Cleaner Engineering and Technology 2 (2021) 100069

Contents lists available at ScienceDirect

Cleaner Engineering and Technology

journal homepage: www.journals.elsevier.com/cleaner-engineering-and-technology

What university attributes predict for graduate employability?

K.B. Aviso[a,*], F.P.A. Demeterio III[b], J.I.B. Janairo[c], R.I.G. Lucas[d], M.A.B. Promentilla[a], R.R. Tan[a], D.E.C. Yu[e]

[a] Chemical Engineering Department, De La Salle University, 2401 Taft Avenue, 0922, Manila, Philippines
[b] Filipino Department, De La Salle University, 2401 Taft Avenue, 0922, Manila, Philippines
[c] Biology Department, De La Salle University, 2401 Taft Avenue, 0922, Manila, Philippines
[d] Department of English and Applied Linguistics, De La Salle University, 2401 Taft Avenue, 0922, Manila, Philippines
[e] Chemistry Department, De La Salle University, 2401 Taft Avenue, 0922, Manila, Philippines

ARTICLE INFO

Keywords:
University rankings
League tables
Research intensity
Employment
Association of Southeast Asian Nations (ASEAN)

ABSTRACT

Research universities play an important role in developing new technologies that can be the basis for economic development, improved quality of life, and reduced environmental impacts. In developing countries, there have been accelerated efforts to transform teaching-oriented higher education institutions into research-intensive universities that can contribute to social development through the generation of new knowledge. There have also been parallel efforts to use internationalization to enrich both education and research. However, the effect of such reforms on the employability of university graduates is unclear. In this work, the influence of different institution attributes on graduate employability is investigated using the hyperbox machine learning technique, which is capable of generating classification models in the form of if/then rules. The analysis focuses on Southeast Asian universities listed in the 2020 Quacquarelli Symonds Asian University Rankings and uses the normalized scores across the different ranking criteria. Five plausible rule-based classifiers are derived and validated. The results show notable association between research and internationalization metrics with employability.

1. Introduction

Universities play an important role as catalysts in green innovation ecosystems (Yang et al., 2020). Research in higher education institutions (HEIs) provides the scientific basis for next generation technologies to enable sustainable development. Partnerships involving academia, industry, and government are critical for accelerating such green innovation. In addition to the knowledge products of academic research, university graduates also shape the future direction of industry through knowledge and attitudes acquired through education (Holdsworth et al., 2019). It is thus important to have a clear understanding of how the dual university roles of research and education interact in the context of the global aspirations for sustainable development.

Artificial intelligence (AI) and machine learning (ML) tools are technological innovations that have tremendous potential to revolutionize higher education, both in strategic educational management and enhancing student learning (Bates et al., 2020). The abundance of educational data brought about by a shift towards records digitization and growing use of learning management systems (Fischer et al., 2020) provides a conducive environment for the accelerated deployment of AI and ML tools for educational applications. ML involves formulating the prediction model using a training dataset followed by evaluating the performance of the model using a test dataset. Mathematical Programming (MP) models can be developed for training in ML wherein the mixed-integer linear programming (MILP) models are a useful class of models for this purpose. MILP models can explore optimal and near-optimal solutions during supervised training by utilizing continuous and integer variables (Voll et al., 2015). Significant applications of MILP in ML include the development of an MILP-based feature selection technique for multiclass discrimination (Iannarilli and Rubin 2003), an MILP model for non-linear data separation (Kim and Ryoo 2007), and a multi-class classification approach based on MILP (Bal and Örkcü 2011). MILP for ML can also help reduce model attributes through a 0–1 programming model (Xu et al., 2011), as well as for pattern recognition (Yan and Ryoo 2017). An MILP approach can likewise be used for optimal rule generation (Rudin and Ertekin 2018), and for binary single group classification (Corrêa et al., 2019). MILP models have also been combined with other ML techniques such as rough sets (Chang et al., 2019) and

* Corresponding author.
E-mail address: kathleen.aviso@dlsu.edu.ph (K.B. Aviso).

https://doi.org/10.1016/j.clet.2021.100069
Received 14 December 2020; Received in revised form 1 February 2021; Accepted 12 March 2021
2666-7908/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Support Vector Machines (SVM) (Labbé et al., 2019).

The previously developed enhanced hyperbox binary classification model built using MILP (Tan et al., 2020) is a promising algorithm to be applied in educational topics and problems, such as in predicting and determining the significant factors associated with graduate employability, which is an important measure of higher education quality (Gyenes 2019). Past works have shown that decision trees algorithm can satisfactorily predict graduate employability (Wang and Li 2016; Tapado et al., 2018). However, a search of the Scopus database reveals that research on this topic with applications for higher education is scarce. In addition, a rule-based algorithm that articulates the decision-making process in a simple manner is desired for this type of AI/ML application. Thus, the enhanced hyperbox binary classification model is appropriate for this problem domain, since it features transparent rule-based decision-making that is presented as a series of if/then statements.

