Yang, T.-C., Liu, Y.-L., & Wang, L.-C. (2021). Using an Institutional Research Perspective to Predict Undergraduate Students' Career Decisions in the Practice of Precision Education. *Educational Technology & Society*, *24* (1), 280-296.

Using an Institutional Research Perspective to Predict Undergraduate Students' Career Decisions in the Practice of Precision Education

Tzu-Chi Yang[1,2], Yih-Lan Liu[2] and Li-Chun Wang[3*]

[1]Department of Mathematics and Information Education, National Taipei University of Education, Taiwan (R.O.C.)// [2]Institute of Education/Center for Teacher Education, National Chiao Tung University, Taiwan (R.O.C.) // [3]Department of Electrical Engineering, National Chiao Tung University, Taiwan (R.O.C.) // tcyang.academic@gmail.com // ylliu@mail.nctu.edu.tw // lichun@g2.nctu.edu.tw

[*]Corresponding author

ABSTRACT: The recently increased importance of practicing precision education has attracted much attention. To better understand students' learning and the relationship between their individual differences and learning outcomes, the bird-eye view possible for educational policymakers and stakeholders from educational data mining and institutional research has gained importance and momentum. The deployment of specific predictive tasks based on institutional data is the most promising solution for dealing with a variety of issues on precision education. Most research in this field is focused on learning performance and related issues, such as at-risk students and drop-out tendencies. Seldom are the relationships between the learning performance and career decisions of students investigated. However, developing a deep understanding of students' career decisions plays an important role in the practice of precision education. In this vein, this paper provides a comprehensive analysis and comparison of the state of the art of predictive techniques for providing a prediction for students' career decisions. The results indicated that it is possible to perform early detection of students' career decisions. The contributions of this study are discussed in terms of their implications for theory, methodology, and application.

Keywords: Precision education, Institutional Research, Students' life planning, Machine learning, Educational data mining

1. Introduction

Due to the massification and diversification of higher education, ensuring broad student success has become a multifaceted challenge for higher education institutions (Arthars et al., 2019). Many researchers (Chen and Wang, 2020) have noted the importance of implementing personalized learning to accommodate students' individual differences, as such differences are closely related to learning performance, attitudes, career decisions, self-efficacy, and success in learning. The provision of learning support without loss of timeliness or personalization is promising (Kokkinos & Saripanidis, 2017; Pardo, 2018; Valladares et al., 2018). The implementation of personalized learning relies on the collection and analysis of learning data drawn from a range of students (Beemer et al., 2018). These records, however, are commonly available in institutional file systems/databases and independently stored in different departments. An institutional database that can store a range of data related to students' individual characteristics and learning is valuable for conducting institution-specific research under conditions of resource constraint (Caison, 2007).

Likewise, better or more comprehensive learning data collection and exploration in relation to learning, from an institutional perspective, has been proposed to support personalized learning (Arthars et al., 2019; Flaig, Simonsmeier et al., 2018). This is because data are easy to find in institutions and can be incorporated to support decision-making (Swing & Ross, 2016). Understanding the rationales and applications of this approach can also be furthered through an objective and reflective examination of historical trends, current practices, and the strengths and weaknesses of teaching and training strategies (Ross & Morrison, 2012). In this vein, the use of an institutional database to deal with personalized learning issues has attracted much attention (Camacho & Legare, 2016; Laguilles, 2016). Education as a whole is also showing progressive movement from a one-size-fits-all approach to a type of precision or personalized education (Lu et al., 2018; Tsai et al., 2020).

Researchers have argued that most personalized learning solutions, in the form of educational tools and flexible learning systems, accommodate individual learners' interactions and learning progress, and they fit the specific needs of the individual learner (Xie et al., 2019; Zawacki-Richter et al., 2019). Nevertheless, data analytics have become inseparable from policy development and governance modes. There is a need to take up new knowledge and provide advanced resources in education that are based on data-led policymaking, such as data-driven precision education (Gulson & Webb, 2017; Williamson, 2019).

Precision education is used to identify students' personal characteristics and individual needs to provide educational researchers and practitioners with the tools to understand students better and to allow for a more effective approach to education (Hart, 2016; Makhluf, 2020). For example, Rump et al. (2017) predicted students' intention to drop out based on learning motivation. Their results indicated that higher education institutions can adopt prophylactic measures that could prevent college students from dropout. Hoffait and Schyns (2017) developed institutional data mining to early detect freshmen likely to face difficulties to allow universities to timely and efficiently provide remediation or reorientation. Cheng et al. (2018) predicted potential high-risk freshmen in three core university courses using institutional databases and demonstrated that this predictive task can ameliorate the problems of freshman unpreparedness by providing intervention with the support of the Office of Student Affairs.

In this light, researchers (Thompson et al., 2019) have investigated the myriad of stressors that students face as they transit into university study, during which they continue to balance competing demands of academic study, social relationships, and personal needs. Researchers have also found that many students are concerned about their educational and career decisions during their college experience (Miller et al., 2018). Hence, greater attention than ever is being placed on the ways in which universities enable their graduates to meet their career goals (Healy et al., 2020). Thus, students' career decisions should be taken into account as they develop strategies for precision education.

On the one hand, students rely on the learning support and guidance provided by universities, especially as relates to career path development. On the other hand, allocating these learning resources to students relies on supportive educational policy making and practice that is associated with institutional research. Such institutional methods must be examined through a bird's eye view. To this end, this study proposes a framework that uses historical trends to predict undergraduate students' career decisions. More specifically, it explores the relationships among individual students' characteristics, learning performance, and career decisions in relation to institutional data in pursuit of this research question: "Is it possible to predict students' career decisions based on individual differences and learning data?" Then, the second research question, namely, "How can the career decisions predictive model can enhance the practice of precision education?" is addressed within the context of precision education for institutional research.

2. Literature review

2.1. Precision education

Precision education can be defined as an approach to research and practice that allow preventive and intervention practices to be tailored to individuals based on the best available evidence. Put another way, it is an approach to creating a learning environment that provides learners with precise instruction, assistance, and resources according to their individual needs, based on the best available evidence (Cook et al., 2018). This approach is inspired by precision medicine, as mentioned by President Barack Obama (Collins & Varmus, 2015; Hart, 2016). Precision medicine focuses on understanding individual variability in disease prevention, care, and treatment in a way that takes into account individual variability in genes, environment, and lifestyle (Collins et al., 2016; Reed & Gates, 2020). In relation to the educational context, the interventions that are most aligned with learners' needs can be precisely evaluated and delivered by linking particular issues (e.g., social, academic, and physical health) to create personalized educational experiences for learners (Burns et al., 2010; Palanica et al., 2019). This approach has attracted much attention and is expected to produce superior outcomes to standardized one-size-fits-all or trial-and-error approaches that are often seen in education (Cook et al., 2018; Gulson & Webb, 2018).

For examples, Bahr (2009) explored the relationship between occurrence and frequency of students' lateral transfer in relation to their completion of a credential and he classified students' behavior pattern into six groups (e.g., vocational, drop-in, noncredit, and exploratory). Prediction of student behavior pattern allows learning resource arrangements to be made to improve long-term outcomes (Bahr, 2010). Laanan and Jain (2016) indicated that the transfer process can be a daunting experience for students who aspire to complete a bachelor's degree. Institution research plays an important role in helping educational institutions evaluate complex processes so that relevant strategies can be developed to deal with issues that arise. Lu et al. (2018) considered precision education as a framework for use in improving learning performance. They collected and analyzed students' learning behaviors and to predict their learning performance from them. From this, certain critical features, such as online learning (e.g., video-viewing and out-of-class practice), homework and quiz scores, and after-school tutoring were found to be closely related to learning performance. Accordingly, these researchers suggested that monitoring and intervening in students' online and off-line learning behaviors in response to their observed learning behavior patterns could be a promising approach to improving learning performance and practicing precision education. Likewise, O'Connor and Daly (2018) examined different combinations of antecedent- and consequent-based strategies for students with escape-motivated problems. They found that students with escape-motivated behavior are a heterogeneous group. This implies that it is possible to develop individualized solutions (e.g., teaching, consequent strategies) to a target student.

