

# Predicting Higher Education Outcomes with Hyperbox Machine Learning: What Factors Influence Graduate Employability?

Kathleen B. Aviso<sup>a</sup>, Jose Isagani B. Janairo<sup>b</sup>, Rochelle Irene G. Lucas<sup>c</sup>, Michael Angelo B. Promentilla<sup>a</sup>, Derrick Ethelbhart C. Yu<sup>d</sup>, Raymond R. Tan<sup>a,\*</sup>

<sup>a</sup>Chemical Engineering Department, De La Salle University, 2401 Taft Avenue, 0922 Manila, Philippines

<sup>b</sup>Biology Department, De La Salle University, 2401 Taft Avenue, 0922 Manila, Philippines

<sup>c</sup>Department of English and Applied Linguistics, De La Salle University, 2401 Taft Avenue, 0922 Manila, Philippines

<sup>d</sup>Chemistry Department, De La Salle University, 2401 Taft Avenue, 0922 Manila, Philippines

[raymond.tan@dlsu.edu.ph](mailto:raymond.tan@dlsu.edu.ph)

A machine learning approach to predict university attributes that influence graduate employability is presented in this work. The machine learning technique used here is the hyperbox model, which is based on the principle of generating if / then rules to predict outcomes. The rule-based hyperbox model can be generated from empirical data using a mixed integer linear programming model. This machine learning approach is applied to the problem of predicting employability of chemical engineering graduates based on institutional attributes. The analysis shows that research intensity and quality do not necessarily result in high employability.

## 1. Introduction

Artificial intelligence (AI) and machine learning (ML) have become commonplace tools in the modern world. Different AI/ML techniques have been developed for use in many practical applications (Jordan and Mitchell, 2015). These techniques offer the prospect of improved decision-making in industry (Makridakis, 2017) and government (Sharma et al., 2020). AI/ML tools such as artificial neural networks (ANN), support vector machines (SVM), and Bayesian networks (BN) are powerful alternatives to classical statistical techniques. There is an emerging body of literature on the use of AI/ML for the analysis of education data. Recent work has been reported for both basic education (Abad and Chaparro Caso López, 2017) and higher education (Alyahyan and Düşteğör, 2020). Institutional attributes such as research intensity and research quality may indirectly influence graduate employability, which is an important measure of higher education effectiveness (Gyenes, 2019). However, a search of the Scopus database indicates that research on the use of AI/ML to analyse the effectiveness of chemical engineering education is rare. Application of AI/ML can lead to new insights to improve educational outcomes and career prospects.

In ML, prediction models are calibrated to fit a set of training data. The training procedure uses a second model to optimize fit. This process is analogous to using least squares optimization to generate a regression equation. Mathematical Programming (MP) models can be developed for training in ML. Mixed-integer linear programming (MILP) models are a useful class of models for this purpose. The use of continuous and integer variables allows MILP models to explore optimal and near-optimal solutions during supervised training (Voll et al., 2015). Notable examples of the use of MILP in ML have been reported. Iannarilli and Rubin (2003) developed an MILP-based feature selection technique for multiclass discrimination. Kim and Ryoo (2007) proposed an MILP model for non-linear data separation, while Bal and Örkücü (2011) developed a multi-class classification approach based on MILP. Xu et al. (2011) proposed a 0-1 programming model for attribute reduction. Yan and Ryoo (2017) also proposed a 0-1 multi-linear programming approach for pattern recognition. Rudin and Ertekin (2018) used an MILP approach for optimal rule generation, while Corrêa et al. (2019) proposed an approach for binary single-group classification. MILP models have also been combined with other ML techniques such as rough sets (Chang et al., 2019) and SVM (Labbé et al., 2019).

Xu and Papageorgiou (2009) developed a hyperbox-based ML approach for classification problems. Compared to many black-box ML techniques, this approach has the advantage of generating transparent results (Yang et al., 2015a), since the hyperboxes can be interpreted as if/then rules (Tan et al., 2020). The original algorithm required repeated re-optimization of MILP models for a satisfactory fit. Maskooki (2013) proposed an improved training algorithm with a reduced number of steps. An improved version of the hyperbox-based ML approach was also developed to account for Type I (false positive) and Type II (false negative) prediction errors (Yang et al., 2015a). Applications of hyperbox-based ML include business performance prediction (Xu and Papageorgiou, 2009), disease diagnosis (Yang et al., 2015b), and geological reservoir classification for CO<sub>2</sub> storage (Tan et al., 2020). The latter work improved the hyperbox-based ML approach for binary classification by (a) using concentric hyperboxes to separate positive and negative samples, and (b) enabling both rule simplification and attribute reduction.