This algorithm is based on the hyperbox classification model created for binary classification (Xu and Papageorgiou 2009). Relative to many black-box ML techniques that suffers from poor model interpretability, this approach highlights the generation of transparent results (Yang et al., 2015b), since the hyperboxes can be interpreted as if/then rules (Tan et al., 2020). The original algorithm needed repeated re-optimization of MILP models to achieve a satisfactory fit, and the training algorithm was further improved with a reduced number of steps (Maskooki 2013). 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., 2015b). Hyperbox-based ML models have found widespread applications such as in business performance prediction (Xu and Papageorgiou 2009), disease diagnosis (Yang et al., 2015a), materials selection for nanotechnology (Janairo et al., 2020), and geological reservoir classification for CO_2 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. This technique was previously applied to predicting employment prospects of chemical engineering graduates in the UK (Aviso et al., 2020); in this work, the employability of graduates across all disciplines from universities in the Association of Southeast Asian Nations (ASEAN) is examined. Only universities listed in the Quacquarelli Symonds Asian University Rankings are included. The rest of this paper is organized as follows. Section 2 gives a literature review on world university rankings. Section 3 gives the formal problem statement, while Section 4 describes the methodology of the hyperbox machine learning technique. Section 5 discusses the results of the analysis of data for ASEAN universities. Implications of these results for university management are discussed in Section 6. Finally, Section 7 gives the conclusions and discusses prospects for future research.

2. Literature review

The rise of world university ranking systems (WURSs) originated from the success of the Academic Ranking of World Universities (ARWU) that was created and launched by Shanghai Jiao Tong University in 2003. ARWU is funded by the Chinese government and was originally designed as a monitoring and benchmarking tool for the improvement of Chinese universities (Pavel, 2015). A year after its creation a privately-owned and commercially-funded WURS emerged, known as the Times Higher Education-Quacquarelli Symonds (THE-QS) World University Rankings (WUR). This venture was a collaboration between two British companies: Times Higher Education (THE) that is into magazine publication, and Quacquarelli Symonds (QS) that is into the analysis of higher educational institutions all over the globe. In 2009, the collaborative venture split into two WURS: THE WUR and QS WUR, with the latter retaining their original joint methodology, and collaboration with the citation indexing company Scopus-Elsevier. THE WUR designed a new methodology and collaborated with the citation indexing company Thomson Reuters before reverting back to Scopus-Elsevier in 2014 (Shahjahan et al., 2020a).

At present, QS and THE are the dominant private and commercial WURSs, and together with other WURSs, they influence the higher education policies of universities and governments all over the globe. In 2019, for example, QS revealed that it had 75,000 academe-based survey respondents, 44,000 employer survey respondents, and covered 4300 universities from 62 countries (Shahjahan et al., 2020a), and its website recorded 85 million visits (Shahjahan et al., 2020b). WURSs are not free from criticisms especially on their methodologies and impact on the global higher educational landscape. Despite their shortcomings, WURSs still serve some purpose, and their data and figures mean things, especially if one keeps away from their final scores and ranks and focus instead on their indicator scores and figures (Soh, 2017). From 2009 to 2016, QS created regional university rankings in order to bring into the picture the fact that there are internal and external differences across regions and universities within the same regions can get more sense and meaning in such rankings. This WURS's regional university rankings was launched with QS Asian University Rankings (QS AUR) in collaboration with the Chosun Ilbo, a Korean newspaper company. QS AUR has 16 indicators, presented in Table 1 with their assigned percentage weights, data gathering method, and rationale.

Graduate employability is difficult to define and quantify (Boden and Nevada, 2010). Some definitions rely on employment outcomes and the ease of securing employment upon graduation. Current research has re-defined graduate employability in relation to individual

Table 1
Indicators for QS AUR (QS Asia University Rankings, 2020).

Indicator	Weight Assigned	Data Gathering Method	Rationale
Academic reputation	30%	Survey among academics from all over the world	Measures the reputation of a universities among the international academic community
Employer reputation	20%	Survey among employers from all over the world	Measures the employability of graduates of universities
Faculty/student ratio	10%	Data coming the universities, or their webpages, or government reports	Measures the contact time and academic support given by universities to their students
International research network	10%	Margalef Index based on Data from Scopus-Elsevier	Measures the international research collaboration of universities
Citations per paper	10%	Data from Scopus-Elsevier	Measures the impact of research of universities
Papers per faculty	5%		Measures the volume of research of universities
Faculty with a PhD	5%	Data coming the universities, or their webpages, or government reports	Measures the universities' commitment to quality education
Proportion of international faculty	2.5%		Measure the international outlook of universities
Proportion of international students	2.5%		
Proportion of inbound exchange students	2.5%		
Proportion of outbound exchange students	2.5%		

characteristics (Azevedo et al., 2012). According to Yorke (2006) employability also depends on transferable soft skills such as communication or interpersonal aptitude. Graduate employability thus encompasses domain knowledge, soft skills, and other personal attributes (Cranmer, 2006). For instance, graduates are expected to possess transferable skills which are highly valued by employers more than their technical knowledge (Fletcher et al., 2017). In addition, trainability, or the ability to acquire new skills over the course of a multi-decade career, is highly valued in the modern world.

Herberta et al. (2020) argue that some HEI programs lack present-day relevance, as they do not seem to address employers' expectation of the required skills and individual propensities of its graduates. As such, graduates are not able to satisfactorily access career opportunities. It appears that there seems to be competencies mismatch between graduates' learning outcomes and the expected skills of employers (Teijeiro et al., 2013). Cavanagh et al. (2015) in their qualitative study involving Australian undergraduate students, found that there was a huge gap between curricula, pedagogical methods and skills to be developed by the students to make them more industry ready. It was also noted that students could not seem to see the relevance between the relationships of their university acquired skills vis-à-vis what is expected from them by the industry. Martin (2014) further argues that university curriculum should be consistent with the demands of the workplace; higher education should thus go beyond just knowledge transmission and ensure the development of skills that the industry needs.