Following the rationale of precision education, researchers (Cook et al., 2018) developed the idea that the practice of precision education, including (a) flexible and adaptive interventions that allow for adaptation and matching to student needs, (b) formatively monitoring and evaluation of data over time that allows stakeholders to further modify interventions given sufficient response or to select an alternative strategy. Then, (c) this amounts to collaboration, allowing the making of timely and appropriate decisions regarding intervention adaptation and matching. This might modify the existing intervention, and implement an alternative one, or simply terminate the intervention outright.

In a nutshell, the practice of precision education involves tailoring mechanisms of prevention and intervention to certain individuals, following the best evidence regarding their needs and potential, established with evidence-based approaches. This allows the selection and provision of more precise treatments for individual learners. This not only enhances learning outcomes but also avoids educational waste, such as the costs in time, finances, and resources (Kazdin & Blasé, 2011).

2.2. Institutional research perspectives on precision education

As the aforementioned, precision education refers to evidence-based innovation in education that takes advantage of big data and data science to enhance traditional educational practices (Christensen & Eyring 2011; Tsai et al., 2020). The critical process here is to analyze big data and develop models to predict student success, meet individual needs, and fulfill educational purposes, such as academic performance, graduation rates, resource consumption, counseling, and many others (Warren et al., 2017). Comprehensive data sources of individuals' characteristics, learning contexts, and portfolios are typically organized and conducted in data/information centers with institutional research professionals (Lonn & Koester, 2019; Rienties & Toetenel, 2016). Further, precision education is a part of a rising uptake of data science and domain knowledge in educational policy, institutional research, and advocacy for data-led policymaking (Gulson & Webb, 2017; Williamson, 2019). Many researchers have argued that it is particularly important to collect, organize, analyze, and utilize comprehensive learning data to support making decisions from a bird's-eye view (e.g., institutional research) to achieve precision education (Bellgard, 2020; Kuch et al., 2020).

Over the past decade, various research approaches and frameworks have been proposed for the practice of precision education in an institutional research perspective (Xing et al., 2015). For example, Márquez-Vera et. al. (2016) proposed a methodology for creating early prediction models for student dropout as soon as possible. They found that this method was sufficiently trustworthy to be used in an early warning system to improve student retention. Du et al. (2019) developed a framework for predicting the learning performance and creating an early warning system for at-risk students to improve learning outcomes. Adekitan and Salau (2019) explored the impact of engineering students' performance in the first three years on their graduation results. If students' learning performance can be predicted, early intervention can be deployed to prevent them from graduating with poor results or from not graduating at all.

Early prediction provides various opportunities to assess information on student learning and to facilitate decision making to serve improvements in education (Daghestani et al., 2020). Most research into this field focuses on learning performance, although student learning could be influenced by many factors, such as career factors (Freeman et al., 2017; Kim et al., 2018). Researchers have found that people seek to become employable by pursuing individually tailored precision education and learning based on their individual needs and what they hope to achieve in life (Santos et al., 2018; Brunila, 2020).

Recently, more attention has been given to producing appropriate methods and contents for career counseling in university students (Peng et al., 2020). This is because individual learning experiences and background factors affect self-efficacy, learning experience, and career path development in students. Students seek to rely on the resources provided by the university, such as learning facilities and career guidance, when they consider the academic or career path they are hoping to follow (Berger et al., 2019; Rivera & Li, 2020). Here, universities are increasingly aware that they have a major responsibility to assist students, not only in academic development but also in career development (Meijers & Kuijpers, 2014). Accommodating the individual need to make career decisions and providing sufficient support to students are an important goal in the practice of precision education (Abdullah et al., 2018; Ulas & Yildirim, 2019). Because precision education relies on educational policy making and practices that are associated with institutional research, it would be promising to explore the methods of practicing precision education in the university with a large-scale view of the university. Thus, we use an institutional research perspective to propose a model for predicting student career decisions and to elaborate how such a model can be used in the practice of precision education.

3. Method

3.1. Data source and data selection

The data used in this study were collected from a de-identified institutional data warehouse supervised by the Big data research center (BDRC) of National Chiao-Tung University in Taiwan. A data warehouse is the repository of the stored data of an institution and is designed to facilitate analysis, and reporting solutions that deal with institutional research issues. In general, it is a duplicate database system that dumps raw data from more than 80% of departments/centers in the university. Generally, the data warehouse supervised by BDRC of the aforementioned university are built with an Extract-Transform-Load (ETL). All data were pre-processed to remove all personally identifying information (e.g., name, student ID number, and contact information), following the privacy policy. A comma-separated-values (csv) file was generated and delivered to the researcher after the dataset application was complete. This study used a sample of 7003 undergraduate students who had enrolled in the university from August 2010 to September 2015, and who had received a bachelor's degree in no more than five years, and were not suspended or did not drop out.

3.2. Conceptual model of data pre-processing

The target university is a research university, known for its science and engineering achievements, especially in the fields of electronics, information communications, and optoelectronics. The university is located near to a Science Industrial Park. After graduation, most university students either seek an academic career or a position in industry. Therefore, in this study, career decisions were classified into four categories, i.e., academic career, engagement in advanced studies in current university, industry employment, and engagement in advanced studies at other universities.

To predict career decisions, various features beyond demographic data were considered, following previous research. For example, many researchers have shown that learning performance may affect career development in college students. Students with lower achievement might seek a different career from those with higher achievement. Thus, learning performance is worth considering as a factor in career decision (Chen, Chen, Hu, & Wang, 2015; Peng et al., 2020; Park & Lee, 2020). In addition to this, previous studies have found that mathematics knowledge and learning achievement in mathematics-related courses might be particularly closely associated with career planning in a college/university setting (Luzzo et al., 1999). This is because mathematics is a coherent discipline with concepts and topics that have important logical and conceptual connections. It is essential for students as they face challenges (problem) in life, reflecting the coherence of discipline. Students usually need to complete formal or informal mathematical education before engaging in any career path they choose (Mielicki et al., 2019). Mathematics knowledge is required for college and is associated with career readiness (Cogan et al., 2019). Accordingly, features that reflect mathematical ability (e.g., level of coursework, score on general scholastic ability test for mathematics, average scores for calculus-related courses) were taken into account in this study.

In brief, demographic data, learning performance, and mathematical knowledge could predict career decisions in the aggregate (Ferrare, & Miller, 2020; Ketenci et al., 2020; Surrette, 2020). The data used included demographic information, learning performance in university, and career decision, as shown in Figure 1. The demographic information was taken into account as a set of initial features prior to enrolling at the university. The career data included career information for each student after receiving their bachelor's degree for one year. Learning performance included the number of credits selected, the number of credits earned, and the average grade received each year. In addition, learning performance was divided into performance on calculus-related courses and on non-calculus-related courses.

