

(iv) Local Interpretation: Apart from understanding global behavior of a model, there is need to understand prediction outcome for an individual instance also [15]. Understanding how each model is using predictor variables for an individual instance enables building trust in a ML model.

The above techniques are implemented using ‘DALEX’ package in R [16].

4. Experimental Setup

4.1 Dataset

The dataset consisted of 650 undergraduate engineering students of Computer Science and Engineering 2019 passing batch at a University in North India. The dataset consisted of details related to student demographics, academic history, current academics, Professional Enhancement Programme activities, and performance in Benchmark tests. Out of total 650 students, 499 are labeled as ‘Placed’ and 151 as ‘Not Placed’, giving a baseline accuracy of 76.8%. Baseline accuracy is the accuracy of a model when majority class is predicted for each instance. Table 1 mentions the features in the dataset along with type and a brief description. ‘Placed’ is the binary target variable with status as ‘Placed’ or ‘Not Placed’. Table 2 mentions the features in the dataset along with type and a brief description.

Table2. Dataset description

Feature	Data Type	Description
XIIthPercentage	Numeric	Percentage in XIIth class
XthPercentage	Numeric	Percentage in Xth class
ComputerProgrammingScore	Numeric	AMCAT Examination
ComputerScienceScore	Numeric	AMCAT Examination
EnglishComprehensionScore	Numeric	AMCAT Examination
LogicalAbilityScore	Numeric	AMCAT Examination
QuantitativeAbility	Numeric	AMCAT Examination
AttendancePercentage	Integer	Attendance in classes
Category	Factor	Categorization of student based on academics
CurrentCGPA	Numeric	Current CGPA of the student
DutyLeaves	Integer	No of duty leaves earned
IsHons	Logical	Is student from Honors programme? Yes or No
ScholarshipStatus	Logical	Whether on scholarship? Yes or No

StandingBacklog	Integer	Number of Fail grades
StandingReappear	Integer	Number of reappear grades
StudentStatus	Factor	PC = Program complete, I = Program Incomplete
Gender	Factor	Male or Female
Placed	Factor	Placement Status

4.2 Experiments

The following experiments are conducted:

I. Learning a classification model to predict placement status of students

II. Evaluating comparative performance of each model

III. Investigating factors affecting placement status of students

5. Results and discussion

5.1 Performance of ML models: Table 3 gives AUC values for each of the three ML algorithms explored. AUC values lie in the range of .80 to .85 which indicates that these models are very close in terms of performance. GBM has highest AUC value followed by GLM and then RF.

Table 3. Model performance

Algorithm	AUC Values
Elastic Net (GLM)	0.8480471
Random Forest (RF)	0.8041734
Gradient Boosting (GBM)	0.8485821

5.2 Understanding model behavior: In order to compare performance of the above three models, only one metric AUC value is not enough.

Residual diagnostics: Residuals are difference between the probability of 'Not Placed' to the binary values (1- Not Placed, 0 -Placed). Looking at the boxplots in Figure 1, GBM have lower median of residuals than GLM and RF. As mentioned earlier, GBM gave highest AUC value also. So, GBM model seems performing better.

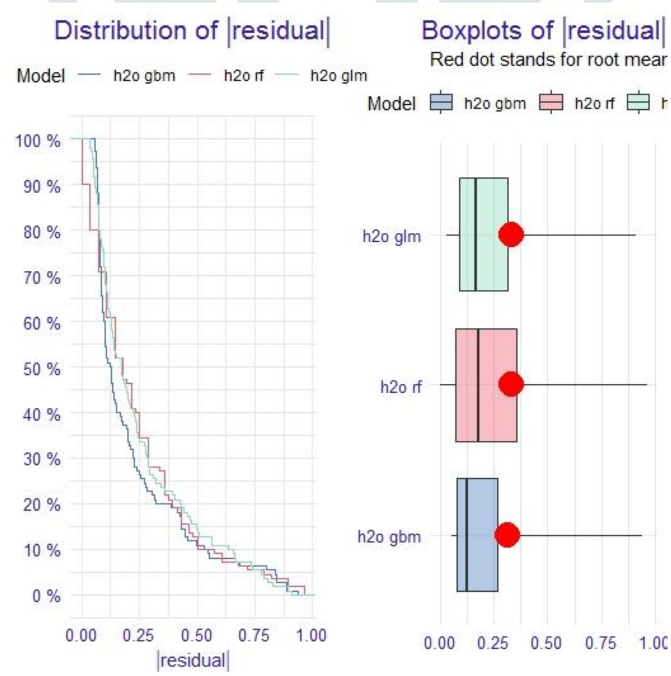

Figure 1. Residual diagnostics

Variable Importance Plot: Taking root mean square error (RMSE), the figure 2 shows variable importance for each of the three models. First, the shift of edge on x-axis indicates the lowest RMSE for GBM model. Second, the variables that are most significant consistently across different models include 'CurrentCGPA' and 'Category'. 'AttendancePercentage' and 'XIIthPercentage' are other predictors treated important by GBM.