Currently, universities are faced with growing challenges of internationalization, rankings, teaching innovations, relevant curricula and industry-ready graduates. HEIs compete in recruiting the best students, hiring the most productive faculty members, collaborating with industry partners, soliciting funds to construct state of the art campuses, and contributing to knowledge generation (HemsleyBrown et al., 2016). Graduates no longer see themselves competing for work, based on their innate competencies and skills but also on the academic reputation of the institution where they graduated. It has been noted that rankings affect university strategies and administrative decisions. Moreover, universities are appraised and assessed through complex factors of influence such as: provision of a rich learning environment, a promise of job opportunities, students' learning experiences, quality of pedagogy, innovative curricula and the university's reputation (Lozano et al., 2018).

To increase their global and national visibility and to establish their relevance in the society through industry related collaborations, HEIs engage themselves in entrepreneurial activities to increase their revenues and sustain their various university activities such as research, teaching and social engagement. This refers to the so-called university's 'third mission' (Degl'Innocenti et al., 2017). These entrepreneurial initiatives are cascaded down to the university's decisions that impact curricular changes that develop skills and competencies for graduates to establish their own businesses. Such innovations help contribute to the reputation of a university in terms of its ability to contribute to national economic growth (Colombo and Piva, 2020).

3. Formal problem statement

The goal of this work is to investigate university attributes that influence graduate employability. Hyperbox machine learning is used because of its capability to generate readily interpretable models in the form of if/then rules. The resulting rule-based models can be used to generate insights on how research and internationalization metrics translate into career benefits for university graduates.

4. Methodology

The hyperbox machine learning technique generates a classification model consisting of disjunctive if/then rules that are calibrated via a MILP model using a set of training data (Janairo et al., 2020). The MILP model for training is described here.

Nomenclature

Indices	
i	Index for box dimension
j	Index for sample
k	Index for box
Parameters	
Φ	Threshold for proportion of false positives
T	Parameter reduction
bt_j	True belongingness of sample j; $bt_j = 1$ if it is a positive sample and $bt_j = 0$ if it is a negative sample
n^N	Total number of negative samples
n^P	Total number of positive samples
$xs_{j,i}$	Performance of sample j in dimension i
$z^L_{k,i}$	An arbitrary low value to create an illusion that the lower limit of box k in dimension i is boundless
$z^U_{k,i}$	An arbitrary high value to create an illusion that the upper limit of box k in dimension i is boundless
Binary Variables	
b_j	Indicates whether sample j is positive ($b_j = 1$) or not ($b_j = 0$)
$c_{j,k}$	Indicates whether sample j belongs in box k ($c_{j,k} = 1$) or not ($c_{j,k} = 0$)
f^N_j	Indicates if sample j has been falsely classified as negative
f^P_j	Indicates if sample j has been falsely classified as positive
$q^L_{i,j,k}$	Indicates if the performance of sample j in dimension i for box k is lower than the box's lower limit
$q^U_{i,j,k}$	Indicates if the performance of sample j in dimension i for box k is higher than the box's upper limit
$y^L_{k,i}$	Indicates whether the lower bound of box k in dimension i is activated ($y^L_{k,i} = 1$) or not ($y^L_{k,i} = 0$)
$y^U_{k,i}$	Indicates whether the upper bound of box k in dimension i is activated ($y^U_{k,i} = 1$) or not ($y^U_{k,i} = 0$)
$z_{k,i}$	Binary variable used for limiting the number of parameters in the generated rules
Variables	
α	Proportion of false positives
β	Proportion of false negatives
$xb^L_{k,i}$	Lower limit of box k in dimension i
$xb^U_{k,i}$	Upper limit of box k in dimension i

In training the model, the over-all objective is to reduce the proportion of false negatives as indicated in Eq. (1) where the variable β can be calculated using Eq. (2). In addition, the proportion of false positives, α, should not exceed the threshold, ϕ, which can be exogenously defined by the user (Eq. (3)). The proportion of false positives, α, can be calculated using Eq. (4). The number of false positives, f^P_j, is given by Eq. (5) while the number of false negatives, f^N_j, is obtained using Eq. (6) where f^N_j and f^P_j are binary variables as indicated in Eq. (7).

$$\min = \beta \tag{1}$$

Subject to:

$$\beta = \frac{\sum_j f^N_j}{n^P} \tag{2}$$

$$\alpha \leq \phi \tag{3}$$

$$\alpha = \frac{\sum_j f^P_j}{n^N} \tag{4}$$

$$f^P_j \geq b_j - bt_j \tag{5}$$

$$f^N_j \geq bt_j - b_j \tag{6}$$

$$f^N_j, f^P_j \in \{0, 1\} \tag{7}$$

Eqs. (8) and (9) establish the outer boundaries of hyperbox k, while Eqs. (10) and (11) establish the inner boundaries of hyperbox k. If the performance of sample j in parameter i lies within the boundaries of hyperbox k then these samples are enclosed within the hyperbox and $c_{j,k} = 1$. Eqs. (12) and (13) account for instances when the performance of sample j in parameter i lies outside of the boundaries of box k. Variables $q^L_{i,j,k}$ and $q^U_{i,j,k}$ will take a value of 1 if sample i lies outside the boundaries of hyperbox k.