In this paper, the hyperbox-based ML technique developed by Tan et al. (2020) is applied to the problem of predicting graduate employability based on high-level university attributes. The rest of this paper is organized as follows. Section 2 discusses the hyperbox concept. Section 3 gives the formulation of the MILP model for generating the hyperbox decision model. Section 4 applies the methodology to predicting employability of chemical engineering graduates in the UK. Section 5 gives the conclusions and prospects for future work.

## 2. Problem statement

The formal problem can be stated as follows:

- Given an information system with a set of criteria (I), which can be further divided into a set of condition attributes (A) and a binary decision set (D);
- Given a set of samples (J) which have known performance levels relative to the set of condition attributes (A) and final classification in the decision set (D);
- Given a predefined limit on the rate of false negatives;
- Given the minimum margin of separation between positive and negative samples for each given criterion;

The objective is to determine the boundaries of the hyperbox which will minimize the number of false positive results based on the training data set. The concept of hyperbox-based ML for binary classification is illustrated in Figure 1 for a problem with two criteria. The hyperboxes should enclose the positive samples while excluding the negative samples. Each hyperbox has an error margin; any sample that falls within this margin is considered as misclassified. It can be seen that one positive sample is misclassified, because it does not fall within any of the hyperboxes. One of the negative samples is also misclassified, since it is enclosed in the error margin of hyperbox 2.

The hyperboxes can also be interpreted as if/then rules (Tan et al., 2020). Hyperbox 1 can be specified using two-sided inequalities in both dimensions. Hyperbox 2 can be specified using one-sided inequalities (upper bounds only), and is said to be semi-infinite in both dimensions. Dimension reduction is illustrated by hyperbox 3, which can be projected to a one-dimensional space along x to make y irrelevant. The next section describes the MILP model (Tan et al., 2020) to generate the hyperboxes.

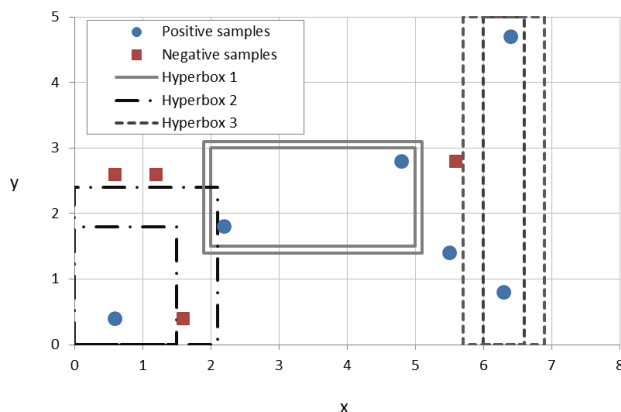


Figure 1: Illustration of hyperbox concept (adapted from Tan et al., 2020)

### 3. MILP for generating Hyperbox Decision Model

The objective of the optimization model is to minimize  $\alpha$ , as shown in Eq(1), where  $\alpha$  is the ratio of the total number of false positives ( $\sum_j FP_j$ ) to the total number of negative samples in the training data. Eq(2) indicates that the ratio of false negatives ( $\sum_j FN_j$ ) to the total number of positive samples in the training data ( $\otimes$ ) should be less than a predefined threshold  $\beta$ . Eq(3) and Eq(4) define  $\alpha$  and  $\otimes$ . A sample  $j$  is counted as a false positive if the hyperbox model classifies it as positive ( $c_j = 1$ ) when its true classification is negative ( $C_j^* = 0$ ), as indicated in Eq(5). The reverse is true for false negatives, as shown in Eq(6).

$$\alpha \quad (1)$$

$$\beta \leq \varepsilon \quad (2)$$

$$\alpha = \frac{\sum_j FP_j}{TN} \quad (3)$$

$$\beta = \frac{\sum_j FN_j}{TP} \quad (4)$$

$$FP_j \geq c_j - C_j^* \quad \forall j \quad (5)$$

$$FN_j \geq C_j^* - c_j \quad \forall j \quad (6)$$

Eq(7) and Eq(8) are meant to determine the outer boundaries of the hyperboxes, while Eq(9) and Eq(10) are meant to determine the inner boundaries.  $X_{ji}$  is the performance of training data point  $j$  in criterion  $i$ ;  $x_{ik}^L$  and  $x_{ik}^U$  are the lower and upper boundaries of hyperbox  $k$  for criterion  $i$ ;  $\Delta$  is the distance between the inner and outer limits of the hyperbox;  $b_{jk}$  is a binary variable which indicates if sample  $j$  is enclosed within hyperbox  $k$ ; and  $M$  is an arbitrary large number.

