

The Effect of Age of Enrollment on the Probability of Graduating from Academic Programs

Gur Keinan 213635899

Yarden Adi 212585848

November 7, 2024

Abstract

Many adolescents worldwide have wondered when they should start their journey in higher education. The writers of this project themselves had decided long ago to join the "Atuda" program, which means starting university at the relatively young age of eighteen. Thus, it is natural to wonder - does the student's age at the start of learning at the university affect the success of the student? An answer to this question can drastically change the landscape of the campuses worldwide and the grades of those studying there. In this Causal Inference project, we aim to explore precisely that. Specifically, we investigate the following causal question: *What is the causal effect of enrolling as an adult student (age 21 or older) on the probability of graduating from an academic program within the allotted time?* Throughout this project, we present and explore the data we use to address this question and determine its suitability for causal analysis tasks, formally present the methods we use to answer the research question, present and discuss the analysis results, and conclude the project. One can find all of the resources used in the project in this GitHub repository.

1 Data Review and Preprocessing

To investigate the relationship between age at enrollment and academic success, we needed data that captures students' academic progress and personal characteristics at the time of enrollment. This section details the dataset we used and the preprocessing steps taken to prepare it for causal analysis.

1.1 Data Review

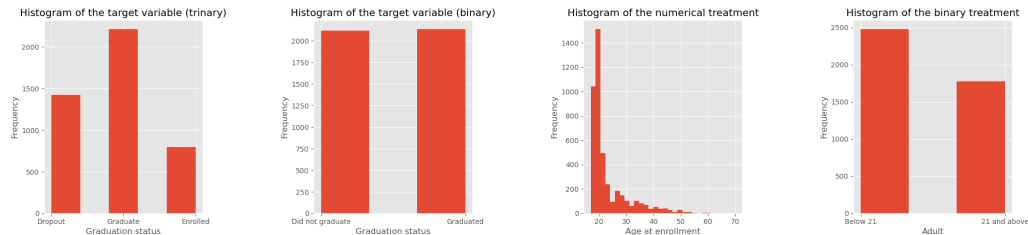
The data used in our project originated from research done in Portugal by the Polytechnic Institute of Portalegre (higher education institute that focuses on technology), as an attempt to provide information to the tutoring team about the risk of students' dropout and failure (Realinho et al., 2022). Commonly, it is used to build machine learning models for predicting academic performance and dropout (see relevant Kaggle competitions in this subject).

The data contains information about students' pre-academic background, age, academic performance, social and economic status, and other relevant variables. It consists of 4424 records and 37 variables, including the enrollment age. The dataset includes a trinary outcome variable indicating whether a student dropped out, graduated, or is still enrolled in the academic program after its predetermined period. A comprehensive description of all features, their types, and possible values is provided in Appendix A.

The data was created by joining three primary data sources:

1. CNAES (National Competition for Access to Higher Education) - contains information about students' academic backgrounds, demographics, and course applications at the time of their enrollment in Portuguese higher education institutions.
2. AMS (Academic Management System) - provides student records data, including demographic information, course enrollments, and academic performance throughout their studies.
3. PORTDATA (Contemporary Portugal Database) - provides macroeconomic data, including unemployment, inflation, and GDP figures.

Figure 1: Distributions of the target (graduation within the allotted time) and treatment (adult enrollment) variables.



Traditionally, each attribute used in the dataset is associated with one of the following classes: demographic, socioeconomic, macroeconomic, academic data at enrollment, and academic data at the end of the first and second semesters. For a complete list of the variables in the dataset and their corresponding classes, please refer to Table 1.

1.2 Data Preprocessing

Treatment and Target variables Following our causal question, we introduce a new binary variable - "enrolling as an adult student (over 21)". This variable acts as the *treatment variable* in our analysis. We omit the numerical age variable for the rest of the analysis. Additionally, as our question focuses on the probability of graduating from an academic program within the allotted time, we introduce another binary variable - "Graduated from an academic program within the expected time frame" rather than the commonly used trinary variable - "graduate/ enrolled/ dropout". This variable is the *target variable* in our analysis. Importantly, we consider students still enrolled beyond the expected completion time as not having graduated within the allotted time, aligning with the standard duration of academic programs in Portuguese higher education. The target and treatment variables distributions are visually presented in Figure 1.

