Decision Support Systems 134 (2020) 113320

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Uplift Modeling for preventing student dropout in higher education

Diego Olaya[a,*], Jonathan Vásquez[b], Sebastián Maldonado[c,d], Jaime Miranda[c,d], Wouter Verbeke[a]

[a] Data Analytics Laboratory, Faculty of Social Sciences & Solvay Business School, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[b] Escuela de Auditoria, Universidad de Valparaíso, Chile
[c] Department of Management Control and Information Systems, School of Economics and Business, University of Chile, Santiago, Chile
[d] Instituto Sistemas Complejos de Ingeniería (ISCI), Chile

ARTICLE INFO

Keywords:
Learning analytics
Uplift modeling
Student dropout
Educational data mining

ABSTRACT

Uplift modeling is an approach for estimating the incremental effect of an action or treatment at the individual level. It has gained attention in the marketing and analytics communities due to its ability to adequately model the effect of direct marketing actions via predictive analytics. The main contribution of our study is the implementation of the uplift modeling framework to maximize the effectiveness of retention efforts in higher education institutions i.e., improvement of academic performance by offering tutorials. The objective is to improve the design of retention programs by tailoring them to students who are more likely to be retained if targeted. Data from three different bachelor programs from a Chilean university were collected. Students who participated in the tutorials are considered the treatment group, otherwise, they are assigned to the nontreatment group. Our results demonstrate the virtues of uplift modeling in tailoring retention efforts in higher education over conventional predictive modeling approaches.

1. Introduction

Student dropout is a genuine concern in private and public institutions in higher education because of its negative impact on the well-being of students and the community in general. Early desertion in undergraduate programs not only causes monetary losses to educational institutions in terms of tuition fees paid either by the students or by the state through scholarships, but also social costs.

Universities have focused in the past years in the design of retention campaigns as a means to prevent student withdrawal. Dropout arises from different context-specific academic and nonacademic factors. Academic achievement and institutional habitus [1], as well as demographics, social interaction, financial constraints, motivation and personality, play a vital role [2]. Each risk factor can be addressed in numerous ways such as academic assistance (e.g., tutoring, counselling and mentoring [3,4]), social engagement and individual attachment to the institution [5–7], purpose for completing school (e.g., vocational education, part-time job placements, internships), and financial assistance.

The success of retention campaigns is subject not only to appropriately understanding the factors associated to student attrition, but also to accurately identifying and targeting students who are most likely to respond to interventions. Unlike research on factors related to student withdrawal, this paper aims to extend the current student dropout literature by introducing uplift modeling as a decision-making tool to support the design of student dropout prevention strategies.

Uplift modeling is a predictive analytics technique that estimates the effect of a *treatment* on the behavior of an individual. Conventional predictive models for churn prediction aim at identifying and targeting individuals who are more likely to attrite. However, targeting on the basis of risk does not consider that each individual responds differently to retention strategies, as risk of dropping out and sensitivity to the intervention are not necessarily related [8].

This study proposes a novel framework for preventing student attrition using uplift modeling. Our main contributions can be summarised as follows:

- We apply uplift modeling to the student dropout problem. Two special considerations are made: addressing self-selection bias and the low risk of triggering student attrition. The former refers to the design of the retention campaign as the university makes an open call to all students who want to participate. The latter alludes to the low risk of targeting students with tailored programs.
- Using data from three bachelor programs from a Chilean university, we demonstrate the virtues of our proposal over conventional predictive modeling. The university currently invites all students to a

* Corresponding author.
E-mail addresses: diego.olaya@vub.be (D. Olaya), jonathan.vasquez@uv.cl (J. Vásquez), sebastianm@fen.uchile.cl (S. Maldonado), jmirandap@fen.uchile.cl (J. Miranda), wouter.verbeke@vub.be (W. Verbeke).

https://doi.org/10.1016/j.dss.2020.113320
Received 22 January 2020; Received in revised form 30 April 2020; Accepted 30 April 2020
Available online 11 May 2020
0167-9236/ © 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

program designed to improve academic performance and engagement. Although the program has a positive effect on student retention, we show the benefits of designing a customized program i.e., targeting only students who are the most likely to be retained by the program.
- Model comprehensibility is addressed by segmenting students according to their estimated uplift, and later observing the characteristics of each segment. This allows us to gain insight into the application and develop better retention policies. The analysis of the variables that are relevant for defining a segment of students to target is discussed in the empirical section.

The remainder of this paper is organized as follows: A literature review on student dropout is presented in Section 2. The proposed framework for student retention using uplift modeling is presented in Section 3. Experimental results are given in Section 4. Finally, Section 5 provides the main conclusions, while also addressing future developments.

2. Prior work on student dropout

Student dropout in higher education has been studied for several years. Dropout occurs when an individual enrolled at an educational institution decides to voluntarily abandon studies [9,10].

The foundations of the research on student dropout are established since the 70's and 80's by [9–12], whose approaches are still used nowadays as a starting point for new developments [13–17]. [9] proposed an interdisciplinary approach by using psychological variables to model student attrition as the interaction between a student and the educational environment. Later, [10] proposed a parsimonious model that reflects the relation between prematriculation attributes and the interaction with the environment (academic and social systems). Last, [12] extended the former approaches by incorporating additional elements related to the interaction between the students and the educational institution.

The early work on student dropout stimulated different research approaches that were widely discussed in the last decade. A first research stream refers to the time of attrition, since associated attributes may vary throughout the academic program [16]. That is, the scope may include freshman to sophomore years [18], from sophomore to junior years [19] or different predetermined periods [17]. By contrast, other studies do not focus on the time perspective, but rather analyze the dropout phenomenon from a systematic standpoint [20].

Furthermore, studies discussing factors associated with dropping out, such as the influence of socioeconomic determinants, have yielded contradictory results. While [15] confirmed the influence of gender in student attrition [21] did not find a significant relationship. Similarly, [17] claimed that low-income students are less likely to abandon their bachelor program, in contrast to [22], which suggested that this group has a higher risk of churn. These divergences may indicate that student dropout depends largely on contextual elements.

Predicting student dropout via statistical and machine learning techniques has gained increasing attention, leading to a rich research area known as learning analytics or educational data mining (EDM) [23]. Recent studies on student dropout mainly focus on applications of different machine learning techniques for this task. Examples include semi-supervised learning [24], unsupervised learning [25], and ensemble learning [26].

As mentioned above, the majority of machine learning applications on student dropout prevention have focused on identifying students with high propensity to attrite, i.e., targeting those at high risk. However, to the best of our knowledge, no study has focused on customizing the assignment of retention actions to students on the basis of their expected sensitivity to an intervention. Therefore, we propose uplift modeling as a tool to support the design of student dropout prevention strategies.

3. The uplift modeling framework and student dropout

This section formally defines uplift modeling, and subsequently presents the main approaches for estimating and evaluating uplift models. At the end, we discuss some considerations that must be addressed when applying uplift modeling to the context of student dropout.

3.1. Uplift modeling

Uplift modeling is a predictive analytics technique that estimates the individual treatment effect (ITE), i.e., the effect of an action or treatment on an outcome of interest. This task differs from the estimation of average causal effects, since it considers that causal effects vary with observable characteristics. Uplift modeling is analogous to the problem of treatment effect heterogeneity [27] and individualized treatment rule estimation [28], as it aims to determine the degree to which treatments have differential causal effects on individuals. The goal is to customize the assignment of treatments by prescribing the action that maximizes a given objective. Therefore, uplift models identify individuals for whom the exposure to an action is expected to lead to a favorable outcome.

Uplift modeling has been applied in a wide variety of domains. Initial applications mainly focused on maximizing the effectiveness of marketing campaigns [29–32]. Nonetheless, uplift modeling has received also attention in the fields of personalized medicine [33], and price optimization [34,35].