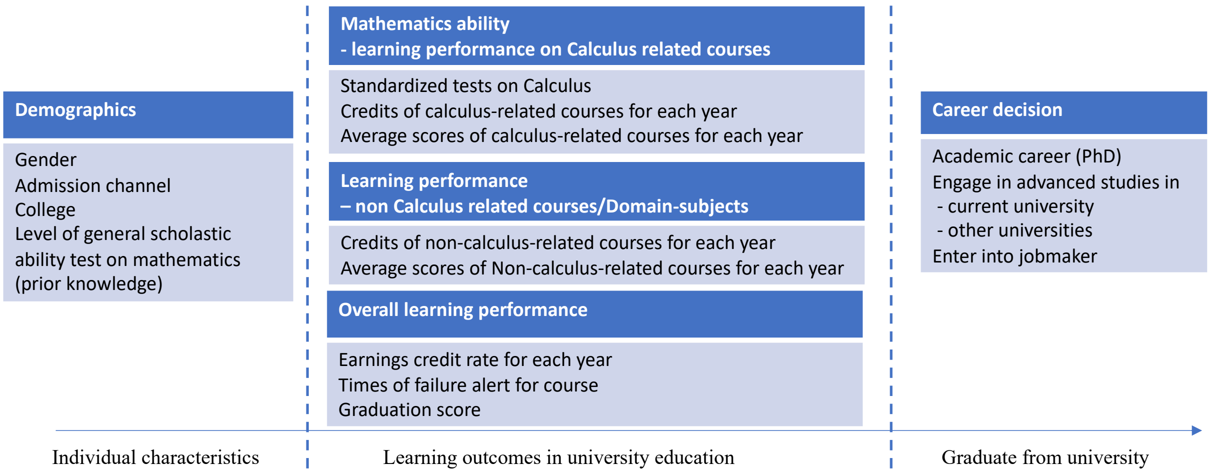

Figure 1. The conceptual model

Calculus-related courses (CRCs) were assessed by two assistants who have both BA and MA degrees in mathematics, and the assessment was validated by a professor who has one-year teaching experience in calculus. A course would be identified as a CRC if (a) calculus was considered basic knowledge in the course (e.g., differential or integral) or (b) if calculus was frequently used for solving problems in the course (e.g., Fourier transformation). Typical calculus-related courses included engineering economics, cryptography, and engineering mathematics. A course not identified as CRC would be classified as an NCRC. All newly recruited students were required to take a standardized test in calculus during the first two semesters, and so the CRC and NCRC values for the first year were replaced by the standardized tests scores for calculus. The overall learning outcomes, including the earnings credit rates for each year, times of failure alert for the course, and graduation score, were incorporated into analysis. All features were processed based on the conceptual model of the features, including calculating values and labelling for specific features. The criteria for labelling features and the schema for the analysed features are given in Tables 1 and 2, respectively. The criteria for labelling features and the schema for the analysed features were given in Tables 1 and 2, respectively.

Table 1. Descriptions of criteria of labelling feature

Feature	Description	Criteria of label	Replaced
Career	Academic career	• Enrolled in the MA program after receiving a BA degree and enrolled in the Ph.D. program after receiving an MA degree • Enrolled in the Five-Year BA-MA Program and Ph.D. program • Enrolled in the Five-Year BA-Ph.D. program • Received a BA degree for one or years and enrolled in the Ph.D. program after receiving an MA degree	AC
	Engage in advanced studies in current university	• Enrolled in the MA program after receiving a BA degree • Enrolled in the MA program after receiving the BA degree for more than one year • Enrolled in the Five-Year BA-MA Program	AS
	Employment	Do not enroll in any further study program and work in a full-time position (according to instance response to Graduation Questionnaire)	EM
	Engage in advanced studies in other universities	Responded that he/she had enrolled a further study program (e.g., MA, PhD.) in other universities	OT
GSATm	Highest level	The original value range of GSAT was a 0-15 ordinal scale, which 15 is the highest score. GSATm was labeled based on the Quartile of the value range obtained from the dataset in this study, i.e., A: GSATm > Q1 B: Q2 ≤ GSATm < Q1 C: Q 3 ≤ GSATm < Q4 D: GSATm ≤ Q4	A
	High level		B
	Middle level		C
	Lowest level		D

Table 2. Structure of the analysed data

Feature	Description	Type of data	Value range of feature
Demographic			
GSATm	Level of general scholastic ability test on mathematics	Ordinal	A: highest, B: high, C: medium, D: low
Gender	Gender of instance	Nominal	0: Female, 1: Male
AC	Admission channel	Nominal	AE: admission by examination and placement, PA: personal application, SP: star Program
College	College of instance	Nominal	ES: College of Science, EE: College of Electrical and Computer Engineering, CS: College of Computer Science, E: College of Engineering, M: College of Management, SS: College of Humanity and Social Science, BT: College of Biological Science and Technology, HS: College of Hakka Studies
Learning performance on Calculus			
STC1	1st Standardized test on Calculus	Numeric	0-100
STC2	2nd Standardized test on Calculus	Numeric	0-100
CRC2 CRC3 CRC4	Credits of calculus-related courses for 2nd/3rd/4th year	Numeric	0-N
CRCs2 CRCs3 CRCs4	Average scores of calculus-related courses for 2nd/3rd/4th year	Numeric	0-100
Learning performance on non-calculus subjects			
NCRC1 NCRC2 NCRC3 NCRC4	Credits of non-calculus-related courses for 1st/2nd/3rd/4th year	Numeric	0-N
NCRCs1 NCRCs2 NCRCs3 NCRCs4	Average scores of Non-calculus-related courses for 1st/2nd/3rd/4th year	Numeric	0-100
Overall learning performance			
ECR1 ECR2 ECR3 ECR4	The earnings credit rate for 1st/2nd/3rd/4th Year	Numeric	0-1
TFA	Times of failure alert for course	Numeric	0-N
GS	Graduation score	Numeric	0-100
Career decision			
CD	Individual career decision	Nominal	AC: Academic, AS: engage in advanced studies in current university, EM: Employment, OT: engage in advanced studies in another university

3.3. Dataset

As seen in the schema, 27 features were selected in this study, and each data row provides these features for a given student. However, no data rows that contained any missing values were taken into consideration and were excluded from the dataset. After this operation, 6255 data rows remained in the dataset.

Admission by examination and placement (AE), personal application (PA), and star program (SP) were three major admission channels that students could use to apply to university. The records we collected in this study only covered students recruited through these admission channels. SP was the latest admission channel launched in Taiwan, and it was used to recruit students from underrepresented high schools. SP provides opportunities for students with good rankings within their senior high schools but who had average uniform high school examination scores. SP is also developed as a means of promoting educational equality for all. The PA channel focuses on scores of nationwide General Scholastic Ability Test (GSAT) and on personal multiple performance at higher schools. That is, students recruited through this channel may have an excellent test scores and have a potential match to one or some of the foci of the universities. On the other hand, students who are not admitted to college via the SP and PA take the Advanced Subjects Test and are then allocated academic majors and universities according to their scores via the AE channel.

Table 3. Career decision of students from different admission channels

Admission channel	Career decision				
	AC	AS	EM	OT	Total
AE	71	805	1657	193	2730
PA	71	1039	1334	203	2649
SP	29	388	390	69	876
Total	171	2232	3381	465	6255

With this, it should be noted that students recruited through these admission channels may have different personal characteristics and backgrounds due to different criteria. Considering the possible heterogeneity of the students from these channels (Lin & Liou, 2019), we first evaluated the distribution (Table 3) by applying the chi-square test ($\chi^2(8) = 104.28$; $p < .001$), and significant differences were found between the career decisions of students from three admission channels. That is to say, students in the three groups exhibited some heterogeneity. For this reason, we generated sub-datasets according to admissions channel, i.e., dataset A: AE, dataset B: SP, and dataset C: SP, to further evaluate the performance of predictive modeling by comparing datasets that split by admission channels and those that did not.

3.4. Implementation of university students' career planning prediction

Many similar precedent studies have examined the relationships between career decisions and individual characteristics of students using conventional statistical models (Chuang & Dellmann-Jenkins, 2010). However, statistical models are usually designed to infer relationships between variables, and machine learning models could be used to maximize predictive accuracy (Tsai et al., 2020). Regarding decisions for practical precision education, recent studies have suggested that supervised machine learning approaches are the most often used to create predictive models for classification and prediction tasks, and these approaches were adopted here as well (Dutt et al., 2017; Rodrigues et al., 2018). Most approaches were model based, driven by an estimation of implicit correlation among analyzed samples and input features and comprising an underlying model that supports the performance of the predictions task. Typical approaches in this area include the artificial neural network (ANN), the support vector machine (SVM), multinomial logistic regression (MLR), decision tree (DT), and rule induction (RI).

In addition to a model-based approach, similarity-based approaches (e.g., k-nearest neighbor, KNN) and probabilistic approaches (e.g., naive Bayes, NB) could be adopted to obtain a comparison of models in a predictive task. The former constitutes the discovery of similar features in provided training data and evaluates the category of a test instance based on the similarity (e.g., the discovery of students which have similar GPA), while the latter exploits probability distribution characteristics that can be observed in training data. To obtain a comprehensive comparison of performance in different predictive models, the seven models described above were also taken into account.