Removal of post-treatment variables The dataset contains multiple variables gathered throughout the student's academic journey, e.g., academic accomplishments, debts, and payment tracking. Despite being informative for machine learning models trying to predict academic dropout, these accomplishments were recorded after the treatment was determined and, therefore, cannot be safely used in the causal inference procedure. We eliminated those variables to maintain the integrity of our results and prevent any post-treatment interference.

Removal of treatment-correlated features During the one-hot encoding of categorical variables, we deliberately excluded the dummy variable representing "Application mode 39 - Over 23 years old" since this feature is strongly correlated with our treatment variable (being 21 or older). Including this dummy variable could have introduced redundancy and potentially biased our analysis.

Clustering categorical values The dataset is of impressive complexity and detail, as evident by the categorical variables with over 30 unique values. To perform meaningful analysis, we manually cluster similar categories into one broader category to represent them instead of simply removing the less frequent value. For example, we merged the values "Armed Forces Professions", "Armed Forces Officers", and "Armed Forces Sergeants" into a single "Armed Forces" category, which functions as a single value in the analysis. We performed this procedure on five variables within the data set - the qualifications and occupations of the parents (mother and father) and the student's previous qualifications.

Pruning categorical outliers Even after performing the previous step, some variables still contained rare values. After careful consideration and visual inspections, we decided to prune some of the rare values to ease the analysis. The pruning resulted in a sample exclusively consisting of individuals with Portuguese nationality, which may limit the generalizability of this study to a broader population.

After the preprocessing stage, the dataset comprises 4249 records, each containing 21 variables, including the treatment and target variables. Five of these variables are numerical, and the rest are categorical.

2 Assumptions for Causal Inference

In this section, we formally present and discuss four assumptions regarding the nature of our data. Combined, those assumptions guarantee the trustworthiness of an observational causal experiment’s results.

2.1 Stable Unit Treatment Value Assumption (SUTVA)

The SUTVA assumption consists of two parts. The first one is *no interference*, which requires that the potential outcomes of each unit are not affected by the treatment assignment of any other unit; the second is *no hidden variations of treatment*, which forbids the existence of different forms or versions of each treatment level, which lead to different potential outcomes.

In our experiment, the treatment variable, "enrolling as an adult student (over 21)", is well-defined and has only two versions. Hence, the second part of the assumption safely holds. The assumption’s first part holds if we presume one student’s adulthood does not affect the probability of his fellow student graduating from the program, which is controversial. There are several aspects in which the varied spectrum of ages on the campus might affect the students’ success. Such elements include but are not limited to 1) peer effects - encountering different perspectives and learning methods (that can be related to one’s age) might affect one’s abilities; and 2) In competitive programs, students’ success may be directly affected by their peers’ accomplishments. Therefore, if age is indeed related to one’s academic achievements, one student’s age can affect other students’ success.

2.2 Consistency

The Consistency assumption states that an individual’s potential outcome under their observed exposure history is the outcome that would actually be observed for that person. Put formally, for a unit that receives treatment T , we observe the corresponding potential outcome $Y = TY_1 + (1 - T)Y_0$. We believe this assumption holds in our study for several reasons:

1. **Data Source Reliability:** Our data comes from three well-established institutional databases (CNAES, AMS, and PORDATA), each with standardized data collection procedures and quality control measures.
2. **Clear Treatment Definition:** The treatment (being 21 or older at enrollment) is precisely defined and measured without ambiguity. Unlike many treatments that might vary in intensity or implementation, age at enrollment is an objective measure that cannot be misinterpreted.
3. **Outcome Measurement:** The graduation outcome is documented in the academic records system (AMS) and follows standardized institutional definitions of what constitutes graduation within the allotted time.
4. **Data Processing Transparency:** All our data preprocessing steps are well documented and reproducible, ensuring that the transformation from raw data to analysis variables maintains the integrity of the treatment and outcome measurements.
5. **Stable Treatment:** Age at enrollment is a stable characteristic that cannot change retrospectively, ensuring that the treatment status remains consistent throughout the study period.