Formally, uplift models predict the ITE in terms of the potential outcomes framework [36]. The problem consists of learning based on a sample of N students independent and identically distributed, whether student i should be treated (i.e., take the tutorials), given the set of pretreatment characteristics, X_i, the binary indicator of treatment, $T \in \{0,1\}$, i.e., $T = 1$ meaning treatment, and the binary outcome variable, $Y \in \{0,1\}$, where $Y = 1$ represents no attrition. The potential outcomes, $Y_i^{(1)}$ and $Y_i^{(0)}$, are the future states of the outcome for the ith student with and without the treatment, respectively. Then, the ITE of treatment against nontreatment on Y is the difference between the two potential outcomes, $Y_i^{(1)} - Y_i^{(0)}$. Since the ITE varies with observables characteristics, it can be defined in terms of the conditional average treatment effect (CATE),

$$CATE_i\text{: } ITE_i\text{: } \tau_i \coloneqq \widehat{P}\left(Y_i \mid X_i, do(T=1)\right) - \widehat{P}\left(Y_i \mid X_i, do(T=0)\right). \quad (1)$$

Eq. (1) defines the $CATE_i$ as a comparison between the conditional likelihoods of no attrition under two different regimes, i.e., treatment and nontreatment. Although predictive modeling consists of estimating outcomes as a function of observed variables, the uplift modeling task is not to predict the outcome variable, but its variation due to the treatment. The $do(\cdot)$ operator [37] is commonly used in causal calculus to indicate that $T = t$ denotes an intervention, i.e., interventional conditional distribution, rather than the observed values taken by T, i.e., observational conditional distribution. Uplift modeling employs machine learning techniques to estimate the potential outcomes. Hence, the difference of the two conditional probabilities is a continuous score known as the uplift score, $\widehat{\tau}_i$. Students whose $\widehat{\tau}_i > 0$ are considered treatment responders and, therefore, should be targeted.

We estimate the CATEs under certain assumptions, since the tutorials are not assigned at random. A first assumption is that the treatment assignment is as good as random once we control for the observed variables, that is, under unconfoundedness, the potential outcomes are independent of the treatment conditional on the observed variables, $Y^{(T)} \perp \perp T \mid X$ [38]. A second assumption, common support, guarantees that the conditional treatment probability is non-zero, $P(Treatment|X) > 0$ which is necessary to find appropriate matches of treated and untreated students. Finally, satisfying the stable unit treatment value assumption (SUTVA) [39] ensures that the potential outcomes are

not influenced by treatments given to other students.

The uplift modeling literature distinguishes between two approaches to estimate the uplift: data preprocessing and data processing [40]. The data preprocessing approach consists of modifying, prior to training, (1) the outcome variable or (2) the input space. In contrast, the data processing approach alludes to methods that estimate the uplift indirectly or directly. Uplift is estimated indirectly when two separate predictive models (i.e., one for the treated and one for the untreated) are trained. The direct estimation refers to modified machine learning algorithms that have been adapted to directly estimate the uplift.

The modified outcome approach (MOA) was introduced by [41,42] to estimate conditional average treatment effects on the treated based on the difference-in-differences estimator and the ITE, respectively. In the uplift modeling literature this approach was presented by [43] and have been extended by [29,44,45]. For example, the MOA by [43] relabels the outcome variable by taking into account the four types of individuals present in the data set: treatment responders (TR), treatment nonresponders (TN), nontreatment responders (NR) and nontreatment nonresponders (NN). The transformed outcome variable considers TR and NN as positive cases, since these individuals are positively affected by the treatment. In contrast, NR and TN are regarded as negative cases, since these individuals either have a unfavourable or no response to the treatment. Hence, the uplift modeling problem is reduced to a conventional classification model, with the uplift computed as follows:

$$\hat{\tau}_i := \hat{P}\,(TR \vee NN \mid X_i) - \hat{P}\,(NR \vee TN \mid X_i). \tag{2}$$

Although the main advantage of the MOA is that it can use existing learning methods entirely off-the-shelf, the approach can also be inefficient, since the information of the treatment indicator is not used more than for the construction of the transformed outcome [46].

The second preprocessing method is the modified covariate approach (MCA), also referred to as S-learner [47] or R-learner [48]. The MCA was presented by [49] to model interactions between the treatment indicator and the observed pretreatment variables. In the uplift literature, Lo [32] proposed to train a single predictive model that includes within the input space aside from the pretreatment variables, the indicator of treatment as a dummy, D, and interaction terms between the dummy and the pretreatment variables, $D \times X$. So that, $\hat{P}\,(Y \mid X, do(T = k)) = f(X, D, D \times X)$. A potential drawback of the MCA is that the enlargement of the input space can result in multicollinearity problems [29].

Estimating uplift indirectly refers to the separate model approach (SMA), also referred to as T-learner [47,48] or Q-learner [28]. The SMA is the most intuitive method, as the predictive models learned on the treated and untreated are used to predict the uplift score of each test case based on the conditional probabilities of treatment and nontreatment. This methodology is simple and implements standard machine learning techniques, but can be suboptimal since the modeling objective of the two predictive models is not to estimate directly the uplift [50]. Nonetheless, [45] concludes that the SMA may perform competitively under certain conditions.

Methods to model uplift directly aim to offset the main drawbacks of the previous approaches. The objective is to implement a single training scheme to estimate the uplift by adapting the objective function of conventional machine learning algorithms (a complete overview is given by [40,44]). The literature on heterogeneous treatment effects has proposed the causal tree [46], causal bayesian additive regression trees (BART) [51], causal forest [52], causal boosting [53], and generalized random forest [54] algorithms with modified splitting procedures that partition the data according to treatment effect heterogeneity.

Similarly in the uplift literature adapted K-nearest neighbour classifiers are proposed by [55,56]. Modifications to the splitting and pruning criteria of decision tree classifiers are found in [31,50,57,58]. Modified random forest algorithms are suggested to offset the instability of a single decision tree by [30,34,35]. For example, [59] employs a modified random forest algorithm to estimate the effect of motivational e-mail campaigns. Last, [60] presented a support vector machine for uplift modeling and [61] proposed a reinforcement learning approach.

The evaluation of uplift models cannot be performed by means of a loss function as conventional predictive models, due to the fundamental problem of causal inference [62]. The difficulty relates to the impossibility of observing the true effect of the treatment for each student. The data set, however, is split into a training set and a test set preserving the proportions of treated and untreated students. The uplift model is constructed on the training set, and later, the model is applied to the test set, to obtain the potential outcomes for each test case. The predicted uplift score is then computed as illustrated in Eq. (1). Subsequently, test cases are ranked according to the predicted uplift score in descending order. Last, test cases are segmented in groups of equal size (i.e., bins) and the segment-wise treatment effect is calculated as the difference between the response rates of treated and untreated subjects. The intuition behind this approach is that a model with an outstanding performance is expected to allocate in top segments students whose propensity to attrite will be reduced subject to participating in the tutorials.

Formally, test cases $\mathbb{D}$ are ranked in descending order according to their uplift scores, $\hat{\tau}$. Let $\pi_k \in \mathbb{D}$ be the k segment of test cases in $\mathbb{D}$, so that the amount of treated and untreated test units within the segment can be calculated respectively as follows:

$$N^{(1)}_{\pi_k} = \sum_{i \in \pi_k} \mathbb{I}\{T_i = 1\} \tag{3}$$

and

$$N^{(0)}_{\pi_k} = \sum_{i \in \pi_k} \mathbb{I}\{T_i = 0\}, \tag{4}$$

where the Iverson bracket $\mathbb{I}$ is equal to one if the logical proposition between the brackets is satisfied. In addition, the number of treated and untreated test cases who do not drop out within the π_k segment is obtained as follows:

$$R^{(1)}_{\pi_k} = \sum_{i \in \pi_k} \mathbb{I}\{Y_i = 1\}\mathbb{I}\{T_i = 1\} \tag{5}$$

and

$$R^{(0)}_{\pi_k} = \sum_{i \in \pi_k} \mathbb{I}\{Y_i = 1\}\mathbb{I}\{T_i = 0\}. \tag{6}$$

Last, the segment-wise uplift [63] is calculated as follows:

$$\tau_{\pi_k} = \left(\frac{R^{(1)}_{\pi_k}}{N^{(1)}_{\pi_k}} - \frac{R^{(0)}_{\pi_k}}{N^{(0)}_{\pi_k}}\right)(N^{(1)}_{\pi_k} + N^{(0)}_{\pi_k}). \tag{7}$$

The performance of an uplift model can be visualized by an uplift curve [50] (see Fig. 1). The uplift curve shows the cumulated segment-wise uplift as a function of the fraction of targeted students. It illustrates the trade-off between the action of targeting larger proportions of the population and the resulting uplift. The overall effect of the treatment is the uplift resulting from targeting 100% of the test set. A straight line connecting the two extremes of the uplift curve (dash line in Fig. 1) serves as a baseline and represents the uplift that is achieved when students are randomly exposed to the treatment (i.e., random selection instead of selection using an uplift model). The farther is the uplift curve above the diagonal line, the better is the model. Moreover, it is expected that the curve has a steep increase until all responders are identified, to later flatten and potentially move downward as treating more students becomes ineffective. Decision-makers can use the uplift curve to decide the optimal proportion of students to target given the straightforward interpretation.