3.5. Evaluation criteria

The confusion matrix was used to describe the performance of a classifier model. The table of a confusion matrix incorporates true predictions and prediction errors. Figure 2 presents a four-class prediction task {A, B, C, D}, where TPA can be defined as the number of observations classified as A where A was predicted, while EAB refers to the number of observations classified as B where A was predicted. The overall accuracy, or kappa (Cohen, 1960), precision, and sensitivity (also called recall) can be calculated to present a measurement for the evaluation of predictive models. The calculation criteria and details of the above measurements are described as follows:

	A	B	C	D
A	TP_A	E_{AB}	E_{AC}	E_{AD}
B	E_{BA}	TP_B	E_{BC}	E_{BD}
C	E_{CA}	E_{CB}	TP_C	E_{CD}
D	E_{DA}	E_{DB}	E_{DC}	TP_D

Figure 2. Illustration of the confusion matrix

- Overall Accuracy = $(TP_A + TP_B + TP_C + TP_D) / (TP_A + TP_B + TP_C + TP_D + E_{AB} + E_{AC} + E_{AC} + E_{BA} + E_{BC} + E_{BD} + E_{CA} + E_{CB} + E_{CD} + E_{DA} + E_{DB} + E_{DC})$
- Kappa = $(P_0 - P_e) / (1 - P_e)$
  where P_0 is the proportion of the observed agreements, P_e is the proportion of agreements expected by chance
- Precision $_{\text{Class A}}$ = $TP_A / (TP_A + E_{AB} + E_{AC} + E_{AD})$
- Sensitivity $_{\text{Class A}}$ = $TP_A / (TP_A + E_{BA} + E_{CA} + E_{DA})$

Here, overall accuracy describes how well a model predicts the career planning of all students investigated. This measure evaluates how often the predictive model is correct. Kappa compares the accuracy of the predictive model with that which a random classification is expected to achieve. The value range is less than or equal to 1, where 1 indicates a perfect prediction by the predictive model, and 0 indicates no better than a random guess. The kappa value is usually considered to measure the success of a predictive model in an imbalanced dataset and thus was taken into account in this study. Precision was defined as the fraction of correct predictions for a certain class. Sensitivity was defined to measure, where a certain class should have been predicted, how many times it was correctly predicted.

To identify the performance of each model for predicting university students' careers, accuracy and kappa were used and precision and sensitivity were used to further evaluate the performance of predictive models for different institutional research purposes.

4. Results and discussions

4.1. Overall performance of the predictive model

The results (Table 4) showed that for high consistency (kappa > 0.7), most models can achieve an accuracy of over 86.76%. That is to say, it is reasonable to expect accurate prediction of university students' career decisions from predictive models having the features proposed in this study. In institutional research, such technologies can be used to provide additional information for the competent authority and stakeholders to address issues related to precision education, such as the arrangement of educational resources and the development of specific strategies. For example, it can predict whether the number of students who choose to enter the job market directly after graduation from university. Educational resources, such as domain-subject curricula, career development instructors, and related programs can be evaluated to determine whether they are sufficient.

The predictive models used in the distinguished dataset demonstrated higher accuracy and consistency than those in the non-distinguished dataset. In other words, the distinguished dataset may produce better performance in predicting career decisions. This result echoed the results of a chi-square test that revealed heterogeneity among the three admissions channel groups. Hence, removing features of the admissions channel improved the accuracy and consistency of the predictive models. These results suggested a need to distinguish different admissions channels for the dataset when performing prediction tasks.

Based on the presented results, it was noticed that by feeding data dataset A (channel: AE) and dataset B (channel: PA), the best performance was obtained by using DT followed by ANN. For feeding dataset C (channel: SP), the best performance was obtained by using ANN followed by DT. That is to say the use of DT and ANN predictive models achieved the highest performance in this prediction task. Most university students in Taiwan were recruited from these three admission channels. This result also suggested that DT and ANN were worth being taken to deal with the career decision prediction tasks.

Table 4. Classifier prediction accuracy and kappa (sorted by accuracy)

Dataset split by admission channel						Original dataset	
Dataset 1 (AE)		Dataset 2 (PA)		Dataset 3 (SP)		Dataset 4 (all channel)	
Classifier	Accuracy/ Kappa	Classifier	Accuracy	Classifier	Accuracy	Classifier	Accuracy
Decision Tree	86.76% / 0.737	**Decision Tree**	86.36% / 0.76	**ANN**	94.75% / 0.911	**Decision Tree**	83.17% / 0.6901
ANN	83.57% / 0.677	**ANN**	78.69% / 0.608	**Decision Tree**	82.99% / 0.703	**ANN**	75.93% / 0.5497
KNN	65.92% / 0.388	KNN	64.34% / 0.373	KNN	72.49% / 0.528	KNN	67.83% / 0.431
Rule Induction	70.10% / 0.355	Rule Induction	66.23% / 0.363	LogisticReg.	68.26% / 0.4385	LogisticReg.	67.05% / 0.3793
LogisticReg.	68.42% / 0.337	LogisticReg.	64.75% / 0.357	Rule Induction	68.26% / 0.4292	Rule Induction	67.63% / 0.3715
SVM	68.12% / 0.309	SVM	64.53% / 0.355	SVM	65.41% / 0.3781	SVM	66.73% / 0.3711
NaiveBayes	49.67% / 0.242	NaiveBayes	43.07% / 0.196	NaiveBayes	42.69% / 0.2309	NaiveBayes	48.46% / 0.2338

Classifier prediction accuracy for DT and ANN over a four-year period was also examined. As shown in Figure 3, increased accuracy was obtained in the third year, and the highest accuracy was found in the fourth year. The accuracy of predictions was evaluated using 10-fold cross-validation, as shown in Table 5. The prediction sought to provide precise learning resources and career guidance to students as earlier as possible, as the earlier the career decision of a student can be identified, the better the support that can be provided. These results indicated that performing such prediction tasks in the third year might be an appropriate time.