2.3 Ignorability - No Unmeasured Confounders

The Ignorability assumption, also known as the assumption of no unmeasured confounders, states that the treatment assignment T is independent of the potential outcomes Y_0, Y_1 given the observed covariates X , that is, $Y_0, Y_1 \perp T \mid X$. This assumption is essential for ensuring that the estimated treatment effect is unbiased and not confounded by factors affecting both the treatment and the outcome.

However, this assumption is inherently unverifiable in practice since it is impossible to confirm whether all relevant confounders have been measured and included in the model. As noted by Hernán and Robins (2006), we cannot account for unmeasured confounders, as we do not observe them. Despite this limitation, research has shown that correlations between observed covariates and unmeasured confounders can reduce the bias associated with missing information (Schulz et al., 2023). This means that if our dataset contains variables from diverse domains that possibly affect the treatment or potential outcomes, we may mitigate the effect of unmeasured confounders.

Based on Alyahyan and Düşteğör (2020), we identify several potential confounder groups that may affect both age of enrollment and the probability of academic success: pre-academic performance (such as high school grades, admission test results, and previous course grades); student demographics (including gender, race/ethnicity, socioeconomic status, and family background); student environment (such as class type, semester length, and type of program); and psychological factors (such as student interest, study behavior, stress, anxiety, time management, self-regulation, and motivation). In addition to these factors, macroeconomic indicators, like unemployment rates and inflation, are also considered necessary.

Our dataset contains most of the relevant classes of confounders, including pre-academic performance, student demographics, student environment, and macroeconomic indicators. However, we lack psychological factors, which could be a significant limitation in our analysis, as these factors may influence both age of enrollment and academic success. Furthermore, although academic progression is also an essential confounder, it is a post-treatment variable and, therefore, cannot be used in the causal analysis without risking post-treatment bias.

2.4 Common Support (Overlap)

The Common Support assumption states that each unit has a non-zero probability of receiving each treatment level, i.e., $\forall x \in X, P(T = 1|X = x) > 0$ and $P(T = 0|X = x) > 0$.

We empirically validate this assumption using propensity scores. We trained a logistic regression model to predict the treatment assignment based on the covariates and used the predicted probabilities as the propensity scores. We then plotted the propensity scores of the treated and control groups to ensure a significant overlap between the two groups. The results are presented in Figure 2. The lowest predicted propensity score of the treated group is 0.0131, and the control group’s is 0.0066. Therefore, we conclude that the common support assumption holds in our data.

The Common Support assumption is particularly plausible in the Portuguese context because, unlike countries with mandatory military service (such as Israel) or strict age-related entry requirements, Portugal maintains a flexible approach to higher education entry. The Portuguese higher education system allows individuals to enroll at various ages, with entry requirements primarily based on previous qualifications rather than age. This flexibility is evident in the diverse admission modes available, including special provisions for adult learners (over 23 years old), holders of various qualification types (from secondary education to specialized technical courses), and multiple application phases. This institutional framework helps ensure that individuals across different age groups have genuine opportunities to access higher education, supporting the validity of the Common Support assumption in our analysis.