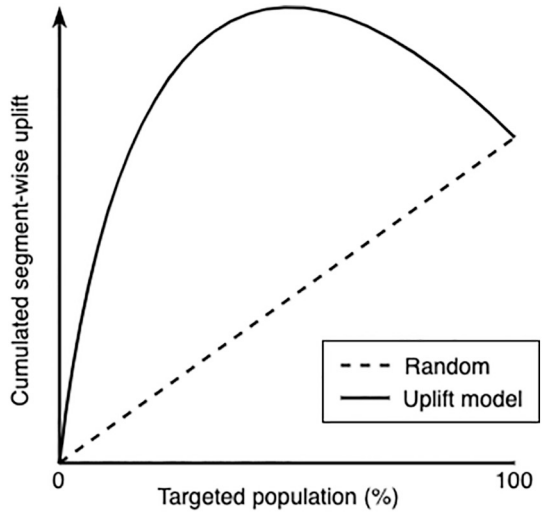


Fig. 1. Performance of an uplift model visualized by an uplift curve. The diagonal dashed line represents the random assignment of the treatment, whereas the curve bent upwards is the uplift curve. The farther the uplift curve from the diagonal line, the better the model.

The Qini measure is a quantitative performance metric that facilitates the comparison of the performance of different uplift models. Similarly to the area under the receiver operating characteristic curve (AUC or AUROC), as computed for evaluating binary classification models, the Qini is the area between the uplift curve and the diagonal. Thus, the larger the Qini, the better the performance of the model.

3.2. Important considerations for student dropout

Uplift modeling considers different segments of individuals depending on their response to an intervention [64]. Treatment responders are labeled as *persuadables*, whereas individuals who are harmed by the treatment are known as *do-not-disturbs*. Moreover, there are also individuals who either will never be persuaded -the *lost causes*- or will respond no matter the action -the *sure things*. Therefore, our interest is in identifying *persuadables* and refraining the treatment to the other categories.

There are important differences between the student dropout task and conventional uplift applications. First, there is a low risk of triggering students to drop out by targeting them with a retention effort, i.e., do-not-disturb students. This is because students with already outstanding performance who take tailored programs will still benefit from the intervention, and neither their performance nor their engagement will be negatively affected.

Furthermore, conventional uplift modeling avoids to treat *lost causes* which could pose an ethical problem. The reason for this is that underperforming students can be considered as “lost causes”, and therefore not be treated. Our objective, however, is not to refrain the action to students who might benefit from it, especially the ones with low academic performance. We seek to design customized programs for a specific segment of students using an uplift model. This does not imply cancelling the current program that is available to all the students.

Another issue that arises in this particular case is self-selection bias. The university makes an open call for all students who are willing to participate in the tutorials. The students, however, are autonomous to decide their participation. Nonetheless, students with grades below average on the standardized test for college admission and/or relatively poor performance in the first semester of the bachelor program are more likely to accept the invitation to the tutorials. This may affect the estimation of causal effects, and provide an erroneous estimation of the uplift. Therefore, we verify the presence of selection bias before training, and if needed, correct the imbalance between the pretreatment characteristics of the treatment groups.

4. Experimental analysis

First, this section describes the data set and the data preprocessing and transformation methods. Later, it presents the preliminary analysis consisting of the assessment and correction of selection bias. Last, it displays and discusses the Qini values of the twelve different uplift models and the uplift curve of the model with the best performance.

4.1. Data set description

We gathered a data set of 3362 students who enrolled between 2012 and 2016 in three bachelor programs of a business school. The complete list of variables is depicted in Table A1 in the appendix. Two main sources of variables were combined to perform this study: prematriculation information and academic performance. These variables were collected during the first year of the programs and their two sources are described next.

- Prematriculation information
  - *Sociodemographic* data consists of information about gender, age, marital status, occupational status, working hours, expected type of funding in higher school, family income level and residence.
  - *Family background* refers to data about the number of family members, educational level of both father and mother, number of parents that are alive, occupational status of both parents, number of members working, number of members enrolled in educational institutions, and an indicator of head of the family (e.g. father, mother, uncle, grandparent, among others).
  - *Standardized admission test* data indicate the scores of the Chilean standardized test (known as PSU) used for university admissions. It includes the subjects of mathematics, verbal, science, and history.
  - *High school features* include the type of high school (i.e., single-gender or mixed-gender education) and the type of funding received by the institution (i.e., public, private, or state-subsidized private).
- Academic information was collected along the first year of the bachelor programs. It consisted of data related to: dropout (i.e., temporary leave or absence, voluntary dropout, and expulsion), academic performance (i.e., final grades and credits for each course), entry type, declared bachelor program preferences, and the participation in the tutorials offered by the program for academic support (PAA).

The PAA seeks to improve the academic performance of students and to reduce the risk of student dropout by offering tutorials in subjects such as economics, mathematics, statistics, and English. The PAA started in 2012, and since then, students are invited each year at the beginning of the second semester to participate. Offering tutorials is a retention strategy whose objective is to prevent student dropout. We introduce a binary variable in the data set to indicate whether a student participated in any of the tutorials. Participants were labeled as treated, whereas nonparticipants as untreated.

The data set comprises 60 variables and includes academic performance information up to the end of the first semester. The outcome variable (i.e., dropout indicator) is defined on the basis of whether a student voluntarily abandoned the bachelor program within the one-year period after finishing the first semester. This time frame is chosen based on the starting point of the tutorials (i.e., at the beginning of the second semester) and the literature on student dropout.

4.2. Data preprocessing and transformation

First, we remove 72 observations without records on academic performance during the first semester (e.g., students who did not take

any course), and two variables with several missing values. Second, we impute missing values with the average or the mode, depending on the variable type. That is, missing values in numerical variables are replaced by the conditional average based on the entry cohort, whereas for categorical variables an additional category, "unknown", is created (those with more than 10% of missing observations) or substituted with the mode. An overview is provided in Table A2 in the appendix. Last, we apply dummy encoding to nominal variables and aggregate similar categories. The final data set includes 60 pretreatment variables.

4.3. Results

This subsection presents the preliminary analysis and the results of the uplift models trained on the student dropout data set. We assess and correct the balance across the pretreatment covariates of the treatment and nontreatment groups to mitigate the effects of selection bias. Then, different uplift models from the data processing and the data preprocessing approaches are employed to estimate the uplift, and later their performance is evaluated.

4.3.1. Preliminary analysis

Predicting the effect of treatments on an outcome of interest cannot be accurately determined when treatments are not randomly assigned to individuals, since the treatment groups may not be comparable. We assess the balance across the pretreatment variables of the treatment and the nontreatment groups to mitigate the effects of selection bias, since the tutorials were not assigned to students at random. The balance assessment can be performed either by considering theoretical evidence or by applying statistical tests. Statistical approaches consists of the estimation of: 1) the normalized difference between the treatment and nontreatment groups for each of the pretreatment variables, and 2) a chi-square test that performs an omnibus test whose null hypothesis states that at least one variable is significantly different between both groups.