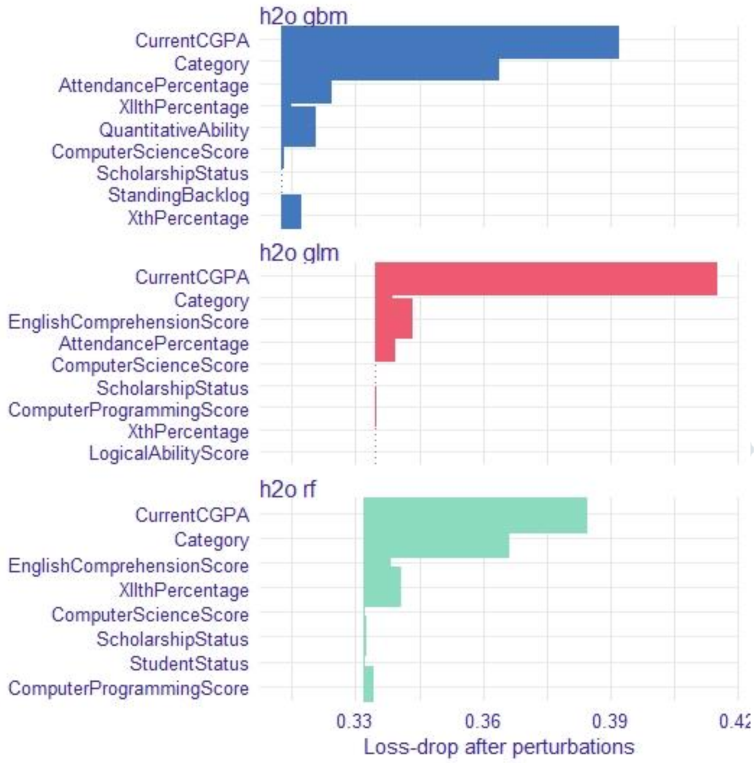

Figure 2. Variable importance

Predictor-Response relationship: Figure 3 shows response of the predictor 'CurrentCGPA' in each of the model. It is observed that each of the three models captures a non-linear relationship between 'Placed' and 'CurrentCGPA'. In case of GLM it is almost an exponential curve. This indicates that although GLM competes the other two in terms of AUC values but it may be using features in a biased manner.

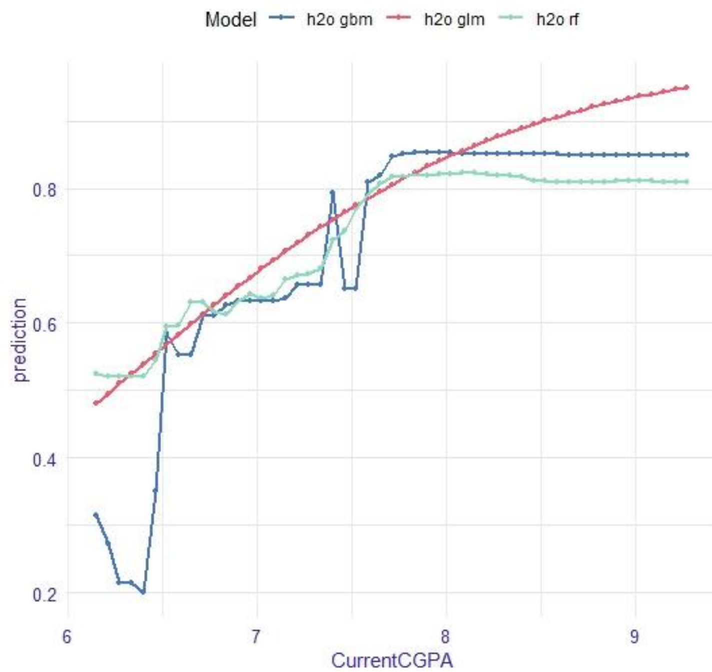

Figure 3. Predictor-response relationship

Local Interpretation: The objective is to compute contributions from individual features and compare across models. Figure 4 shows the local interpretation for an individual instance. There are three categories of contributions: Zero, Positive and Negative. A positive contribution indicates that the feature is contributing towards outcome by model for target variable. 'ScholarshipStatus', 'Gender' and 'IsHons' are having zero contribution. 'QuantitativeAbility' and 'ComputerProgrammingScore' are having positive contribution.