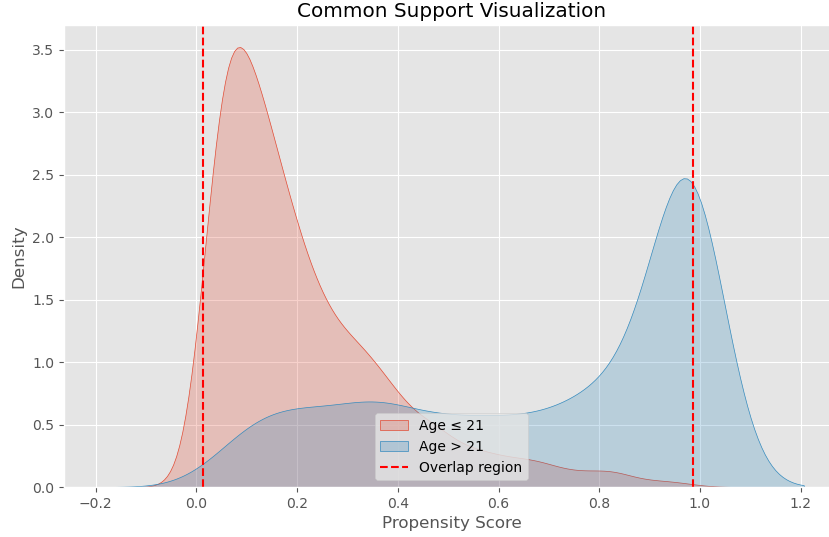
However, it is essential to note that not everyone receives similar opportunities to acquire higher education in Portugal. According to the OECD Review of Inclusive Education in Portugal (OECD, 2022), despite significant improvements in access and attainment, inclusion challenges persist for disadvantaged groups such as low-income students and those with immigrant backgrounds.

These disparities suggest that while there may be overall support for the common support assumption, certain groups might have systematically different opportunities or probabilities of accessing higher education and, therefore, receiving the treatment. This might challenge the common support assumption, particularly for vulnerable populations.

3 Causal Analysis Methodology

After validating the assumptions for causal inference, we describe the methodology used to perform the causal analysis. We present the relevant measures of causal effects and the methods we use to estimate them.

Figure 2: Common support of the propensity scores



3.1 Measures

To quantify the causal effect of the treatment on the outcome, one usually tries to estimate the Average Treatment Effect (ATE). However, in some cases, estimating the Average Treatment Effect on the Treated (ATT) or the Average Treatment Effect on the Control (ATC) can be more informative. We formally present those measures below.

Average Treatment Effect (ATE) The difference between the expected outcome under treatment and the expected outcome under control. Formally, it is defined as $ATE = E[Y_1 - Y_0]$.

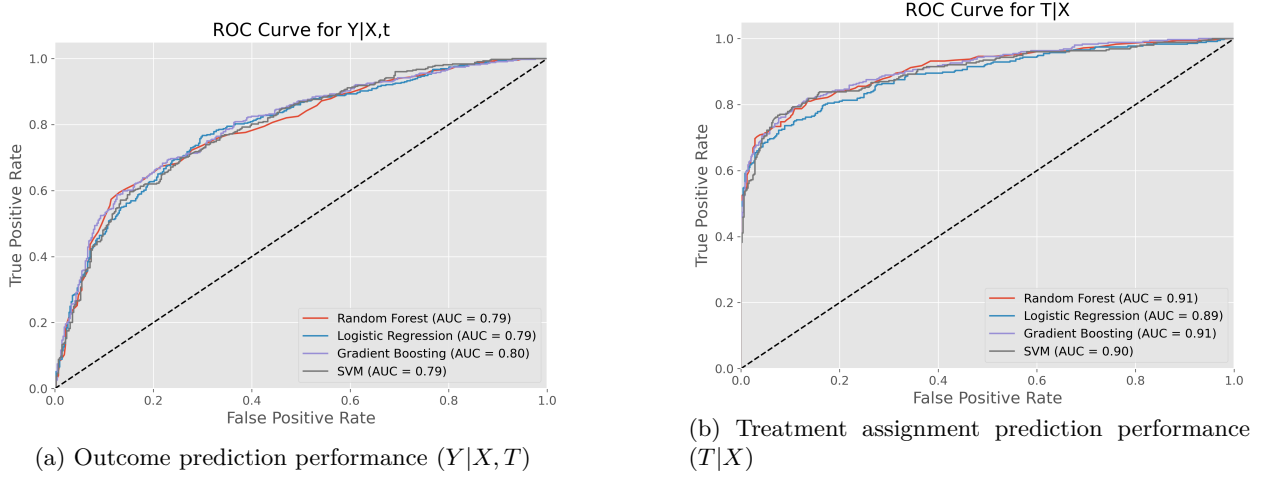
Average Treatment Effect on the Treated (ATT) The difference between the expected outcome under treatment and the expected outcome under control, but only for the treated units. Formally, it is defined as $ATT = E[Y_1 - Y_0 | T = 1]$.