$$xs_{j,i} \geq xb^L_{k,i} - \Delta - M(1 - c_{j,k}) \tag{8}$$

$$xs_{j,i} \leq xb^U_{k,i} + \Delta + M(1 - c_{j,k}) \tag{9}$$

$$xs_{j,i} \geq xb^L_{k,i} - M(1 - c_{j,k}) \tag{10}$$

$$xs_{j,i} \leq xb^U_{k,i} + M(1 - c_{j,k}) \tag{11}$$

$$xs_{j,i} \leq xb^L_{k,i} - \Delta + M\left(1 - q^L_{i,j,k}\right) \tag{12}$$

$$xs_{j,i} \geq xb^U_{k,i} + \Delta - M\left(1 - q^U_{i,j,k}\right) \tag{13}$$

Eqs. (14)–(17) identify whether the upper and lower boundaries are activated for parameter i in hyperbox k with $y^L_{k,i} = 0$ if the lower bound $xb^L_{k,i}$ is activated and $y^U_{k,i} = 0$ if the upper bound $xb^U_{k,i}$ is activated. The magnitudes of $z^L_{k,i}$ and $z^U_{k,i}$ are exogenously defined to represent a situation indicating that the limit for the corresponding parameter in the hyperbox is boundless.

$$xb^L_{k,i} \geq z^L_{k,i} - M\left(1 - y^L_{k,i}\right) \tag{14}$$

$$xb^L_{k,i} \leq z^L_{k,i} + My^L_{k,i} \tag{15}$$

$$xb^U_{k,i} \leq z^U_{k,i} + M\left(1 - y^U_{k,i}\right) \tag{16}$$

$$xb^U_{k,i} \geq z^U_{k,i} - My^U_{k,i} \tag{17}$$

The variable $z_{k,i}$ represents the total number of parameters eliminated from the decision rules and this can be obtained by using Eq. (18)–20 while Eq. (21) defines the desired degree of parameter reduction.

$$z_{k,i} \geq \left(1 - y^L_{k,i}\right) + \left(1 - y^U_{k,i}\right) - 1 \tag{18}$$

$$z_{k,i} \leq \left(1 - y^L_{k,i}\right) \tag{19}$$

$$z_{k,i} \leq \left(1 - y^U_{k,i}\right) \tag{20}$$

$$z_{k,i} \geq \tau \tag{21}$$

Eqs. (22) and (23) indicate that a sample j is not enclosed by a box if its performance lies outside the limits of hyperbox k even in just one parameter. Eqs. (25) and (26) indicate that a sample is classified to be positive if it is contained within the boundaries of at least one hyperbox. Eq. (27) indicates all other binary variables.

$$q^L_{i,j,k} + q^U_{i,j,k} \leq M(1 - c_{j,k}) \tag{22}$$

$$\sum_i q^L_{i,j,k} + q^U_{i,j,k} \geq (1 - c_{j,k}) \tag{23}$$

$$xb^U_{k,i} \geq xb^L_{k,i} \tag{24}$$

$$c_{j,k} \leq Mb_j \tag{25}$$

$$c_{j,k} \geq b_j \tag{26}$$

$$b_j, c_{j,k}, q^L_{i,j,k}, q^U_{i,j,k}, \ y^L_{k,i}, \ y^U_{k,i}, \ z_{k,i} \in \{0,1\} \tag{27}$$

5. Results

The ASEAN is a regional economic bloc consisting of ten countries of varying size and level of development. In decreasing order of economic development, the countries that make up ASEAN are Singapore, Brunei, Malaysia, Thailand, Indonesia, Vietnam, Philippines, Lao, Cambodia, and Myanmar. Other than Singapore (a highly industrialized city-state) and Brunei (a small, resource-rich country), most ASEAN countries have only recently emphasized the need for their universities to increase contribution to national and regional development through scientific research.

The hyperbox ML technique is applied to the data from the current QS AUR (QS Asia University Rankings, 2020), where there are 48 ASEAN universities listed. Ten (10) conditional attributes were considered with 1 decision attribute. These attributes are summarized in Table 2. Scores for each criterion are normalized on a scale ranging from 0 to 1, and do not represent the actual values of the raw data.

Of the 48 data samples, 36 are randomly selected as training data (representing 75% of data set) while the remaining 12 are used for validation (25% of data set). The normalized scores of the training and validation data can be found in the Appendix. The following model assumptions are implemented. Since the decision attribute is continuous, a cut-off score of 75 is used to distinguish between graduates with high employability and those that are not. If D1 $\geq$ 75, then these samples were considered of high employability and their corresponding $bt_j = 1$. However, for institutions with D1 $<$ 75, then the corresponding value of $bt_j = 0$. The characteristics of the training and validation data are summarized in Table 3.

Eq. (1) was then solved with model parameters summarized in Table 4.

5.1. Rule-based classifiers 1 and 2

Multiple optimal solutions are obtained upon solving the MILP model. The first solution is summarized in Table 5. This decision rule set achieved a performance of $\alpha = 0.24$ and $\beta = 0$ for the training data.