Formally, the normalized difference Δ_x is defined as the difference in averages by treatment status, scaled by the square root of the sum of the variances divided by two (Eq. (8)). The $\overline{X}_t$ and s_t are the sample mean and the sample standard deviation of the x pretreatment variable for the treatment group, respectively. Analogously, $\overline{X}_c$ and s_c represent the mean and standard deviation for the untreated individuals. A rule of thumb in the literature is that a normalized difference larger than one quarter is an indication of imbalance in that particular pretreatment variable.

$$\Delta_x = \frac{\overline{X}_t - \overline{X}_c}{\sqrt{(s_t^2 + s_c^2)/2}}, \tag{8}$$

Table 1 displays ten of the pretreatment variables with the largest normalized differences. We observe considerable differences between the students who participated in the tutorials and those who did not in terms of their family income, PSU scores in mathematics, type of healthcare affiliation, parents educational level, among others. This disparity may indicate that the beneficiaries of the program were mainly students from low-income families. Therefore, an unbiased uplift estimation requires that we reduce the imbalance between the treated and untreated students by making as homogeneous as possible both groups.

In addition, we performed the omnibus test proposed by [65]. This implementation is available in the RItools package in R. The p-value of 5.55e-156 indicates that at least one pretreatment variable has imbalanced. This result is consistent with the normalized differences analysis.

This study employs propensity score matching (PSM) to reduce the effect of imbalanced pretreatment variables on the uplift estimation [38]. Among methods, such as multivariate regression, synthetic control, and instrumental variable estimators, PSM is one of the most widely used approaches for causal inference in observational studies since it: separates the adjustment of confounding and the estimation of treatment effects phases, excludes from the analysis individuals for whom no comparison can be made, allows to formally verify whether the resulting data set is balanced [66], and is not sensitive to the number of pretreatment variables [67]. Nonetheless, PSM requires large samples to achieve overlap between the treated and untreated individuals, as well as it only controls for observed confounding variables [68].

PSM seeks to balance the overall distribution of the pretreatment variables by pairing "similar" treated and untreated students. The measure of similarity is the propensity score (PS). The PS is the likelihood of an individual to be treated as a function of the pretreatment characteristics, $P(T = 1|X_i)$ [38]. Thus, treated and untreated individuals whose PS is relatively equal are considered similar and matched. This is done by means of nearest neighbour matching which selects the untreated individual whose PS is closest to the PS of a treated individual.

Table 2 illustrates the improvement in balance for the previously shown ten pretreatment variables after using PSM. The imbalance of some variables such as gross family income, PSU score in mathematics and private school remains large, but it is considerably reduced by PSM. The p-value of the omnibus test (0.987) demonstrates that the matched set is statistically balanced.

4.3.2. Uplift modeling techniques

This study includes a selection of uplift models from the data preprocessing and the data processing approaches. Two well-established learners were chosen for classification: random forest and boosted trees (i.e., *xgboost*), as it has been empirically observed that ensemble methods reduce considerably the risk of overfitting without

Table 1
Top ten unbalanced pretreatment variables.

Variable	Untreated		Treated		
	($N^{(0)}$ = 2,676)		($N^{(1)}$ = 614)		
	Mean	(s.d.)	Mean	(s.d.)	Δ_x
Gross family income	0.17	(0.95)	−0.75	(0.85)	1.02
PSU score mathematics	0.17	(0.92)	−0.76	(0.98)	0.98
Private school	0.61	(0.49)	0.19	(0.40)	0.94
Private healthcare	0.72	(0.45)	0.34	(0.47)	0.80
Father educ. level	0.14	(0.94)	−0.61	(1.03)	0.77
Mother educ. level	0.13	(0.94)	−0.59	(1.04)	0.73
Public healthcare	0.25	(0.43)	0.58	(0.49)	0.72
Public school	0.15	(0.36)	0.45	(0.50)	0.68
School type (governance)	0.01	(0.08)	−0.05	(0.12)	0.61
Human sciences as school's field	0.98	(0.12)	0.83	(0.38)	0.56

Table 2
Top ten unbalanced pretreatment variables after PSM.

Variable	Untreated		Treated		
	($N^{(0)}$ = 490)		($N^{(1)}$ = 490)		
	Mean	(s.d.)	Mean	(s.d.)	Δ_x
Gross family income	−0.45	(0.96)	−0.63	(0.88)	0.20
PSU score mathematics	−0.33	(0.97)	−0.52	(0.86)	0.20
Private school	0.34	(0.47)	0.24	(0.43)	0.21
Private healthcare	0.48	(0.50)	0.41	(0.49)	0.15
Father educ. level	−0.28	(1.08)	−0.48	(0.99)	0.19
Mother educ. level	−0.27	(0.99)	−0.50	(1.06)	0.22
Public healthcare	0.45	(0.50)	0.52	(0.50)	0.14
Public school	0.31	(0.46)	0.37	(0.48)	0.13
School type (governance)	−0.04	(0.11)	−0.05	(0.11)	0.15
Human sciences as school's field	0.94	(0.24)	0.93	(0.25)	0.02

compromising the bias error [69]. The baseline uplift methodology is the SMA. The MOA and MCA are implemented as in [32,43], respectively. Last, uplift is estimated directly by implementing the CTS [35], KL, ED, Chi [58], Xlearner and Rlearner [70] algorithms.

4.3.3. Uplift models performance

We use ten-fold crossvalidation to evaluate predictive performance, which is a well-established approach for model validation. The data set is split into ten folds of the same size. Stratification is applied to preserve within each fold the observed overall response rate. In every iteration, one fold is left out for testing and the remaining folds are used to train the model. The overall performance is the average of the results of each round and the standard deviation of the results indicates the stability of the model.

Section 3.1 defines uplift modeling as the estimation of the net effect of a treatment on an outcome of interest. The predicted individual uplift score can be seen as an indicator of how sensitive an individual is to the treatment. Hence, the treatment assignment consists of targeting individuals whose uplift score is positive. A good performing uplift model identifies accurately treatment responders and prioritizes the treatment allocation to those whose net treatment effect is the largest. This study empirically evaluates the performance of uplift modeling by ranking in descending order test cases based on their predicted uplift scores, later segmenting the sorted test set in four bins of equal size, and calculating the "observed bin uplift" as indicated in Eq. (7). The larger the observed bin uplift, the larger the net effect of the treatment within that particular bin. Hence, best performing uplift models would have a larger observed uplift in top bins than in subsequent bins.

Response modeling and uplift modeling aim to optimally identify individuals to maximize the effect of a targeting decision by employing predictive analytics. The difference lies in that response models build a predictive model only using the treatment group, while uplift models incorporate the information available in both the treatment and the non-treatment groups into the analysis. That is, uplift modeling targets students based on the predicted net effect of the tutorials, whereas targeting in response modeling is made on the basis of their likelihood to attrite.

We corroborate the advantage of uplift models over response modeling in treatment customization by contrasting the predicted uplift achieved when the tutorials are assigned as suggested by the MOAxgboost uplift model (Panel A) and by two conventional response models: the random forest (Panel B) and the XGboost (Panel C) algorithms (Fig. 2). We display the results of the MOAxgboost, as it will be seen below that this approach outperforms other uplift modeling methodologies trained on the student drop out data set. The performance of these models is evaluated according to the per bin observed uplift. In Fig. 2, the y axis shows the observed uplift, whereas the x axis indicates the bins. We can see that the uplift model prioritizes the assignment of tutorials to students who are expected to be positively affected. As more students are targeted, the effect of the program decreases and even become negative or null in the last bins. By contrast, response models fail in ranking correctly treatment responders, as the effect of the tutorials in the first quartiles is inferior or even negative than the one predicted by the MOAxgboost. The study by [8] aligns with this finding as it concludes that individuals whose risk of churning is high, are not in all cases the best targets for retention campaigns.