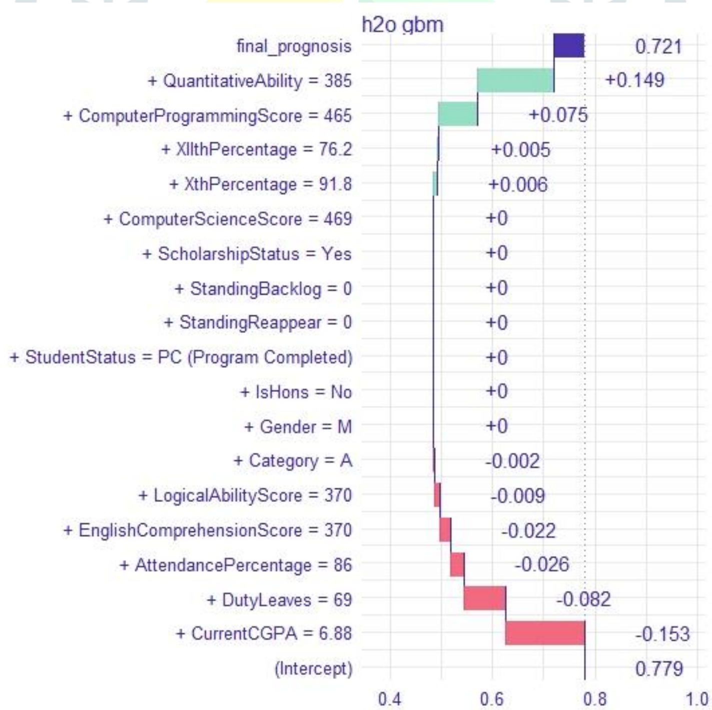

Figure 4. Local Interpretation for instance #1

6. Conclusion & Future Work

GBM gave the best performance among the three classifiers explored. Explanations using DALEX are helpful in understanding behavior of complex ML-based models. 'CurrentCGPA', 'AttendancePercentage', 'QuantitativeAbility' and 'ComputerProgrammingScore' are among most important factors that affect placement status of an undergraduate engineering student in computer science. Factors like 'Gender' are not found affecting placements much. This indicates that placement related processes are fair to each gender.

As a future work, additional features related to student's participation need be explored. These include participation of students in placement preparation activities, competitive events, hackathons, and feedback from faculty teaching these students. Another direction is to explore advanced ML algorithms to improve the prediction accuracy. Additional human interpretability techniques in machine learning need be explored in order to understand model behavior better.

7. Acknowledgements

The authors are thankful to the following at Lovely Professional University, Phagwara for allowing access to the dataset and using for this research study.

- Division of Research and Development
- Division of Student Career Services

8. References

[1] Tom M. Mitchell, "The Discipline of Machine Learning", 2006, CMU, http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf

[2] Kiri L. Wagstaff. (2012) "Machine Learning that Matters" Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK, 2012

[3] Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. (2016) "Why Should I Trust You? Explaining the Predictions of Any Classifier." KDD 2016, San Francisco, CA, USA, http://dx.doi.org/10.1145/2939672.2939778

[4] Zachary C. Lipton. (2016) "The Mythos of Model Interpretability", ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY, USA

[5] http://uc-r.github.io/dalex

[6] http://mhrd.gov.in/sites/upload_files/mhrd/files/document-reports/Manual_for_ Universities_23012013.pdf

[7] Pal, A. K., & Pal, S. (2013). Classification model of prediction for placement of students. International Journal of Modern Education and Computer Science, 5(11), 49.

[8] Manvitha, P., & Swaroopa, N. (2019). Campus Placement Prediction Using Supervised Machine Learning Techniques. International Journal of Applied Engineering Research, 14(9), 2188-2191.

[9] Rao, K. S., Swapna, N., & Kumar, P. P. (2018). Educational data mining for student placement prediction using machine learning algorithms. Int. J. Eng. Technol. Sci., 7(1.2), 43-46.

[10] Thangavel, S. K., Bkaratki, P. D., & Sankar, A. (2017, January). Student placement analyzer: A recommendation system using machine learning. In 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 1-5). IEEE.

[11] Mishra, T., Kumar, D., & Gupta, S. (2016). Students' Employability Prediction Model through Data Mining. International Journal of Applied Engineering Research, 11(4), 2275-2282.

[12] https://www.coursera.org/lecture/machine-learning/regularized-logistic-regression-4BHEy

[13] L. Breiman (2001), "Random forests.", Machine learning, 45(1): 5–32

[14] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

[14] David Baehrens, Timon Schroeter, Stefan Harmeling (2010) "How to Explain Individual Classification Decisions." Journal of Machine Learning Research 11 (2010) 1803-1831

[15] https://cran.r-project.org/bin/windows/base/