These hyperboxes can be translated into disjunctive if/then rules as follows:

Rule 1a: IF (International Research Network $>$ 57.03) AND (5 $<$ Papers per faculty $<$ 77.1) AND (International Faculty $>$ 33.7) AND (Citations per paper $>$ 41.9) THEN (Employer Reputation $=$ High).OR

Rule 1b: IF (International research network $>$ 36.1) AND (Inbound exchange $>$ 40.3) AND (Outbound exchange $>$ 22.1) AND (5 $<$ Papers per faculty $<$ 77.1) AND (Citations per paper $>$ 41.9) THEN (Employer Reputation $=$ High).

Note that both rules indicate that high levels of internationalization and research impact, coupled with moderate research productivity, are associated with high employability. When tested on the validation data,

Table 2
Condition and decision attributes.

Attribute Type	Code	Description
Condition	A1	International Research Network
	A2	Inbound Exchange
	A2	Outbound Exchange
	A4	Papers per Faculty
	A5	International Students
	A6	International Faculty
	A7	Faculty with PhD
	A8	Citations per Paper
	A9	Faculty Student Ratio
	A10	Academic Reputation
Decision	D1	Employer Reputation

Table 3
Distribution of samples between training and validation.

	True Positive	True Negative	Total
Training	7	29	36
Validation	3	9	12
Total	19	38	48

Table 4
Settings of model parameters.

Description	Model Parameter	Value
Number of hyperboxes	N_k	2
Tolerance for fraction of false positives	ϕ	0.25
Distance between inner and outer boxes	Δ	5
Arbitrary bounds for boundless limits	$z^L_{k,i}$, $z^U_{k,i}$	−10,000, 10,000
Minimum number of parameters to be eliminated in decision rule	τ	5

Table 5
Decision rules of optimal solution (classifier 1).

Attribute	$k = 1$		$k = 2$	
	$x^L_{i,1}$	$x^U_{i,1}$	$x^L_{i,2}$	$x^U_{i,2}$
A1	57.03	–	36.1	–
A2	–	–	40.3	–
A3	–	–	22.1	–
A4	5	77.1	5	77.1
A5	–	–	–	–
A6	33.7	–	33.7	–
A7	–	–	–	–
A8	41.9	–	–	–
A9	–	–	–	–
A10	–	–	–	–

these rules are able to correctly classify 2 of 3 positive samples and 9 of 9 negative samples resulting in $\alpha = 0$ and $\beta = 0.33$. One positive sample activated Rule 1a while the other activated Rule 1b. The performance of the rules with respect to training and validation are summarized in Tables 6 and 7, respectively.

An alternative rule set (Solution 3) which has similar training performance to Solution 1, but performs better in classifying the validation data, is shown in Table 8.

The hyperboxes can be translated into decision rules as follows:

Rule 2a: IF (Papers per faculty < 69.6) AND (32.3 < Faculty staff with PhD < 69.1) AND (47.2 < Citations per paper < 49.7) THEN (Employer reputation = High).OR

Rule 2b: IF (International research network > 28.5) AND (Papers per faculty < 77.1) AND (Faculty student ratio > 44.9) THEN (Employer reputation = High).

For both rules, papers published by faculty has been identified to be a significant indicator, with the first rule further emphasizing the importance of research by identifying paper citation and faculty with PhD degrees as additional indicators. These indicators are linked with each other since PhD degrees signify adequate training for conducting research. When used on the validation data, these rules are able to classify the samples without error, resulting in $\alpha = 0$ and $\beta = 0$. The positive samples were all identified using Rule 2b. The performance of the rules with respect to training and validation are summarized in Tables 9 and 10, respectively.

Table 6
Performance of Classifier 1 on Training Data (α = 0.24; β = 0).

Training Data	Predicted Positive	Predicted Negative	Total
Actual Positive	7	0	7
Actual Negative	7	22	29
Total	14	22	36

Table 7
Performance of Classifier 1 on Validation Data (α = 0; β = 0.33).

Validation	Predicted Positive	Predicted Negative	Total
Actual Positive	2	1	3
Actual Negative	0	9	9
Total	2	10	12

Table 8
Decision rules for alternative solution (classifier 2).

Attribute	$k = 1$		$k = 2$	
	$x^L_{i,1}$	$x^U_{i,1}$	$x^L_{i,2}$	$x^U_{i,2}$
A1	–	–	28.5	–
A2	–	–	–	–
A3	–	–	–	–
A4	–	69.6	–	77.1
A5	–	–	–	–
A6	–	–	–	–
A7	32.3	69.1	–	–
A8	47.2	49.7	–	–
A9	–	–	44.9	–
A10	–	–	–	–

5.2. Rule-based classifiers 3–5

Alternative decision rule sets can be generated to focus more on generating the lower limits to the attributes considered. Eq. (28) is included as an additional constraint which restricts the model from defining an upper bound to the attributes.

$$y^U_{k,i} = 0 \quad (28)$$

A total of 20 solutions are obtained with this modified MILP. The performance of theses classifiers during training and validation are summarized in Table 11. It is important to note that these are all degenerate solutions, which indicates that all 20 rule sets performed equally with respect to the objective function of minimizing β. For brevity, only Solutions 13, 16 and 19 are presented in this work. These are chosen because of their superior performance in validation.