The Qini metric summarizes the performance of the different uplift modeling approaches employed in our experiments. The larger the Qini, the better the performance of the uplift model. Table 3 presents the Qini values for each technique at different targeting percentages, since uplift models may perform differently depending on the targeted fraction of students. The Qini values of the model with the best performance are in bold. Nonetheless, it is observed that the MOAxgboost achieves an optimal performance in targeting both small and large groups of students. That is, the MOAxgboost is the most appropriate approach to personalize the assignment of tutorials to students. Moreover, the variability of the Qini values for all models increases as more students

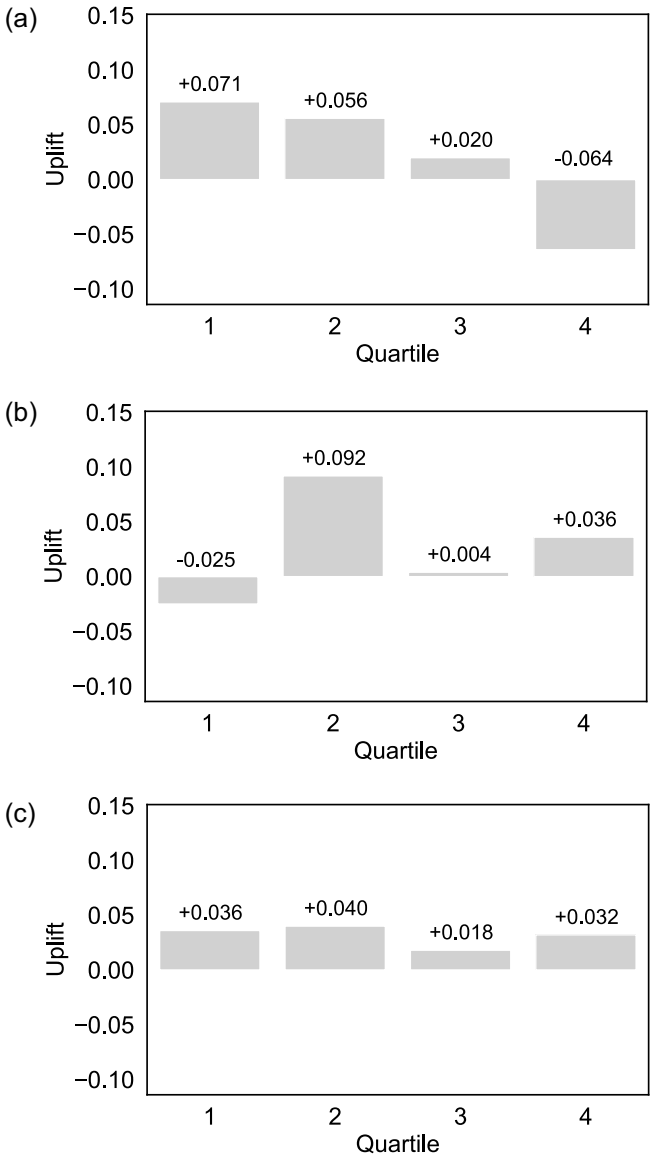


Fig. 2. The MOAxgboost uplift model (Panel A) contrasted against conventional random forest (Panel B) and extreme gradient boosting trees (XGboost) (Panel C) algorithms. It is observed that the uplift model effectively: 1) identifies treatment responders and 2) prioritizes the assignment of the treatment.

are selected to participate in the tutorials, indicating the instability of uplift modeling when targeting large samples.

As it was previously mentioned, the performance of uplift models is visualized by means of an uplift curve. The uplift curve illustrates the cumulated uplift as a function of the proportion of targeted individuals. Fig. 3 shows the uplift curve of the MOAxgboost. Overall, targeting according to the predictions of the MOAxgboost model boosts the effect of the program compared to targeting at random. The uplift curve increases irregularly up to targeting 80% of the sample, as the uplift model correctly identifies students who will not drop out due to their participation in the tutorials. Subsequently, it moves downward as the tutorials would be given to students whose intention to attrite will not change. The advantage of using an uplift curve is that it favors model comprehensibility and supports program designers in choosing the optimal fraction of students to target.

The results of the MOAxgboost model motivate the usage of uplift modeling as a support tool to assist decision-makers in designing retention strategies to reduce student dropout. The MOAxgboost uplift technique

Table 3
Performance evaluation: Qini values. The Qini values of the model with the best performance are in bold.

Model	at 10%	at 30%	at 50%	at 100%
SMARF	−0.0306 (0.0658)	−0.2063 (0.5318)	−0.4595 (1.2152)	−0.0948 (2.3726)
SMAxgboost	−0.0383 (0.1511)	−0.1708 (0.6702)	−0.2082 (1.1038)	−0.5273 (2.7071)
MOARF	−0.0587 (0.0518)	−0.2428 (0.2523)	−0.3337 (0.6460)	0.1160 (1.4222)
MOAxgboost	**0.0385** (0.1123)	**0.2253** (0.5894)	**0.5095** (1.0739)	**1.2031** (2.0885)
MCARF	−0.1308 (0.1841)	−0.4825 (0.8737)	−0.6002 (1.4692)	−0.7283 (2.6170)
MCAxgboost	−0.0475 (0.1020)	−0.1592 (0.3887)	−0.2714 (0.8371)	−0.5329 (1.5594)
XLearner	−0.0142 (0.0037)	−0.0474 (0.1636)	−0.0627 (0.4795)	0.0423 (1.6597)
RLearner	−0.0262 (0.0746)	−0.1062 (0.5967)	−0.2776 (1.1711)	−0.3170 (2.1355)
KL	−0.0618 (0.1083)	−0.4044 (0.3630)	−0.6504 (0.7124)	−0.9005 (1.3740)
ED	−0.0817 (0.0561)	−0.4356 (0.2727)	−0.6618 (0.7010)	−0.7375 (1.6394)
Chi	−0.0912 (0.0888)	−0.6302 (0.3289)	−1.1111 (0.5995)	−1.2944 (1.4747)
CTS	−0.0587 (0.1162)	−0.2853 (0.5716)	−0.5726 (0.8403)	−0.7693 (1.7291)

performs well in the majority of analyzed segments and, therefore, personalizing the assignment of tutorials according to its predictions can prevent future cases of student dropout. We, however, advice training different uplift modeling approaches, as there is no single uplift modeling technique with outstanding performance for all problems (i.e., No free lunch theorem) [40].

Last, we are also interested in analyzing in terms of the pretreatment variables to what extend students identified by the MOAxgboost uplift model as treatment responders differ from those who are treatment nonresponders. Our interest lies in the accuracy of the model, but also in its comprehensibility. Therefore, we form four different profiles on the basis of the quartiles shown in Fig. 2, Panel A. Since test set students are ranked according to their likelihood to respond positively to the program, i.e., individual uplift scores of the MOAxgboost model, segments of students with high treatment effects are those at the top.

The student profiles result from averaging the values of the ten most important predictors of the MOAxgboost. The importance scores of each attribute are obtained by fitting over ten folds a MOAxgboost, and subsequently averaging the values. We use the *Gain* criterion, as it indicates how valuable is the attribute at the splits during the construction of the trees. Although more variables can be selected, inspecting few variables allows us to maintain clarity in visualizations.

Fig. 4 illustrates the variation of the average across the four quartiles for the chosen variables. Students with the highest likelihood to respond to the retention program are those in the top segments. We can conclude from the figure that the main differences among treatment responders and nonresponders are observed in their PSU scores in mathematics, the number of members in the family, attendance to a private school, the overall performance in the first semester, and the performance in English courses.

The studies by [15,17,19] indicate the association of academic performance variables from the first semester of bachelor programs and student dropout. Although the model suggests targeting students with relatively good performance, their proficiency in mathematics is among the lowest. Particularly, math test scores are related to dropout during the first academic year [71]. In addition, treatment responders are part of households with relatively few family members, as well as graduated from nonprivate high schools. This indicates that retention campaigns can take a proactive rather than a reactive approach, since prematriculation information may also be used for treatment customization.

The radar charts in this study intend to facilitate the interpretation of uplift modeling estimates for decision-makers. They are valuable to understand the needs of students and emphasize the attributes that differentiate treatment responders from nonresponders.

5. Conclusions

This article applies the uplift modeling framework to the problem of student dropout prevention. We demonstrate that focusing retention efforts, i.e., offering tutorials, on students with the largest likelihood to be retained due to the intervention boosts the effect of the program. We are able to reach higher uplift with the best machine learning model designed for this purpose, than with the alternative of targeting at random.