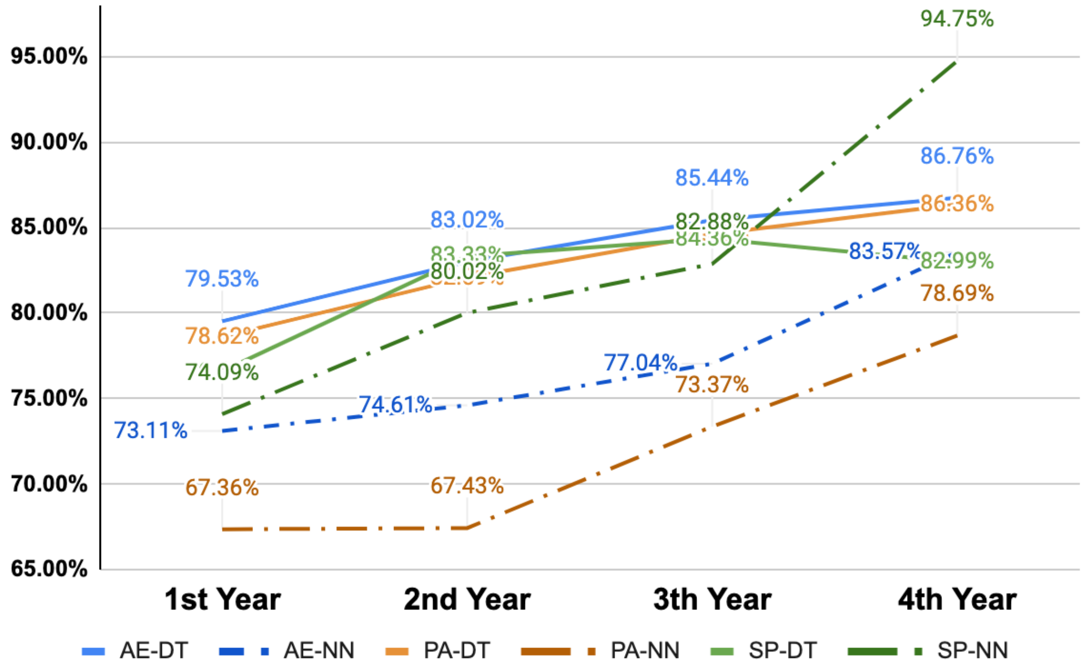

Figure 3. Classifier prediction accuracy of DT and ANN in four years

Table 5. Cross-validation results for DT and NN in each year

		1st year			2nd year			3rd year			4th year		
		AE	PA	SP	AE	PA	SP	AE	PA	SP	AE	PA	SP
DT	Accuracy	62.88%	56.93%	58.33%	62.62%	58.52%	59.36%	64.16%	57.91%	60.27%	64.09%	56.97%	59.70%
	Kappa	0.2451	0.2297	0.2806	0.2526	0.2584	0.2955	0.2703	0.256	0.3122	0.2763	0.2447	0.3026
NN	Accuracy	64.45%	60.94%	58.45%	64.23%	60.48%	55.37%	63.90%	61.69%	56.39%	64.75%	61.16%	59.13%
	Kappa	0.2736	0.2813	0.269	0.2843	0.2776	0.2142	0.2674	0.304	0.2423	0.2961	0.2922	0.2966

4.2. Precision and Sensitivity in university students' career decisions

Moreover, we further explored the precision and sensitivity of DT and ANN. Table 6 represents the confusion matrix for DT and ANN in predicting students' career decisions. Because this is a four-class classification problem, precision and sensitivity were calculated separately for each of the classes.

The precision of the prediction model indicates the rate at which a class was correctly predicted, and the sensitivity indicates that given a certain **c**lass should have been predicted, how many times it actually was correctly predicted. The overall precision for each class ranged from 0.5 to as high as 0.96, while sensitivity ranged from 0.4 to 0.97; however, in some classes, the models appeared to be inadequate or unable to provide predictions.

With the use of ANN, the precision and sensitivity of class AC was 0 for AE and PA students. This suggests that ANN could fail to identify students with a tendency toward enrolment in the PhD program if they were recruited from the PA or AE channel. In other words, educators and stakeholders might need to select a predictive model that is more appropriate for the goal of the prediction task.

Table 6. Confusion matrices for each entrance channel

Decision tree							ANN						
Channel = AE		Actual				Class precision	Channel = AE		Actual				Class precision
		AC	AS	EM	OT				AC	AS	EM	OT	
Predicted	AC	37	12	16	6	**0.889**	Predicted	AC	0	35	31	5	0
	AS	1	654	145	5	0.86		AS	0	693	108	4	0.775
	EM	4	53	1596	4	0.839		EM	0	105	1537	15	0.874
	OT	0	17	98	78	0.881		OT	1	61	83	48	0.667
Class sensitivity		**0.521**	0.812	0.963	0.404		Class sensitivity		0	**0.861**	0.928	0.249	
Channel = PA		Actual				Class precision	Channel = PA		Actual				Class precision
		AC	AS	EM	OT				AC	AS	EM	OT	
Predicted	AC	37	22	12	0	**0.902**	Predicted	AC	0	39	32	1	0
	AS	1	923	107	8	0.841		AS	0	873	165	1	0.778
	EM	2	106	1223	3	0.877		EM	0	125	1209	0	0.794
	OT	1	46	53	103	0.904		OT	117	85	117	1	0.5
Class sensitivity		0.521	0.888	**0.917**	0.507		Class sensitivity		0	0.84	**0.906**	0.005	
Channel = SP		Actual				Class precision	Channel = SP		Actual				Class precision
		AC	AS	EM	OT				AC	AS	EM	OT	
Predicted	AC	8	14	5	2	**0.889**	Predicted	AC	15	6	4	4	0.833
	AS	1	341	45	1	0.859		AS	2	383	3	0	0.946
	EM	0	25	361	4	0.84		EM	1	9	378	54	**0.964**
	OT	0	17	19	33	0.825		OT	1	7	7	54	0.885
Class sensitivity		0.276	0.879	**0.926**	0.478		Class sensitivity		**0.517**	0.986	**0.969**	0.783	

The measures of precision and sensitivity in particular provide valuable information for stakeholders as they select predictive models to deal with questions regarding precision education in relation to institutional research. Many researchers have noted that as students consider which academic or career path they would like to take, they rely on the resources and guidance provided by schools/colleges (Schwartz et al., 2016; Xie & Reider, 2014). More and more, universities are acknowledging that their strong responsibility to guide and support students as they begin their career development (Meijers & Kuijpers, 2014). Thus, correctly predicting career decision making could help educational institutions provide relevant support and resources to help students develop their career plans.

For example, suppose that a university is seeking to provide adequate employee training resources for students (e.g., practical courses, career consulting). A high accurate prediction for students who tend to enter the job market can ensure appropriate delivery of relevant resources and promotions. If this is ensured, career development resources can enjoy maximum utilization.

For a relevant example, supposed a university is seeking to enhance its PhD programs and provide adequate early training resources for students who look for an academic career by enrolling in a PhD program. The more students with such tendencies that can be identified, the better and more effective the early training will become.

In this vein, choosing a model with high sensitivity is particularly appropriate. In a nutshell, high accuracy (greater than 80%) was obtained for most predictive models. DT and ANN demonstrated the highest performance in predicting the career decisions of university students. In relation to precision education, both precision and sensitivity should be considered to help stakeholders choose appropriate predictive models for strategy development.

Conduct-appropriate predictive models might not have the highest contribution. More information related to students' career decisions is required. ANN simulates human nerves in a black-box manner, and so its predictive process is difficult to explain. Thus, little information was provided from this model. However, a decision tree was constructed by calculating estimated measurements (e.g., information gain or the Gini index) among analyzed features. Accordingly, the importance of each feature, classification rule, and predictive process in the DT model was explored. Thanks to this, using the DT model can easily provide a visualized result, namely, a tree-structured graph. This result can not only help educational decision-makers and stakeholders understand the predictive process but also help them extract information from its results (Ellis, 2019). The following section explains how to read and draw out information from DTs.

4.3. Classification rules and information extraction from decision trees

A decision tree is a series of nodes, starting at the base with a single node (root), passing through decision nodes and extending to terminal nodes (leaf nodes) that represent the categories that the tree can classify. It works like a flow chart, choosing a path at each split until a decision is ultimately made at the leaf nodes. These rules provide a detailed explanation for classification queries and extract useful information from DTs. For example, Figure 4 represents the predictive process for dataset A, and we can derive the following rules, given a student recruited by admission examination and placement.
R01: IF [ECRate 2 <= 79.54%], THEN CSL = EM
R02: IF [ECRate 2 <= 79.54%] AND [College = SS, HS], THEN CSL = EM
R03: IF [ECRate 2 <= 79.54%] AND [College = CS] AND [GS<=79.05], THEN CSL = EM

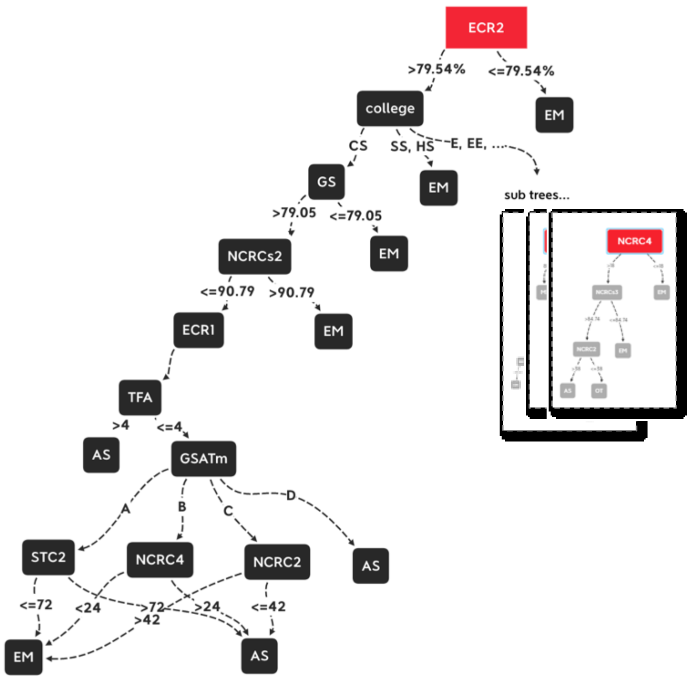

Figure 4. An example of decision tree in predicting career decisions of AE students

Traveling from the root down single or multiple paths, we can break down the tree into smaller and smaller subsets, incremental travel subsets with decision nodes and leaf nodes, to develop a final result. Each decision node has two or more branches and represents a test. The leaf nodes represent predictions (or classifications) for the paths. In this study, the predictions were academic (AC), engaging in advanced study at the current university (AS), employment (EM), and engaging in advanced studies at another university (OT).