Average Treatment Effect on the Control (ATC) The difference between the expected outcome under treatment and the expected outcome under control, but only for the control units. Formally, it is defined as $ATC = E[Y_1 - Y_0 | T = 0]$.

Each of the three measures provides a different perspective on the causal effect of the treatment. The ATE measures the average effect of the treatment on the entire population. In contrast, the ATT and ATC measures provide insights into the effect of the treatment on the treated and control units, respectively. The ATT and ATC measures are notably helpful when the treatment assignment is not controlled (not random), as in observational studies.

Bootstrap Confidence Intervals To assess the uncertainty in our estimated treatment effects, we employed bootstrap resampling to calculate confidence intervals for the ATE, ATT, and ATC. We perform 1000 bootstrap iterations by default, where in each iteration, we resample the entire dataset with replacement, maintaining the original sample size. We then apply our causal inference methods to this resampled dataset, computing the ATE, ATT, and ATC. After collecting these bootstrap estimates, we calculate the 95% confidence intervals using the percentile method, taking each effect's 2.5th and 97.5th percentiles of the bootstrap distribution.

Figure 3: Performance comparison of classification models



3.2 Methods

The fundamental problem of causal inference is that directly observing causal effects is impossible. For any given individual, we can only observe the outcome under one treatment condition - either treated or untreated. We never observe both potential outcomes for the same unit simultaneously, making it impossible to directly calculate individual treatment effects. In the remainder of this section, we present the estimation methods we used in our analysis, explain how they are trying to handle the mentioned problem, present the assumptions each requires for truthfully estimating the causal effects, and detail the implementation of each method.

3.2.1 Covariate Adjustment

Covariate adjustment methods use statistical models to regress the missing potential outcomes based on the observed covariates and the treatment. The regressed values are then used to estimate the average effects. There are different covariate adjustment methods, where the separation between them is usually found in if and how to separate the treatment variable from the covariates while regressing the missing potential outcomes and how to use the regressed values for the average effect estimation. We used the methods *S-Learner* and *T-Learner* for our analysis.

S-Learner Addresses the treatment variable T as one of the covariates. First, one must learn a statistical model for predicting the conditional average treatment effect (CATE) as a function of the covariates X and the treatment T , i.e., $\hat{\mu}(X, T) = \hat{\mathbb{E}}[Y^{obs}|X, T]$. Any ML method can be used for this purpose, but importantly, the whole population is used to learn the model. Then, use the learned model to estimate the average treatment effects as follows:

- *ATE*: Average the predicated causal effect on all of the population:

$$\widehat{ATE}^{SL} = \frac{1}{n} \sum_{i=1}^n \hat{\mu}(x_i, 1) - \hat{\mu}(x_i, 0)$$

- *ATT*: Average the predicted causal effect on the treatment group:

$$\widehat{ATT}^{SL} = \frac{1}{\sum_{i=1}^n t_i} \sum_{i=1}^n t_i [\hat{\mu}(x_i, 1) - \hat{\mu}(x_i, 0)]$$

- *ATC*: Average the predicted causal effect on the control group:

$$\widehat{ATC}^{SL} = \frac{1}{\sum_{i=1}^n 1 - t_i} \sum_{i=1}^n (1 - t_i) [\hat{\mu}(x_i, 1) - \hat{\mu}(x_i, 0)]$$

Hahn (1998) demonstrated that the S-Learner is consistent under the assumptions made in Section 2 and assuming that the model $\mu(X, T)$ is correctly specified. However, it may be prone to model misspecification, as noted by Rubin (1979). If the relationship between the covariates and the outcome is incorrectly modeled, it can lead to biased estimates of the treatment effects.