Self-selection bias is tested and corrected as part of the modeling process to avoid bias in the uplift estimation. Subsequently, we train twelve different uplift modeling approaches to predict students'

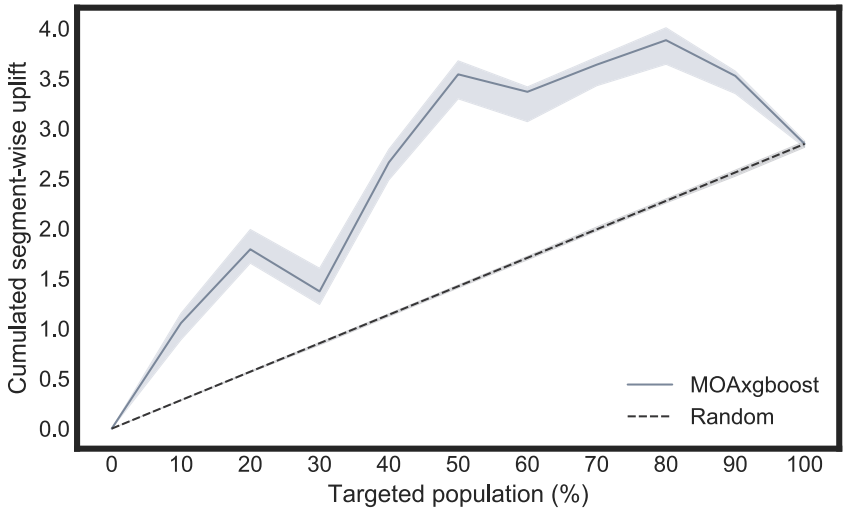


Fig. 3. Uplift curve MOAxgboost. The MOAxgboost model boosts the effect of the program compared to targeting at random, as students who will not drop out due to the intervention are prioritized. The uplift curve bends downward, as the tutorials are assigned to students whose intention to attrite will not change.

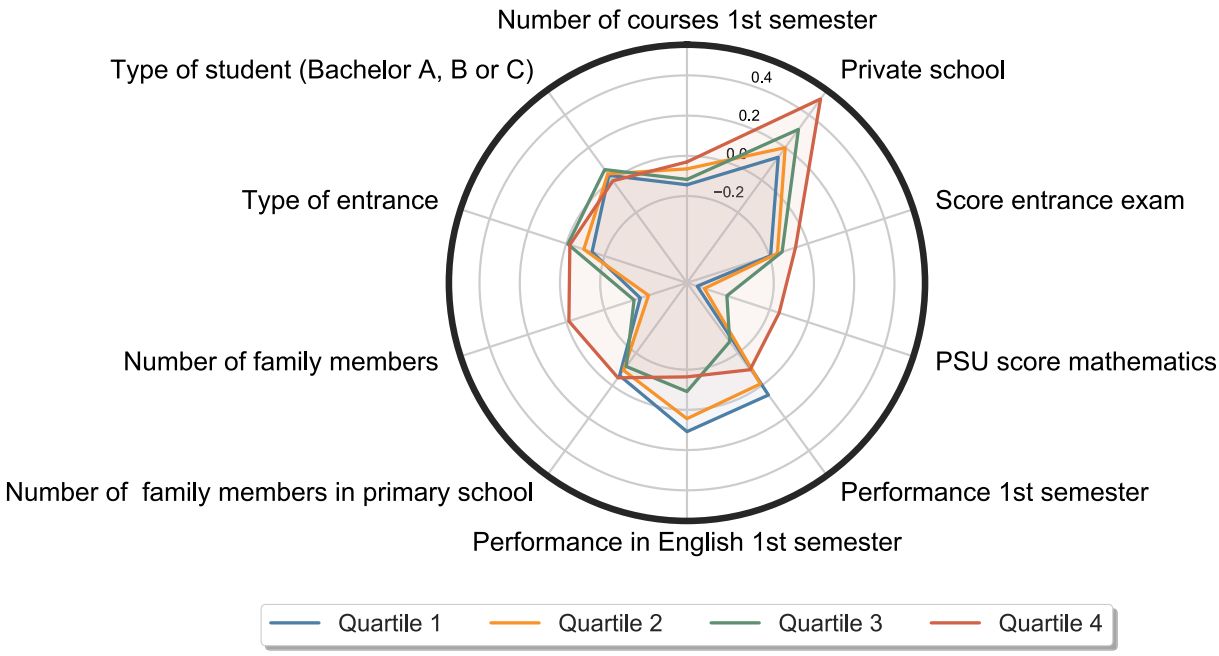


Fig. 4. The radar chart profiling allows us to examine the characteristics of students with high and low expected treatment effects. Students who are expected to benefit from the tutorials are those who present the characteristics of the upper quartiles.

response to the retention strategy, and assess feature relevance to better understand the characteristics of students who are likely to be retained with such program. This knowledge translates into a better design of tailored retention efforts. Particularly, the importance of pre-matriculation attributes indicates that the design of retention efforts can take a proactive rather than a reactive approach.

There are several opportunities for future research. First, a further step is to target students according to the uplift model, and subsequently corroborate the effectiveness of the customized targeting assignment. This task, however, requires setting aside a holdout set of students who should not be treated, as estimating causal effects requires the comparison of alike individuals. Particularly, in terms of their characteristics and likelihood to respond to the intervention. Second, student dropout is a context-specific phenomenon and retention strategies comprise, but are not limited to, offering tutorials. Therefore, applying the uplift modeling framework to different institutional contexts, i.e., data collection on prematriculation and academic information at other universities, would enrich the understanding on the effectiveness and limitations of this approach in the customization of retention programs. Third, incorporating academic information from subsequent semesters may enhance model estimates and the comprehension of long-term program effects. Last, profit metrics for business analytics can be adapted to assess the benefits and costs of student dropout, as retaining students leads to social benefits and positive externalities.

Acknowledgements

This work was supported in part by Innoviris, the Brussels Region Research Funding Agency, Brussels, Belgium. The third author gratefully acknowledges financial support from CONICYT PIA/BASAL AFB180003 and FONDECYT, grant 11607384.

Appendix A

Table A1
Data set variables.

Category	Subcategory	Variables
Academic	Academic	Number of courses 1st semester, year of entry, preferred application to bachelor program, performance 1st semester, performance in statistics 1st semester, performance in mathematics 1st semester, performance in economics 1st semester, performance in English 1st semester, type of entrance, and type of student (Bachelor A, B or C).
Prematriculation	Family background	Mother as head of family, head of family unknown, father as head of family, father educ. Level, number of parents alive, number of hours dedicated to work, mother educ. Level, number of family members in other schools, number of family members in college, number of family members in secondary school, number of working family members, number of family members in primary school, number of family members in preschool, number of family members, number of family members studying, occupation father, and occupation mother.
	High school attributes	Private school, public school, male school, female school, human sciences as school's field, graduation year from school, state-subsidized private school, and school type (governance).
	Sociodemographic	Family support, central region, southern Region, parent support, independent student, private healthcare, public healthcare, financing source unknown, study support unknown, loan as 2nd funding source, scholarship as 2nd funding source, unknown 1st funding source, parents support as 1st funding source, scholarship as 1st funding source, parents support as 2nd funding source, gender female, single, gross family income, work schedule, and working student.
	Standardized admission test	Score entrance exam, PSU score lang. & comu., PSU score mathematics, PSU score hist. & scien., ranking score, high school grades, and age during 1st semester.

Table A2
Missing values before variable transformation.