Valuable information can be explored in this process. For instance, as shown in Figure 5, the root decision node for ECR2 has two branches, with one greater than 79.54% and ECR2 equal to or less than 79.54%, the student will tend to enter the job market after receiving the BA degree. Traveling down the decision tree for the other two channels (PA and SP), it appears that a student recruited from channel PA or SP tended to enter the job market if their the ECR3 was lower than 91.18% or ECR2 was lower than 72.73%, respectively. These results suggest that the rate at which university students earned credits plays an important role in their career decisions, in particular their rates during their second and third years. Additionally, most students in the College of Social Science (SS) and Hakka Studies (HS) were observed to enter the job market regardless of demographic and learning performance features (e.g., gender, level of general scholastic ability test on mathematics, rate of earning credits, or graduation score).

With regards to institutional research, these results indicated that students who tended to enter the job market after receiving a BA probably demonstrated lower rates of earning credits during the second and third years, except for those in SS and HS. In this vein, an early investigation can be performed to develop a deeper understanding of students' career planning. In this way, institutions can arrange and provide adequate resources for students to help them develop their career plans and achieve their goals.

Likewise, the time of failure alert (TFA), level of general scholastic ability test on mathematics (GSATm), and non-calculus-related credit in the fourth year (NCRC4) for students are useful statistics to identify students who would will seek to apply to a master's program, regardless of admission channels. Accordingly, more academic training resources (e.g., advanced courses for domain subjects) can be arranged and presented to students.

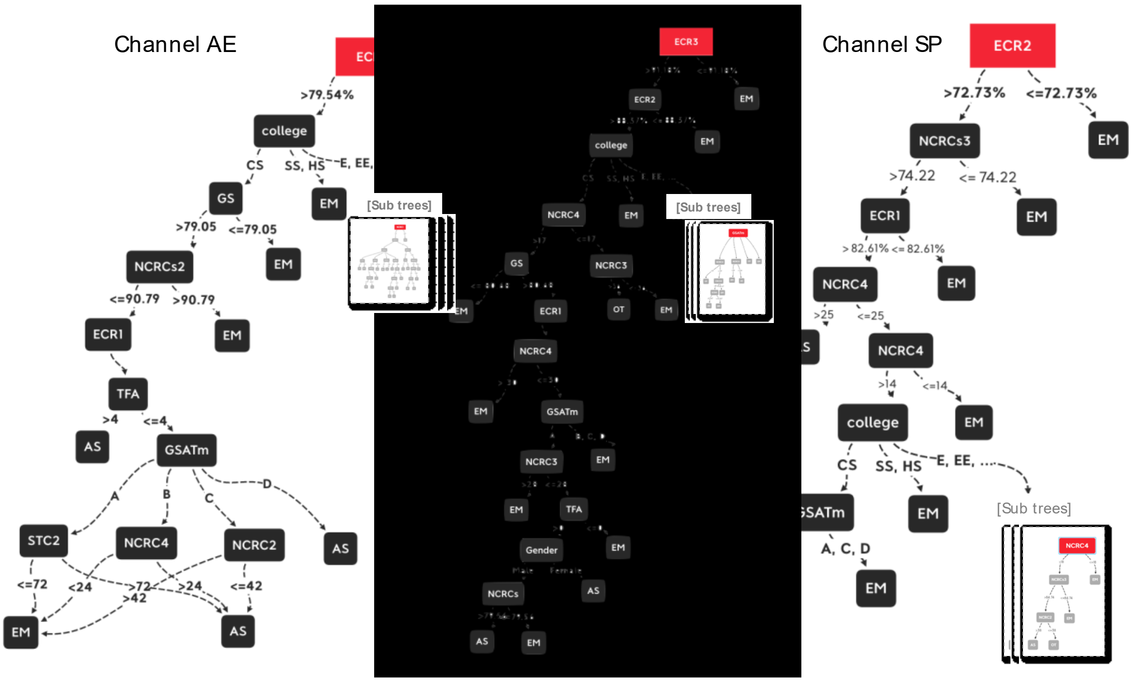

Figure 5. An example of decision tree for students recruited from three admission channel

4.4. Classification rules and information extraction from the decision table

The classification rules and information can also be represented as decision tables to facilitate decision making and promote precision education strategies. For example, Table 7 represents a decision table of classification rules to identify students who might seek to enter a master's program at their current university. It can be found that for the college of management, the GSATm, NCRC1, and NCRC4 were important features to identify whether a student had a tendency to apply to a master's program. On the other hand, NCRC2, NCRC3, NCRC4, and STC1 were useful for identifying students who were willing to enroll in a graduate institute in the current university for the college of biological science and technology. These findings suggested that it is possible to investigate the career decisions of students at different time points for different colleges. Specifically, the mathematical prior knowledge (i.e., GSATm), credits enrolled in the first year and last year before they graduate were closely related to the tendency of career decision for college of management.

Prior mathematical knowledge and credits enrolled in the first year might not significantly predict students' career decisions in the college of biological science and technology. Tracing learning performance for calculus and how many credits were enrolled in each year after the second year in university were needed to evaluate whether a student tends to engage in advanced studies in the current university. Thus, various features can be considered to investigate students' career decisions in different contexts to support help educational decision-makers and stakeholders as they evaluate precision education strategies for their institutions.

The previous sections have demonstrated a serial process of extracting information that is valuable for the practice of precision education using an institutional research perspective. For example, based on the above results, early prediction can be performed during the first year of study to allow the institution to obtain a better picture of how many students would engage in advanced studies at the university. In this way, teaching and learning resources can be evaluated for freshmen and promoted precisely in relation to how many credits students had enrolled in the previous year. To facilitate practical and interpretable applications of precision education, we presented a conceptual framework (Figure 6) for the research community in this field to explore the practice of precision education based on the comprehensive process proposed in this study.

Table 7. Examples of decision tables

ECR2	College	NCRC1	NCRC4	GSATm	STC1	NCRC3	NCRC2	CD
<= 0.795								EM
> 0.795	HC、SS							EM
	M	<=43						EM
		>43	<=21					EM
			>21	B or C				AS
				A				EM
		<=44		D				EM
		>44						AS
	BS		<=21					EM
			>21		> 67			EM
					<=67	<=34		EM
						>34	>41	AS
				> 59			<=41	AS
					<=59			EM

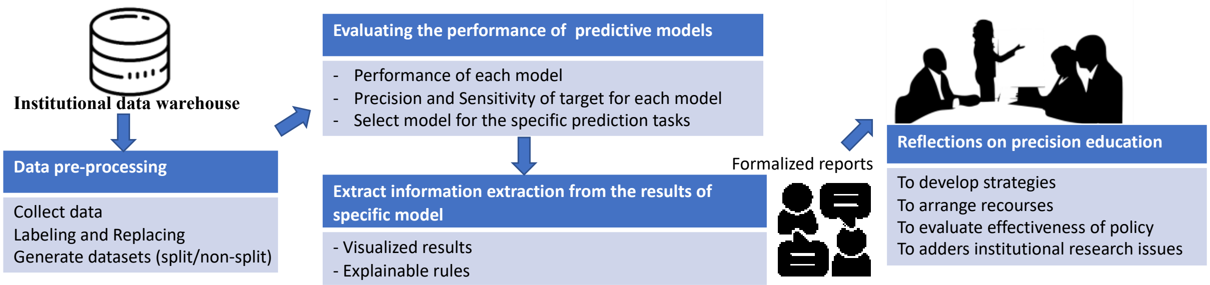

Figure 6. Conceptual framework of prediction tasks from the perspectives of institutes

5. Conclusions

Questions in institutional research for higher education have been transformed from reactive to proactive in decision and policymaking in the past decade. The detection and accommodation of individual needs play an important role in higher education because it presents an opportunity to address precision education. Among the various individual needs, career development and training are attracting attention as beneficiaries of the accurate prediction of students' career decisions.