The decision rules corresponding to Solution 13 (Classifier 3) are shown in Table 12.

The hyperboxes can be translated into the following decision rules:

Rule 3a: IF (International research network > 31.1) AND (Inbound Exchange > 56.4) AND (Outbound exchange > 54.7) THEN (Employer Reputation = High)OR

Rule 3b: IF (Papers per faculty > 21.9) AND (International faculty > 33.7) AND (Faculty staff with PhD > 8.48) THEN (Employer Reputation = High)

Both rules indicate that research and internationalization metrics are associated with high employability. For the validation data, 2 of 3 positive samples have been identified correctly with one positive sample correctly classified by Rule 3a, and the other classified using Rule 3b.

The decision rules corresponding to Solution 16 (Classifier 4) are shown in Table 13. The hyperboxes in Table 13 can be translated into the following decision rules:

Table 9
Performance of Classifier 2 on training data (α = 0.24; β = 0).

Training Data	Predicted Positive	Predicted Negative	Total
Actual Positive	7	0	7
Actual Negative	7	22	29
Total	14	22	36

Table 10
Performance of Classifier 2 on validation data (α = 0; β = 0).

Validation	Predicted Positive	Predicted Negative	Total
Actual Positive	3	0	3
Actual Negative	0	9	9
Total	3	9	12

Table 11
Summary of performance of 20 different decision rule sets.

Solution number	Training		Validation	
	α	β	α	β
1	0	0	0	0.6667
2	0.1379	0	0	0.6667
3	0	0	0	1
4	0	0	0	1
5	0.1379	0	0	0.6667
6	0.1724	0	0	1
7	0.06897	0	0	0.6667
8	0.03448	0	0	1
9	0.2414	0	0	0.6667
10	0	0	0	0.6667
11	0.06897	0	0	0.6667
12	0.03448	0	0	0.6667
13	0.2069	0	0	0.3333
14	0.06897	0	0	0.6667
15	0.06897	0	0	0.6667
16	0.06897	0	0	0.3333
17	0	0	0	0.6667
18	0.1379	0	0	0.6667
19	0.2069	0	0	0.3333
20	0.06897	0	0	0.6667

Table 12
Decision rules for solution 13 (classifier 3).

Attributes	$k = 1$		$k = 2$	
	$x_{i,1}^L$	$x_{i,1}^U$	$x_{i,2}^L$	$x_{i,2}^U$
A1	31.1	–	–	–
A2	56.4	–	–	–
A3	54.7	–	–	–
A4	–	–	21.9	–
A5	–	–	–	–
A6	–	–	33.7	–
A7	–	–	8.48	–
A8	–	–	–	–
A9	–	–	–	–
A10	–	–	–	–

Table 13
Decision rules of solution 16 (classifier 4).

Attribute	$k = 1$		$k = 2$	
	$x_{i,1}^L$	$x_{i,1}^U$	$x_{i,2}^L$	$x_{i,2}^U$
A1	–	–	66	–
A2	45.2	–	–	–
A3	–	–	–	–
A4	–	–	24.6	–
A5	–	–	–	–
A6	–	–	–	–
A7	–	–	21.2	–
A8	–	–	57.8	–
A9	34	–	26.1	–
A10	58.7	–	–	–

Rule 4a: IF (Inbound exchange > 45.2) AND (Faculty student ratio > 34) AND (Academic reputation > 58.7) THEN (Employer reputation = High).OR

Rule 4b: IF (International research network > 66) AND (Papers per faculty > 24.6) AND (Faculty staff with PhD > 21.2) AND (Citations per paper > 57.8) AND (Faculty student ratio > 26.1) THEN (Employer reputation = High).

In addition to the influence of research and internationalization metrics, these rules also signify the roles of academic reputation, faculty student ratio (a proxy measure of teaching quality) and faculty staff with PhD (an indicator of human resources quality) in predicting for high levels of employability. This classifier also shows the association between institutional reputation as perceived by the academic community and by prospective employers in industry. Two of the positive samples in the validation data are correctly classified using Rule 4a.

The decision rules corresponding to Solution 19 (Classifier 5) shown in Table 14 and can be translated into the following decision rules:

Rule 5a: IF (Faculty staff with PhD > 13.7) AND (Faculty student ratio > 34) THEN (Employer reputation = High).OR

Rule 5b: IF (International faculty > 94.4) AND (Faculty student ratio > 45.1) THEN (Employer reputation = High).

These rules highlight the role of faculty quantity, quality, and diversity in determining the employability of university graduates. One of the positive samples in the validation data is classified correctly using Rule 5a while the other positive sample is classified by Rule 5b.

6. Discussion of implications

A summary of the attributes which were activated in the different rule sets presented is shown in Table 15. It can be seen that among all attributes, A5 (International Students) does not appear in any of the decision rules, indicating that this aspect has negligible influence on employment prospects. On the other hand, A4 (Papers per faculty) appears most frequently, followed by A1 (International Research Network) and A9 (Faculty Student Ratio). These indicators suggest that internationalization, research, and quality of education are important for generating graduates with high employability.