Variable	Missing values (%)	Replacement strategy
Age during 1st semester	0.27	mean
Father educ. level	16.14	"Unknown"
First funding source	13.01	"Unknown"
Gender	0	–
Graduation year from school	2.01	mean
Gross family income	2.13	mode
Head of family	12.67	"Unknown"
Healthcare type	2.13	"Unknown"
High school grades	8.57	mean
Male/Female/Mix school	2.80	mode
Marital Status	0	–
Max. score science-history	13.28	mean
Mother educ. level	13.62	"Unknown"
Number of family members in preschool	2.13	mean
Number of family members in primary school	2.13	mean
Number of courses 1st semester	0	–
Number of family members	11.88	mean
Number of family members in college	2.13	mean
Number of family members in other schools	2.13	mean
Number of family members in secondary school	2.13	mean
Number of family members studying	2.13	mean
Number of hours dedicated to work	2.13	mean
Number of working family members	2.13	mean
Occupation father	18.18	"Unknown"
Occupation mother	14.19	"Unknown"
Parents alive	11.82	"Unknown"
Performance 1st semester	0.03	mean
Performance in economics 1st semester	1.43	mean
Performance in English 1st semester	14.71	mean
Performance in mathematics 1st semester	2.34	mean
Performance in statistics 1st semester	90	removed
Preferred application to bachelor program	10.43	mean
PSU score lang. & comu.	13.28	mean
PSU score mathematics	13.28	mean
Ranking score	24.86	mean
Region	2.37	mode
School field	2.71	mode
School type (funding)	2.71	mode
School type (governance)	2.71	mode
Score entrance exam	8.45	mean
Second funding source	22.22	"Unknown"
Support	14.77	"Unknown"
Type of entrance	0	–
Type of student (Bachelor A, B or C)	0	–
Work schedule	98.69	removed
Working student	10.09	mode
Year of entry	0	–

References

[1] L. Thomas, Student retention in higher education: the role of institutional habitus, J. Educ. Pol. 17 (4) (2002) 423–442.
[2] T. Dharmawan, H. Ginardi, A. Munif, Dropout detection using non-academic data, 2018 4th International Conference on Science and Technology (ICST), IEEE, 2018, pp. 1–4.
[3] C.J. Bland, A.L. Taylor, S.L. Shollen, A.M. Weber-Main, P.A. Mulcahy, Faculty Success through Mentoring: A Guide for Mentors, Mentees, and Leaders, R&L Education, 2009.
[4] S. Larose, D. Cyrenne, O. Garceau, M. Harvey, F. Guay, F. Godin, G.M. Tarabulsy, C. Deschênes, Academic mentoring and dropout prevention for students in math, science and technology, Mentor. Tutor. 19 (4) (2011) 419–439.
[5] N. Zepke, L. Leach, Improving student engagement: ten proposals for action, Act. Learn. High. Educ. 11 (3) (2010) 167–177.
[6] L. Thomas, Building student engagement and belonging in higher education at a time of change, Paul Hamlyn Found. 100 (2012) 1–99.
[7] M. Yorke, The development and initial use of a survey of student belongingness, engagement and self-confidence in UK higher education, Assess. Eval. High. Educ. 41 (1) (2016) 154–166.
[8] E. Ascarza, Retention futility: targeting high-risk customers might be ineffective, J. Mark. Res. 55 (1) (2018) 80–98.
[9] W.G. Spady, Dropouts from higher education: an interdisciplinary review and synthesis, Interchange 1 (1) (1970) 64–85.
[10] V. Tinto, Dropout from higher education: a theoretical synthesis of recent research, Rev. Educ. Res. 45 (1) (1975) 89–125.
[11] J.P. Bean, Conceptual models of student attrition: how theory can help the institutional researcher, New Dir. Inst. Res. 1982 (36) (1982) 17–33.
[12] J.P. Bean, Interaction effects based on class level in an explanatory model of college student dropout syndrome, Am. Educ. Res. J. 22 (1) (1985) 35–64.
[13] R. Chen, S.L. DesJardins, Investigating the impact of financial aid on student dropout risks: racial and ethnic differences, J. High. Educ. 81 (2) (2010) 179–208.
[14] J. Forsman, C. Linder, R. Moll, D. Fraser, S. Andersson, A new approach to modelling student retention through an application of complexity thinking, Stud. High. Educ. 39 (1) (2014) 68–86.
[15] A. Fortin, L. Sauvé, C. Viger, F. Landry, Nontraditional student withdrawal from undergraduate accounting programmes: a holistic perspective, Acc. Educ. 25 (5) (2016) 437–478.
[16] B.M. Kehm, M.R. Larsen, H.B. Sommersel, Student dropout from universities in europe: a review of empirical literature, Hungarian Educ. Res. J. 9 (2) (2019) 147–164.
[17] J. Vásquez, J. Miranda, Student desertion: What is and how can it be detected on time? Data Science and Digital Business, Springer, 2019, pp. 263–283.
[18] A.L. Caison, Analysis of institutionally specific retention research: a comparison between survey and institutional database methods, Res. High. Educ. 48 (4) (2007) 435–451.
[19] C.H. Yu, S. DiGangi, A. Jannasch-Pennell, C. Kaprolet, A data mining approach for identifying predictors of student retention from sophomore to junior year, J. Data Sci. 8 (2) (2010) 307–325.
[20] G. Johnes, R. McNabb, Never give up on the good times: student attrition in the UK, Oxf. Bull. Econ. Stat. 66 (1) (2004) 23–47.