To answer our first research question, various machine learning approaches were used as predictive models to perform predictive analyses to determine the career decisions of students with consideration for their demographics, mathematical ability, and overall learning performance. A maximum accuracy of 94.75% was achieved, and a high accuracy, kappa, precision, and sensitivity were obtained for most predictive models in this study. The results indicate that students' career decisions are predictable with reasonable accuracy.

In the pursuit of the second research question, this study sought to stimulate future research that can promote the practice of precision education. The results of this study show that the institutional database can be used to obtain precise information on the individual needs of students. This can enable specific strategies or policies to be proposed from an institutional research perspective. In this way, institutions can develop more precise interventions for their students and enable their individual needs to be met in order to reach the goal of precision education.

Overall, this study's contributions fall under three categories: theory, methodology, and application. First, it deepens the theoretical understanding of the relationships among students' demographic, mathematical knowledge, learning performance, and career decision aspects. Specifically, the predictive model identified by this study revealed that using datasets distinguished by admission channels can produce maximum performance for prediction tasks, suggesting that students recruited along different admission channels were heterogeneous. Moreover, the rates of earning credit for the second or third year and prior mathematical knowledge (i.e., GSATm) were critical features in predicting students' career decisions, regardless of their admissions channel. These findings echo previous studies, which indicated that mathematical knowledge could be related to students' career aspirations and decisions (Lazarides et al., 2020).

Second, for methodology, this study performed prediction tasks by deploying seven commonly used supervised machine learning approaches. These approaches successfully extracted relationships from students' features and career decisions. The results of this study suggested that the proposed framework (including its data structure) is appropriate for predicting students' career decisions.

Finally, a comprehensive analysis and comparison of the application of major machine learning techniques were performed in this study. It was shown that particular machine learning techniques provided for optimized prediction of different targets or purposes (e.g., specific career decisions of students in different colleges) and additional information could be obtained according to a predictive model, such as drawing decision trees or creating decision tables.

Like precision medicine, optimal precision education is a long way off (Reardon & Stuart, 2017). Emerging educational data mining techniques can be included in future analysis and comparison, such as deep learning approaches. Moreover, adequate datasets for learning data that contain a diversity of courses or learning activities and individual characteristics (e.g., motivation, interest, and preferences) are worth being made available to assemble advanced frameworks and perform prediction tasks. It would be valuable to determine whether such educational applications can help online teachers by applying the insights of precision education.

Acknowledgement

This study is supported by the Ministry of Science and Technology of the Republic of China, Taiwan, under contract numbers MOST-108-2511-H-152-015-MY2. The authors also gratefully thank Big Data Research Center (BDRC) of National Chiao-Tung University for allowing to use its data.

References

Abdullah, N., Hussin, N., Shonubi, O. A., Ghazali, S. R., & Talib, M. A. (2018). Career decision-making competence, self-knowledge, and occupational exploration: A Model for university students. *Journal of Technical Education and Training*, *10*(1), 71-81.

Adekitan, A. I., & Salau, O. (2019). The Impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, *5*(2), e01250. doi:10.1016/j.heliyon.2019.e01250

Arthars, N., Dollinger, M., Vigentini, L., Liu, D. Y. T., Kondo, E., & King, D. M. (2019). Empowering teachers to personalize learning support. In D. Ifenthaler, D. K. Mah, & J. K. Yau (Eds.), *Utilizing learning analytics to support study success* (pp. 223-248). Cham, Switzerland: Springer.

Bahr, P. R. (2009). College hopping: Exploring the occurrence, frequency, and consequences of lateral transfer. *Community College Review, 36*(4), 271–298.

Bahr, P. R. (2010). The Bird's eye view of community colleges: A Behavioral typology of first-time students based on cluster analytic classification. *Research in Higher Education, 51*(8), 724–749.

Beemer, J., Spoon, K., Fan, J., Stronach, J., Frazee, J. P., Bohonak, A. J., & Levine, R. A. (2018). Assessing instructional modalities: Individualized treatment effects for personalized learning. *Journal of Statistics Education*, *26*(1), 31-39.

Bellgard, M. I. (2020). ERDMAS: An Exemplar-driven institutional research data management and analysis strategy. *International Journal of Information Management*, 50, 337-340.

Brunila, K. (2020). Interrupting psychological management of youth training. *Education Inquiry*, *11*(4), 302-315.

Burns, M. K., Codding, R. S., Boice, C. H., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review*, *39*(1), 69-83.

Caison, A. L. (2007). Analysis of institutionally specific retention research: A Comparison between survey and institutional database methods. *Research in Higher Education*, *48*(4), 435-451.

Camacho, D. J., & Legare, J. M. (2016). Shifting gears in the classroom—Movement toward personalized learning and competency-based education. *The Journal of Competency-Based Education*, *1*(4), 151-156.

Chen, S. Y., & Wang, J. H. (2020). Individual differences and personalized learning: A Review and appraisal. *Universal Access in the Information Society*, 1-17.

Cheng, T. M., Hou, H. Y., Agrawal, D. C., Liaw, C. F., & Chen, R. C. (2018). Remedial education for high-risk university freshmen – The Case of a university of technology in Taiwan. *Asia Pacific Journal of Education,* 38, 241-256.

Christensen, C. M., & Eyring, H. J. (2011). *The Innovative university: Changing the DNA of higher education from the inside out*. San Francisco, CA: John Wiley & Sons.

Cogan, L. S., Schmidt, W. H., & Guo, S. (2019). The Role that mathematics plays in college-and career-readiness: Evidence from PISA. *Journal of Curriculum Studies*, *51*(4), 530-553.

Collins, F. S., & Varmus, H. (2015). A New initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793–795.

Collins, H., Calvo, S., Greenberg, K., Neall, L. F., & Morrison, S. (2016). Information needs in the precision medicine era: How genetics home reference can help. *Interactive Journal of Medical Research*, *5*(2), e13. doi: 10.2196/ijmr.5199

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology*, 66, 4-10.

Daghestani, L. F., Ibrahim, L. F., Al-Towirgi, R. S., & Salman, H. A. (2020). Adapting gamified learning systems using educational data mining techniques. *Computer Applications in Engineering Education*, *28*(3), 568-589.

Du, X., Yang, J., & Hung, J. L. (2020). An Integrated framework based on latent variational autoencoder for providing early warning of at-risk students. *IEEE Access*, *8*, 10110-10122.

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic review on educational data mining. *IEEE Access*, *5*, 15991-16005.

Ellis, L. (2019). Artificial intelligence for precision education in radiology–Experiences in radiology teaching from a UK foundation doctor. *The British Journal of Radiology*, *92*(1104), 20190779. doi:10.1259/bjr.20190779

Fan, J. (2016). The Role of thinking styles in career decision-making self-efficacy among university students. *Thinking Skills and Creativity*, *20*, 63-73.

Ferrare, J. J., & Miller, J. M. (2020). Making sense of persistence in scientific purgatory: A Multi-institutional analysis of instructors in introductory science, technology, engineering, and mathematics (STEM) courses. *The Journal of Higher Education*, *91*(1), 113-138.