$$X_{ji} > x_{ik}^L - \Delta - M(1 - b_{jk}) \quad \forall j, i \quad (7)$$

$$X_{ji} < x_{ik}^U + \Delta + M(1 - b_{jk}) \quad \forall j, i \quad (8)$$

$$X_{ji} > x_{ik}^L - M(1 - b_{jk}) \quad \forall j, i \quad (9)$$

$$X_{ji} < x_{ik}^U + M(1 - b_{jk}) \quad \forall j, i \quad (10)$$

Eq(11) and Eq(12) account for the possibility that boundaries may not exist for criterion  $i$  in hyperbox  $k$ .  $Z_{ik}^L$  and  $Z_{ik}^U$  approximate  $-\infty$  and  $+\infty$  (or very large negative and positive values);  $b_{ik}^L$  and  $b_{ik}^U$  are binary variables which indicate whether the lower boundary ( $b_{ik}^L = 1$ ) or upper boundary ( $b_{ik}^U = 1$ ) for criterion  $i$  in hyperbox  $k$  disappears. Eq(13) and Eq(14) account for samples that lie outside of the outer boundaries of the hyperbox  $k$ ;  $q_{ijk}^L$  and  $q_{ijk}^U$  are binary variables indicating that the performance of sample  $i$  in criterion  $j$  is less than the lower limit ( $q_{ijk}^L = 1$ ) or more than the upper limit ( $q_{ijk}^U = 1$ ) set for hyperbox  $k$ . Eq(15) and Eq(16) define when a sample is enclosed in hyperbox  $k$ ;  $b_{jk}$  is a binary variable which takes a value of 1 when sample  $j$  is enclosed within hyperbox  $k$  and 0 otherwise. Eq(17) considers a sample to belong in the positive decision set ( $c_j = 1$ ) if it is enclosed by at least one hyperbox. Eq(18) defines all the binary variables in the model.

$$Z_{ik}^L - M(1 - b_{ik}^L) \leq x_{ik}^L \leq Z_{ik}^L + Mb_{ik}^L \quad \forall i, k \quad (11)$$

$$Z_{ik}^U - Mb_{ik}^U \leq x_{ik}^U \leq Z_{ik}^U + M(1 - b_{ik}^U) \quad \forall i, k \quad (12)$$

$$X_{ji} \leq x_{ik}^L - \Delta + M(1 - q_{ijk}^L) \quad \forall j, i \quad (13)$$

$$X_{ji} \geq x_{ik}^U + \Delta - M(1 - q_{ijk}^U) \quad \forall j, i \quad (14)$$

$$\sum_i (q_{ijk}^L + q_{ijk}^U) \leq M(1 - b_{jk}) \quad \forall j, k \quad (15)$$

$$\sum_i (q_{ijk}^L + q_{ijk}^U) \geq (1 - b_{jk}) \quad \forall j, k \quad (16)$$

$$\sum_k b_{jk} \leq Mc_j \quad \forall j, k \quad (17)$$

$$b_{jk}, b_{ik}^U, b_{ik}^L, Q_{ijk}^U, Q_{ijk}^L, C_j \in \{0,1\} \quad (18)$$

There are two phases in the developed methodology. The first phase is to split the data set into the training and validation sets. The training data set is then used to generate the hyperboxes. The second phase tests the resulting binary classifier on the validation data set.

#### 4. Case study: Predicting employability

This case study analyzes the employability of chemical engineering graduates of different UK universities. Industry expectations are an important measure of educational outcomes (Gyenes, 2019). The data used were reported by Gonzalez-Garay et al. (2019). Five university attributes are considered: (1) entry standards, (2) research intensity, (3) staff to student ratio, (4) budget per student, and (5) research quality. The decision attribute is employability, which is transformed into binary form using a threshold value of 80 %. The total number of samples was 25. From this data set, 15 samples were used as training data (Table 1), while the other 10 samples were used as validation data (Table 2).

The hyperbox-based ML approach was used to derive empirical if/then rules from Table 1. Supervised training was done by solving the MILP using the optimization software LINGO. Predictive performance was gauged using the validation data in Table 2. Different classifiers can also be generated by using integer cuts. Expert knowledge can be used to determine if the model produces genuine insights, or merely detects spurious patterns. These alternative classifiers are not shown here due to space constraints.