These results have significant practical implications for the management of universities, particularly in emerging economies of similar level of development as the majority of the ASEAN countries. Many institutions in the region were established for the primary purpose of training an educated workforce in aid of national development. Education is thus traditionally seen as being the core function of universities. The research function has only been recently added, as a result of belated recognition of the role of scientific research as an essential investment to further drive growth. The rapid transition towards increased research intensity has led to tension between the educational and research functions in many institutions.

Although research is understood as being essential from a strategic standpoint, implementation at the grassroots level leads to competition for limited human, financial, and infrastructure resources. Similarly, internationalization is often perceived as having a cosmetic rather than fundamental function. The results of this work may alter the cost-benefit analysis calculus by offering preliminary empirical evidence that the investment of institutional resources in research and internationalization leads to benefits that also accrue to university graduates.

7. Conclusion

In this work, associations between institutional attributes and graduate employability in ASEAN universities were analyzed using hyperbox machine learning. Five different plausible rule-based classifiers were generated and trained using data from the 2020 QS Asia University Rankings. Like all ML techniques, even though the classification is imperfect, general patterns in the data are revealed by the rules. Research and internationalization metrics were shown to be associated with high levels of employability, along with attributes indicative of the willingness of institutions to commit resources to higher education and to human resource development. It should be noted that these rules do not necessarily imply causality; these associations may be the result of underlying

Table 14
Decision rules for Solution 19 (Classifier 5).

Attribute	$k=1$		$k=2$	
	$x_{i,1}^L$	$x_{i,1}^U$	$x_{i,2}^L$	$x_{i,2}^U$
A1	–	–	–	–
A2	–	–	–	–
A3	–	–	–	–
A4	–	–	–	–
A5	–	–	–	–
A6	–	–	94.4	–
A7	13.7	–	–	–
A8	–	–	–	–
A9	34	–	45.1	–
A10	–	–	–	–

Table 15
Summary of activated attributes in selected decision rule sets.

	Classifier 1		Classifier 2		Classifier 3		Classifier 4		Classifier 5		N
	A	B	A	B	A	B	A	B	A	B	
A1	✓	✓		✓	✓			✓			5
A2		✓			✓		✓				3
A3		✓			✓						2
A4	✓	✓	✓	✓		✓		✓			6
A5											0
A6	✓	✓				✓				✓	4
A7			✓			✓		✓	✓		4
A8	✓		✓					✓			3
A9				✓			✓	✓	✓	✓	5
A10							✓				1

links to hidden causal attributes that are not reflected in the ranking criteria. Nevertheless, the rules are sufficiently plausible to suggest links between institutional research and internationalization on one hand, and employment prospects on the other.

While research, internationalization, monetary and human resources are university attributes that directly benefit the graduates and translate into university reputation that becomes the main factor for employers in determining employability, future studies can aim on another factor that contributes to university reputation – the alumni. The alumni sector serves as tangible evidence of university attributes. Collectively, the accomplishments and social impacts of the alumni enrich the university's heritage that could undoubtedly affect university reputation. It will be interesting to determine the proportion of the alumni factor in university reputation that influences the employability of new graduates.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clet.2021.100069.

References

Aviso, K.B., Janairo, J.I.B., Lucas, R.I.G., Promentilla, M.A.B., Yu, D.E.C., Tan, R.R., 2020. Predicting higher education outcomes with hyperbox machine learning: what factors influence graduate employability? Chem. Eng. Trans. 81, 679–684.

Azevedo, A., Apfelthaler, G., Hurst, D., 2012. Competency development in business graduates: an industry-driven approach for examining the alignment of our undergraduate business education with industry requirements. Int. J. Manag. Educ. 10, 12–28.

Bal, H., Örkcü, H., 2011. A new mathematical programming approach to multi-group classification problems. Comput. Oper. Res. 38, 105–111.

Bates, T., Cobo, C., Mariño, O., Wheeler, S., 2020. Can artificial intelligence transform higher education? Int. J. Educ. Technol. High. Educ. 17 https://doi.org/10.1186/s41239-020-00218-x.

Boden, R., Nevada, M., 2010. Employing discourse: universities and graduate 'employability'. J. Educ. Pol. 25, 37–54.

Cavanagh, J., Burston, M., Southcombe, A., Bartram, T., 2015. Contributing to a graduate-centred understanding of work readiness: an exploratory study of Australian undergraduate students' perceptions of their employability. Int. J. Manag. Educ. 13, 278–288.

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, 620.

Colombo, M.G., Piva, E., 2020. Start-ups launched by recent STEM university graduates: the impact of university education on entrepreneurial entry. Res. Pol. 49, 103993.

Corrêa, R.C., Blaum, M., Marenco, J., Koch, I., Mydlarz, M., 2019. An integer programming approach for the 2-class single-group classification problem. Electron. Notes Theor. Comput. Sci. 346, 321–331.

Cranmer, S., 2006. Enhancing graduate employability: best intentions and mixed outcomes. Stud. High Educ. 31, 169–184.

Degl' Innocenti, M., Matousek, R., Tzeremes, N., 2017. The interconnections of academic research and universities' "third mission": evidence from the UK. Res. Pol. 48, 103793.

Fischer, C., Pardos, Z.A., Baker, R.S., Williams, J.J., Smyth, P., Yu, R., Slater, S., Baker, R., Warschauer, M., 2020. Mining big data in education: affordances and challenges. Rev. Res. Educ. 44, 130–160.