Flaig, M., Simonsmeier, B. A., Mayer, A. K., Rosman, T., Gorges, J., & Schneider, M. (2018). Reprint of "Conceptual change and knowledge integration as learning processes in higher education: A Latent transition analysis." *Learning and Individual Differences*, *66*, 92-104.

Freeman, V. F., Lenz, J. G., & Reardon, R. C. (2017). Career course impact on college students' career decision and affective states. *VISTAS Online*, *32*(1), 1-14.

Gulson, K. N., & Webb, P. T. (2017). Mapping an emergent field of "computational education policy": Policy rationalities, prediction and data in the age of Artificial Intelligence. *Research in Education*, *98*(1), 14-26.

Hart, S. A. (2016). Precision education initiative: Moving toward personalized education. Mind, *Brain, and Education, 10*(4), 209–211.

Healy, M., Hammer, S., & McIlveen, P. (2020). Mapping graduate employability and career development in higher education research: a citation network analysis. *Studies in Higher Education*, 1-13.

Hoffait, A. S., & Schyns, M. (2017). Early detection of university students with potential difficulties. *Decision Support Systems*, *101*, 1-11.

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on psychological science*, *6*(1), 21-37.

Ketenci, T., Leroux, A., & Renken, M. (2020). Beyond student factors: A Study of the impact on stem career attainment. *Journal for STEM Education Research*, *3*, 368-386.

Kim, E., Benson, S., & Alhaddab, T. A. (2018). A Career in academia? Determinants of academic career aspirations among PhD students in one research university in the US. *Asia Pacific Education Review*, *19*(2), 273-283.

Kokkinos, C. M., & Saripanidis, I. (2017). A Lifestyle exposure perspective of victimization through Facebook among university students. Do individual differences matter? *Computers in Human Behavior*, 74, 235-245.

Kuch, D., Kearnes, M., & Gulson, K. (2020). The Promise of precision: datafication in medicine, agriculture and education. *Policy Studies, 41*(5), 1-20.

Laanan, F. S., & Jain, D. (2016). Advancing a new critical framework for transfer student research: Implications for institutional research. *New Directions for Institutional Research*, *170*, 9-21.

Laguilles, J. S. (2016). Collecting and using postgraduate outcomes data at a private college. *New Directions for Institutional Research*, *169*, 25-36.

Lazarides, R., Dicke, A. L., Rubach, C., & Eccles, J. S. (2020). Profiles of motivational beliefs in math: Exploring their development, relations to student-perceived classroom characteristics, and impact on future career aspirations and choices. *Journal of Educational Psychology*, *112*(1), 70-92.

Lin, J. J., & Liou, P. Y. (2019). Assessing the learning achievement of students from different college entrance channels: A Linear growth curve modelling approach. *Assessment & Evaluation in Higher Education*, *44*(5), 732-747.

Lonn, S., & Koester, B. (2019). Rearchitecting data for researchers: A Collaborative model for enabling institutional learning analytics in higher education. *Journal of Learning Analytics*, *6*(2), 107-119.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of Students' academic performance in blended learning. *Educational Technology & Society*, *21*(2), 220-232.

Makhluf, H. A. (2020). Precision education: Engineering learning, relevancy, mindset, and motivation in online environments. In *Exploring Online Learning Through Synchronous and Asynchronous Instructional Methods* (pp. 202-224). Hershey, PA: IGI Global.

Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A Case study with high school students. *Expert Systems*, *33*(1), 107-124.

Mielicki, M. K., Martinez, M. V., DiBello, L. V., Lee-Hassan, A. W., & Pellegrino, J. W. (2019). Assessing mathematics knowledge and skill: What college students actually know and can do? In O. Latkin-Troitschanskaia (Eds.), *Frontiers and Advances in Positive Learning in the Age of InformaTiOn (PLATO)* (pp. 253-265). Wiesbaden, Germany: Springer.

Miller, A. K., Osborn, D. S., Sampson Jr, J. P., Peterson, G. W., & Reardon, R. C. (2018). The Impact of a college career course on students' career decision states. *The Career Development Quarterly*, *66*(4), 371-377.

O'Connor, M. A., & Daly III, E. J. (2018). Selecting effective intervention strategies for escape-maintained academic-performance problems: Consider giving'em a break! *Journal of school psychology*, *66*, 41-53.

Palanica, A., Docktor, M. J., Lee, A., & Fossat, Y. (2019). Using mobile virtual reality to enhance medical comprehension and satisfaction in patients and their families. *Perspectives on Medical Education*, *8*(2), 123-127.

Pardo, A. (2018). A Feedback model for data-rich learning experiences. *Assessment & Evaluation in Higher Education*, *43*(3), 428-438.

Peng, Y., & Wang, Q. (2020). The Impact of mindful agency coaching and motivational interviewing on the development of positive learning dispositions in undergraduate students: A Quasi-experimental intervention study. *Journal of Educational and Psychological Consultation*, *30*(1), 63-89.

Reardon, S. F., & Stuart, E. A. (2017). Editors' introduction: Theme issue on variation in treatment effects. *Journal of Research on Educational Effectiveness, 10*(4), 671-674.

Reed, D. K., & Gates, C. (2020). Don't fail to plan for summer reading interventions. *Preventing School Failure: Alternative Education for Children and Youth, 64*(3), 223-229.

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A Cross-institutional comparison across 151 modules. *Computers in Human Behavior*, *60*, 333-341.

Rodrigues, M. W., Isotani, S., & Zárate, L. E. (2018). Educational data mining: A Review of evaluation process in the e-learning. *Telematics and Informatics*, *35*(6), 1701-1717.

Ross, S. M., & Morrison, G. R. (2012). Constructing a deconstructed campus: Instructional design as vital bricks and mortar. *Journal of Computing in Higher Education*, *24*(2), 119-131.

Rump, M., Esdar, W., & Wild, E. (2017). Individual differences in the effects of academic motivation on higher education students' intention to drop out. *European Journal of Higher Education*, *7*(4), 341-355.

Santos, A., Wang, W., & Lewis, J. (2018). Emotional intelligence and career decision-making difficulties: The Mediating role of career decision self-efficacy. *Journal of Vocational Behavior*, *107*, 295-309.

Surrette, T. N. (2020). Influence of mentoring and professional communities on the professional development of a cohort of early career secondary mathematics and science teachers. *School Science and Mathematics*, *120*(3), 175-188.

Swing, R. L., & Ross, L. E. (2016). A New vision for institutional research. *Change: The Magazine of Higher Learning*, *48*(2), 6-13.

Thompson, M. N., Her, P., Fetter, A. K., & Perez-Chavez, J. (2019). College student psychological distress: Relationship to self-esteem and career decision self-efficacy beliefs. *The Career Development Quarterly*, *67*(4), 282-297.

Ulas, O., & Yildirim, I. (2019). Influence of locus of control, perceived career barriers, negative affect, and hopelessness on career decision-making self-efficacy among Turkish university students. *International Journal for Educational and Vocational Guidance*, *19*(1), 85-109.

Valladares, M., Ramírez-Tagle, R., Muñoz, M. A., & Obregón, A. M. (2018). Individual differences in chronotypes associated with academic performance among Chilean University students. *Chronobiology international*, *35*(4), 578-583.

Warren, J. J., Clancy, T. R., Delaney, C. W., & Weaver, C. A. (2017). Big-data enabled nursing: Future possibilities. In *Big Data-Enabled Nursing* (pp. 441-463). New York, NY: Springer.

Williamson, B. (2019). Digital policy sociology: Software and science in data-intensive precision education. *Critical Studies in Education*, 1-17.

Williamson, B. J. (2020). New digital laboratories of experimental knowledge production: Artificial intelligence and education research. *London Review of Education*, *18*(2), 209-200.

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, *47*, 168-181.

Yoo, J. (2019). A Study on AI education in graduate school through IPA. *Journal of The Korean Association of Information Education*, *23*(6), 675-687.

Zawacki-Richter, O., Juarros, V. M., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 1-27. doi:10.1186/s41239-019-0171-0