Table 1: Training data (adapted from Gonzalez-Garay et al., 2019)

	Entry Standards	Research Intensity	Staff/Student	Budget per student	Research Quality	Employability
Portsmouth	134	0.580	0.052	3	2.46	1
Surrey	152	0.800	0.065	4	2.98	0
Teesside	113	0.260	0.061	4	2.63	0
Herriot Watt	165	0.860	0.065	5	3.30	1
Sheffield	157	0.880	0.045	5	3.06	1
Bath	203	0.880	0.052	4	3.08	0
Imperial College	233	0.990	0.057	9	3.34	1
Edinburgh	197	0.910	0.067	8	3.30	0
Swansea	141	0.830	0.053	4	3.29	1
Queens Belfast	154	1.000	0.063	5	2.99	1
Leeds	189	0.830	0.078	5	2.97	1
West of Scotland	161	0.450	0.056	4	2.48	0
Lancaster	157	1.000	0.082	7	3.08	1
Cambridge	235	1.000	0.089	10	3.38	1
Hull	111	0.530	0.058	2	2.82	1

Using the training data, several scenarios were considered by varying the value of  $\sum$  from 1.0 to 0.0 in increments of 0.2, with  $\otimes = 0.05$  and using three hyperboxes. The resulting rules were then tested on the validation data. The resulting values for  $\langle$  and  $\otimes$  both for the training and validation data are summarized in Table 3.

The three hyperboxes shown in Table 4 can be translated into three rules as follows:

- Rule 1: IF ( $140.2 \leq \text{Entry Standards} \leq 189.0$ ) AND ( $0.49 \leq \text{Research Intensity}$ ) AND ( $4.4 \leq \text{Budget per Student} \leq 7.6$ ) THEN (Employability = 1)
- Rule 2: IF ( $0.09 \leq \text{Staff/Student Ratio}$ ) AND ( $\text{Research Quality} \leq 2.58$ ) THEN (Employability = 1)
- Rule 3: IF ( $\text{Entry Standards} \leq 134.0$ ) AND ( $\text{Staff/Student Ratio} \leq 0.06$ ) AND ( $\text{Research Quality} \leq 3.24$ ) THEN (Employability = 1)

These rules are disjunctive. Since there are three hyperboxes, a sample is classified as Employability = 1 if it satisfies at least one of the rules indicated above. This set of rules gives good prediction performance, with 78 % balanced accuracy, 56 % sensitivity, and 100 % specificity. One key finding is that high employability does not necessarily result from high research intensity or quality. A possible reason for this surprising result is that the benefits of university research accrue more to postgraduate rather than undergraduate programs. This result needs to be examined further in future work.

Table 2: Validation data (adapted from Gonzalez-Garay et al., 2019)

	Entry Standards	Research Intensity	Staff/Student	Budget per student	Research Quality	Employability
Strathclyde	225	0.870	0.040	3	3.03	1
Nottingham	174	0.860	0.064	7	3.16	1
Loughborough	160	0.930	0.067	4	3.08	0
Birmingham	190	1.000	0.056	9	3.14	1
Newcastle	162	0.740	0.059	5	3.02	1
UCL	187	0.980	0.066	9	3.12	1
Bradford	124	0.280	0.078	2	2.69	1
London South Bank	108	0.910	0.037	2	2.59	1
Aston	127	0.610	0.053	5	2.83	1
Manchester	186	0.970	0.047	5	3.20	1

Table 3: Performance of model for training and validation data sets with varying  $\Sigma$ 

$\Sigma$	Training Data		Validation Data	
	$\langle$	$\otimes$	$\langle$	$\otimes$
1.0	0.0	1.0	0.0	1.0
0.8	0.0	0.80	0.0	0.89
0.6	0.0	0.50	0.0	0.78
0.4	0.0	0.30	0.0	0.44
0.2	0.0	0.10	0.0	0.89
0.0	0.0	0.0	0.0	0.67

The best validation performance occurs at  $\Sigma = 0.4$ . The dimensions of the different hyperboxes are summarized in Table 4.

Table 4: Lower and upper bounds for criteria

	Box 1		Box 2		Box 3	
	LL	UL	LL	UL	LL	UL
Entry Standards	140.2	189.0				134.0
Research Intensity	0.49					
Staff/Student			0.09			0.06
Budget per student	4.4	7.6				
Research Quality				2.58		3.24

## 5. Conclusions

This paper has applied hyperbox-based ML for predicting the employability of chemical engineering graduates based on UK university rankings. The rules generated had satisfactory predictive ability, as characterized by 78 % balanced accuracy, 56 % sensitivity, and 100 % specificity. Results show that research intensity and quality do not necessarily result in high employability. The application highlights the capability of this approach to generate classification rules which offer better insights than conventional black-box ML approaches. In the future, the geographic scope can be broadened to larger regions, or even to a global scale, using public data on world university rankings. Future work also can focus on the use of the same approach for other applications. This technique can be useful for classification problems where transparent, interpretable rule-based models are needed.