[21] M. Ferreira, Gender issues related to graduate student attrition in two science departments, Int. J. Sci. Educ. 25 (8) (2003) 969–989.
[22] M. Saldaña, O. Barriga, An adaptation of tinto's attrition model to the universidad católica de la santísima concepción, chile, Rev. Ciencias Soc. 16 (4) (2016) 616–628.
[23] P.L. Peterson, E. Baker, B. McGaw, International Encyclopedia of Education, Elsevier Ltd., 2010.
[24] G. Kostopoulos, S. Kotsiantis, P. Pintelas, Estimating student dropout in distance higher education using semi-supervised techniques, Proceedings of the 19th Panhellenic Conference on Informatics, Athens, Greece, 2015, pp. 38–43.
[25] N. Iam-On, T. Boongoen, Generating descriptive model for student dropout: a review of clustering approach, Human-Centric Comput. Inf. Sci. 7 (1) (2017) 1.
[26] N. Iam-On, T. Boongoen, Improved student dropout prediction in thai university using ensemble of mixed-type data clusterings, Int. J. Mach. Learn. Cybern. 8 (2) (2017) 497–510.
[27] K. Imai, M. Ratkovic, et al., Estimating treatment effect heterogeneity in randomized program evaluation, Ann. Appl. Stat. 7 (1) (2013) 443–470.
[28] M. Qian, S.A. Murphy, Performance guarantees for individualized treatment rules, Ann. Stat. 39 (2) (2011) 1180.
[29] K. Kane, V.S. Lo, J. Zheng, Mining for the truly responsive customers and prospects using true-lift modeling: comparison of new and existing methods, J. Market. Anal. 2 (4) (2014) 218–238.
[30] L. Guelman, M. Guillén, A.M. Pérez-Marín, Random forests for uplift modeling: an insurance customer retention case, International Conference on Modeling and Simulation in Engineering, Economics and Management, Springer, 2012, pp. 123–133.
[31] B. Hansotia, B. Rukstales, Incremental value modeling, J. Interact. Mark. 16 (3) (2002) 35.
[32] V.S. Lo, The true lift model: a novel data mining approach to response modeling in database marketing, ACM SIGKDD Explor. Newslett. 4 (2) (2002) 78–86.
[33] S. Jaroszewicz, P. Rzepakowski, Uplift modeling with survival data, ACM SIGKDD Workshop on Health Informatics (HI-KDD14), New York City, 2014.
[34] Y. Zhao, X. Fang, D. Simchi-Levi, A practically competitive and provably consistent algorithm for uplift modeling, 2017 IEEE International Conference on Data Mining (ICDM), IEEE, 2017, pp. 1171–1176.
[35] Y. Zhao, X. Fang, D. Simchi-Levi, Uplift modeling with multiple treatments and general response types, Proceedings of the 2017 SIAM International Conference on Data Mining, SIAM, 2017, pp. 588–596.
[36] D.B. Rubin, Estimating causal effects of treatments in randomized and nonrandomized studies, J. Educ. Psychol. 66 (5) (1974) 688.
[37] J. Pearl, Causality, Cambridge University Press, 2009.
[38] P.R. Rosenbaum, D.B. Rubin, The central role of the propensity score in observational studies for causal effects, Biometrika 70 (1) (1983) 41–55.
[39] D.B. Rubin, Bayesian inference for causal effects: the role of randomization, Ann. Stat. (1978) 34–58.
[40] F. Devriendt, D. Moldovan, W. Verbeke, A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: a stepping stone toward the development of prescriptive analytics, Big Data 6 (1) (2018) 13–41.
[41] A. Abadie, Semiparametric difference-in-differences estimators, Rev. Econ. Stud. 72 (1) (2005) 1–19.
[42] B. Zhang, A.A. Tsiatis, E.B. Laber, M. Davidian, A robust method for estimating optimal treatment regimes, Biometrics 68 (4) (2012) 1010–1018.
[43] Y.-T. Lai, K. Wang, D. Ling, H. Shi, J. Zhang, Direct marketing when there are voluntary buyers, Sixth International Conference on Data Mining (ICDM'06), IEEE, 2006, pp. 922–927.
[44] R.M. Gubela, S. Lessmann, S. Jaroszewicz, Response transformation and profit decomposition for revenue uplift modeling, Eur. J. Oper. Res. 283 (2) (2020) 647–661.
[45] K. Rudaś, S. Jaroszewicz, Linear regression for uplift modeling, Data Min. Knowl. Disc. 32 (5) (2018) 1275–1305.
[46] S. Athey, G. Imbens, Recursive partitioning for heterogeneous causal effects, Proc. Natl. Acad. Sci. 113 (27) (2016) 7353–7360.
[47] S.R. Künzel, J.S. Sekhon, P.J. Bickel, B. Yu, Metalearners for estimating heterogeneous treatment effects using machine learning, Proc. Natl. Acad. Sci. 116 (10) (2019) 4156–4165.
[48] X. Nie, S. Wager, Quasi-oracle estimation of heterogeneous treatment effects, arXiv (2019) 1712.04912.
[49] L. Tian, A.A. Alizadeh, A.J. Gentles, R. Tibshirani, A simple method for estimating interactions between a treatment and a large number of covariates, J. Am. Stat. Assoc. 109 (508) (2014) 1517–1532.
[50] N.J. Radcliffe, P.D. Surry, Real-world uplift modelling with significance-based uplift trees, White Paper TR-2011-1, Stochastic Solutions (2011) 1–33 http://stochasticsolutions.com/pdf/sig-based-up-trees.pdf , Accessed date: 3 March 2019.
[51] J.L. Hill, Bayesian nonparametric modeling for causal inference, J. Comput. Graph. Stat. 20 (1) (2011) 217–240.
[52] S. Wager, S. Athey, Estimation and inference of heterogeneous treatment effects using random forests, J. Am. Stat. Assoc. 113 (523) (2018) 1228–1242.
[53] S. Powers, J. Qian, K. Jung, A. Schuler, N.H. Shah, T. Hastie, R. Tibshirani, Some methods for heterogeneous treatment effect estimation in high dimensions, Stat. Med. 37 (11) (2018) 1767–1787.
[54] S. Athey, J. Tibshirani, S. Wager, et al., Generalized random forests, Ann. Stat. 47 (2) (2019) 1148–1178.
[55] F. Alemi, H. Erdman, I. Griva, C.H. Evans, Improved statistical methods are needed to advance personalized medicine, Open Transl. Med. J. 1 (2009) 16.
[56] L. Guelman, Optimal Personalized Treatment Learning Models With Insurance Applications, Universitat de Barcelona, Barcelona, Spain, 2015.
[57] D.M. Chickering, D. Heckerman, A decision theoretic approach to targeted advertising, Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers Inc., 2000, pp. 82–88.
[58] P. Rzepakowski, S. Jaroszewicz, Decision trees for uplift modeling with single and multiple treatments, Knowl. Inf. Syst. 32 (2) (2012) 303–327.
[59] S. Debaere, F. Devriendt, J. Brunneder, W. Verbeke, T. De Ruyck, K. Coussement, Reducing inferior member community participation using uplift modeling: evidence from a field experiment, Decis. Support. Syst. 113077 (2019).
[60] L. Zaniewicz, S. Jaroszewicz, Support vector machines for uplift modeling, 2013 IEEE 13th International Conference on Data Mining Workshops, IEEE, 2013, pp. 131–138.
[61] C. Li, X. Yan, X. Deng, Y. Qi, W. Chu, L. Song, J. Qiao, J. He, J. Xiong, Reinforcement learning for uplift modeling, arXiv (2019) 1–22 1811.10158.
[62] P.W. Holland, Statistics and causal inference, J. Am. Stat. Assoc. 81 (396) (1986) 945–960.
[63] P. Gutierrez, J.-Y. Gérardy, Causal inference and uplift modelling: a review of the literature, International Conference on Predictive Applications and APIs, 2017, pp. 1–13.
[64] N.J. Radclifte, R. Simpson, Identifying who can be saved and who will be driven away by retention activity, Journal of Telecommunications Management 1 (2) (2008) 168–176.
[65] J. Bowers, M. Fredrickson, B. Hansen, RItools: Randomization inference tools, R package version 0.1-11, 2010.
[66] E.L. Fu, R.H. Groenwold, C. Zoccali, K.J. Jager, M. van Diepen, F.W. Dekker, Merits and caveats of propensity scores to adjust for confounding, Nephrol. Dial. Transplant. 34 (10) (2019) 1629–1635.
[67] R.B. DAgostino Jr., Propensity scores in cardiovascular research, Circulation 115 (17) (2007) 2340–2343.
[68] D.B. Rubin, Estimating causal effects from large data sets using propensity scores, Ann. Intern. Med. 127 (8_Part_2) (1997) 757–763.
[69] Z. Khan, A. Gul, A. Perperoglou, M. Miftahuddin, O. Mahmoud, W. Adler, B. Lausen, Ensemble of optimal trees, random forest and random projection ensemble classification, ADAC (2019) 1–20.
[70] Z. Zhao, T. Harinen, Uplift modeling for multiple treatments with cost optimization, arXiv (2019) 1–10 1908.05372.
[71] K. Kori, M. Pedaste, E. Tõnisson, T. Palts, H. Altin, R. Rantsus, R. Sell, K. Murtazin, T. Rüütmann, First-year dropout in ict studies, 2015 IEEE Global Engineering Education Conference (EDUCON), IEEE, 2015, pp. 437–445.



Diego Olaya is a Ph.D. candidate at the Vrije Universiteit Brussel. He received his bachelor's degree in Economics from the Universidad Nacional de Colombia, and he holds a M.Sc. in Management Science from the Vrije Universiteit Brussel (VUB). His research interests include business analytics, data mining, and the intersection of machine learning and causal inference.



Jonathan Vásquez received his B.S. and M.S. degree from the University of Chile in 2012 and 2016 respectively. He is currently academic at Ingenieria en Información y Control de Gestión, Escuela de Auditoría, Universidad de Valparaiso. His research focuses on the application of data mining and machine learning in student desertion, students' performance, educational data science, and profiles classification.



Sebastián Maldonado received his B.S. and M.S. degree from the University of Chile, in 2007, and his Ph.D. degree from the University of Chile, in 2011. He is currently Full Professor at the Department of Management Control and Information Systems, School of Economics and Business, University of Chile. His research interests include statistical learning, data mining and business analytics. Sebastián Maldonado has published more than 70 scientific contributions including more than 50 Thomson Reuters' ISI papers in the last ten years.



Jaime Miranda is currently Associate Professor and Head at the Department of Management Control and Information Systems, University of Chile. He has an Industrial Engineer B.S. degree, a M.S. degree in Operations Management and a PhD. degree in Engineering Systems from the University of Chile. In 2016 was president of the Association of Latin-Iberoamerican Operational Research Societies (ALIO). His research interests include Operation Research and Business Analytics. He has published several articles and implemented OR systems in Chile.



Wouter Verbeke, Ph.D., is associate professor of data analytics at Vrije Universiteit Brussel (Brussels, Belgium). He graduated in 2007 as a Civil Engineer and obtained a Ph.D. in applied economics at KU Leuven in 2012. His research is situated in the field of prescriptive and profit-driven data analytics and is driven by real-life business applications in fraud, customer relationship, credit risk, supply chain, and human resources management. In 2014, he won the distinguished EURO award for best article published in the European Journal of Operational Research in the category 'Innovative Applications of O.R'. His work has been published in established international scientific journals such as IEEE Transactions on Knowledge and Data Engineering, Information Sciences and European Journal of Operational Research. He has authored two books, entitled 'Fraud Analytics Using Descriptive, Predictive & Social Network Techniques' and 'Profit-driven Business Analytics', published by Wiley.