Fletcher, A.J., Sharif, A.W.A., Haw, M.D., 2017. Using the perceptions of chemical engineering students and graduates to develop employ ability skills. Educ. Chem. Eng. 18, 11–25.

Gyenes, Z., 2019. Improve process safety in undergraduate education. Chem. Eng. Trans. 77, 397–402.

Hemsley-Brown, J., Melewar, T.C., Nguyen, B., Wilson, E.J., 2016. Exploring brand identity, meaning, image, and reputation (BIMIR) in higher education: a special section. J. Bus. Res. 69, 3019–3022.

Herberta, I.P., Rothwella, A.T., Gloverb, J.T., Lamberta, S.A., 2020. Graduate employability, employment prospects and work readiness in the changing field of professional work. Int. J. Manag. Educ. 18, 1–13.

Holdsworth, S., Sandri, O., Thomas, I., Wong, P., Chester, A., McLaughlin, P., 2019. The assessment of graduate sustainability attributes in the workplace: potential advantages of using the theory of planned behaviour (TPB). J. Clean. Prod. 238, 117929.

Iannarilli, F.J., Rubin, P.A., 2003. Feature selection for multiclass discrimination via mixed-Integer linear programming. IEEE Trans. Pattern Anal. Mach. Intell. 25, 779–783.

Janairo, J.I.B., Aviso, K.B., Promentilla, M.A.B., Tan, R.R., 2020. Enhanced hyperbox classifier model for nanomaterial discovery. A&I 1, 299–311.

Kim, K., Ryoo, H.S., 2007. Nonlinear separation of data via mixed 0-1 integer and linear programming. Appl. Math. Comput. 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 Appl. Math. 261, 276–304.

Lozano, J.M., Bofarull, I., Waddock, S., Prat-i-Pubill, Q., 2018. Avoiding the iron cage of business school rankings. High Educ. Pol. 33, 135–157.

Martin, R.D., 2014. The importance of communication competency for employability. Procedia Soc. Behav. Sci. 139, 387–394.

Maskooki, A., 2013. Improving the efficiency of a mixed integer linear programming based approach for multi-class classification problem. Comput. Ind. Eng. 66, 383–388.

Pavel, A.P., 2015. Global university rankings-a comparative analysis. Procedia Econ. Fin. 26, 54–63.

QS Asian University Rankings, 2020. www.topuniversities.com/university-rankings/asian-university-rankings/2020. (Accessed 5 December 2020).

Rudin, C., Ertekin, Ş., 2018. Learning customized and optimized lists of rules with mathematical programming. Math. Program. Comput. 10, 659–702.

Shahjahan, R.A., Sonneveldt, E.L., Estera, A.L., Bae, S., 2020a. Emoscapes and commercial university rankers: the role of affect in global higher education policy. Crit. Stud. Educ. https://doi.org/10.1080/17508487.2020.1748078 (in press).

Shahjahan, R.A., Estera, A.L., Bae, S., Sonneveldt, E.L., 2020b. Imagining 'Asian'higher education: visual campus gaze and global university rankings (GURs) websites. Compare. https://doi.org/10.1080/03057925.2020.1746176 (in press).

Soh, K., 2017. The seven deadly sins of world university ranking: a summary from several papers. J. High Educ. Pol. Manag. 39, 104–115.

Tan, R.R., Aviso, K.B., Janairo, J.I.B., Promentilla, M.A.B., 2020. A hyperbox classifier model for identifying secure carbon dioxide reservoirs. J. Clean. Prod. 272, 22181.

Tapado, B.M., Acedo, G.G., Palaoag, T.D., 2018. Evaluating information technology graduates employability using decision tree algorithm. In: Proceedings of the 9th International Conference on E-Education, E-Business, E-Management and E-Learning, pp. 88–93.

Teijeiro, Mercedes, Rung, P., Freirea, M.J., 2013. Graduate competencies and employability: the impact of matching firms' needs and personal attainments. Econ. Educ. Rev. 34, 286–295.

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.

Wang, J., Li, L., 2016. Research on the college graduate employment education based on data mining technology. Anthropol. 23, 231–235.

Xu, G., Papageorgiou, L.G., 2009. A mixed integer optimisation model for data classification. Comput. Ind. Eng. 56, 1205–1215.
Xu, Y., Wang, L., Zhang, R., 2011. A dynamic attribute reduction algorithm based on 0-1 integer programming. Knowl. Base Syst. 24, 1341–1347.
Yan, K., Ryoo, H.S., 2017. 0-1 multilinear programming as a unifying theory for LAD pattern generation. Discrete Appl. Math. 218, 21–39.
Yang, L., Ainali, C., Kittas, A., Nestle, F.O., Papageorgiou, L.G., Tsoka, S., 2015a. Pathway-level disease data mining through hyper-box principles. Math. Biosci. 260, 25–34.
Yang, L., Liu, S., Tsoka, S., Papageorgiou, L.G., 2015b. Sample re-weighting hyper box classifier for multi-class data classification. Comput. Ind. Eng. 85, 44–56.
Yang, Z., Chen, H., Du, L., Lin, C., Lu, W., 2020. How does alliance-based government-university-industry foster cleantech innovation in a green innovation ecosystem? J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2020.124559 (in press).
Yorke, M., 2006. Employability in Higher Education: what it Is – what it Is Not. The Higher Education Academy, York, UK.