## References

- Abad F.M., Chaparro Caso López A.A., 2017, Data-mining techniques in detecting factors linked to academic achievement, School Effectiveness and School Improvement, 28, 39-55.
- Alyahyan E., Düştögör D., 2020, Predicting academic success in higher education: Literature review and best practices, International Journal of Educational Technology in Higher Education, 17, Article 3, 1-21.
- Bal H., Örkücü H.H., 2011, A new mathematical programming approach to multi-group classification problems, Computers & Operations Research, 38, 105-111.

- Chang W., Yuan X., Wu Y., Zhou S., Lei J., Xiao Y., 2019, Decision-making method based on mixed integer linear programming and rough set: A case study of diesel engine quality and assembly clearance data, *Sustainability*, 11, Article 620, 1-21.
- Corrêa R.C., Blaum M., Marengo J., Koch I., Mydlarz M., 2019, An integer programming approach for the 2-class single-group classification problem, *Electronic Notes in Theoretical Computer Science*, 346, 321-331.
- Gonzalez-Garay A., Pozo C., Galan-Martin A., Brechtelsbauer C., Chachuat B., Chadha D., Hale C., Hellgardt K., Kogelbauer A., Matar O.K., McDowell N., Shah N., Guillen-Gosalbez G., 2019, Assessing the performance of UK universities in the field of chemical engineering using data envelopment analysis, *Education for Chemical Engineers*, 29, 29-41.
- Gyenes Z., 2019, Improve process safety in undergraduate education, *Chemical Engineering Transactions*, 77, 397-502.
- Iannarilli F.J., Rubin P.A., 2003, Feature selection for multiclass discrimination via mixed-integer linear programming, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 25, 779-783.
- Jordan M.I., Mitchell T.M., 2015, Machine learning: Trends, perspectives, and prospects, *Science*, 349, 255-260.
- Kim K., Ryoo H.S., 2007, Nonlinear separation of data via mixed 0-1 integer and linear programming, *Applied Mathematics and Computation*, 193, 183-196.
- Labbé M., Martínez-Merino L.I., Rodríguez-Chía A.M., 2019, Mixed integer linear programming for feature selection in support vector machine, *Discrete Applied Mathematics*, 261, 276-304.
- Makridakis S., 2017, The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms, *Futures*, 90, 46-60.
- Maskooki A., 2013, Improving the efficiency of a mixed integer linear programming based approach for multi-class classification problem, *Computers & Industrial Engineering*, 66, 383-388.
- Rudin C., Ertekin Ş., 2018, Learning customized and optimized lists of rules with mathematical programming, *Mathematical Programming Computation*, 10, 659-702.
- Sharma G.D., Yadav A., Chopra R., 2020, Artificial intelligence and effective governance: A review, critique and research agenda, *Sustainable Futures*, 2, Article 100004, 1-6.
- Tan R.R., Aviso K.B., Janairo J.B., Promentilla M.A.B., 2020, A hyperbox classifier model for identifying secure carbon dioxide reservoirs, *Journal of Cleaner Production* (in press).
- Voll P., Jennings M., Hennen M., Shah N., Bardow A., 2015, The optimum is not enough: A near-optimal solution paradigm for energy systems synthesis, *Energy*, 82, 446-456.
- Xu G., Papageorgiou L.G., 2009, A mixed integer optimisation model for data classification, *Computers & Industrial Engineering*, 56, 1205-1215.
- Xu Y., Wang L., Zhang R., 2011, A dynamic attribute reduction algorithm based on 0-1 integer programming, *Knowledge-Based Systems*, 24, 1341-1347.
- Yan K., Ryoo H.S., 2017, 0-1 multilinear programming as a unifying theory for LAD pattern generation, *Discrete Applied Mathematics*, 218, 21-39.
- Yang L., Liu S., Tsoka S., Papageorgiou L.G., 2015a, Sample re-weighting hyper box classifier for multi-class data classification, *Computers & Industrial Engineering*, 85, 44-56.
- Yang L., Ainali C., Kittas A., Nestle F.O., Papageorgiou L.G., Tsoka S., 2015b, Pathway-level disease data mining through hyper-box principles, *Mathematical Biosciences*, 260, 25-34.