T-Learner To address the potential issues of model misspecification in the S-Learner, we also employ the T-Learner approach. The T-Learner uses separate models for the treatment and control groups, potentially reducing bias from model misspecification. It uses observations in the treatment group to estimate the response under treatment, $\hat{\mu}_1(x) = \hat{\mathbb{E}}[Y^{obs}|X, 1]$, and observations in the control group to estimate the response under control, $\hat{\mu}_0(x) = \hat{\mathbb{E}}[Y^{obs}|X, 0]$. Any machine learning method can be used to get these estimates. Calculating the average effects continues similarly to the S-Learner with $\hat{\mu}(X_i, T)$ replaced with $\hat{\mu}_T(x)$. Greenland and Robins (1986) showed that the T-Learner is consistent under the assumptions made in Section 2 and robust against misspecification of the outcome model. This robustness comes from the separate modeling of the treatment and control groups, which allows for different functional forms in each group.

Implementation Details Before applying our models, we standardized all numerical features using StandardScaler and one-hot encoded categorical variables. We utilized Gradient Boosting classifiers for both the S-Learner and T-Learner approaches. This choice was informed by our analysis of various classifiers’ performance in predicting the outcome variable Y given the covariates X and treatment T , as illustrated in Figure 3a. The Gradient Boosting classifier demonstrated the highest Area Under the Curve (AUC) score, indicating its superior predictive performance for our dataset. For the S-Learner, we trained a single Gradient Boosting model on the entire dataset, including the covariates and the treatment variable. For the T-Learner, we trained separate Gradient Boosting models for the treated and control groups. We used these models to predict outcomes for all individuals under both treatment conditions, allowing us to estimate average effects.

3.2.2 Propensity-Based Methods

One of the problems with covariate adjustment methods is the sensitivity to model specification and the sparsity of the covariates when handling high-dimensional data (Zhao et al., 2020). The propensity score methods of Rosenbaum and Rubin (RR) aim to address the fundamental problem by adjusting for the propensity score rather than potentially high-dimensional covariates. RR demonstrated that under the assumption of no unmeasured confounding, this is enough for unbiased estimation of causal effects. Such estimation can be accomplished using simple non-parametric methods (Abdia et al., 2017).

Inverse Probability Weighting (IPW) Relies on building a classification model to estimate the probability of the exposure observed (i.e., the propensity score) for a particular individual and using the predicted probability as a weight in subsequent analyses. Our analysis uses the Horvitz–Thompson estimator (Horvitz and Thompson, 1952). Intuitively, it estimates the target variable’s mean in the treatment and control groups by weighting the observations based on their “membership”. One must first train a classification model for predicting the propensity score $\hat{e}(x) = \hat{P}(T = 1|X = x)$. Then, use it to weigh each of the samples in the following way to receive each of the average effects:

- *ATE*: Compare the weighted means of the treatment and control groups, where units with higher (lower) propensity scores have more influence in the treatment (control) group.

$$\widehat{ATE}^{IPW} = \frac{1}{n} \sum_{i=1}^N \frac{t_i y_i}{\hat{e}(x)} - \frac{1}{n} \sum_{i=1}^N \frac{(1 - t_i) y_i}{1 - \hat{e}(x)}$$

- *ATT*: Compare the mean of the treatment group with a weighted mean of the control group, where units with higher propensity scores (more likely to be a part of the treatment group) have more influence (Lechner, 2001).

$$\widehat{ATT}^{IPW} = \frac{\sum_{i=1}^N t_i y_i}{\sum_{i=1}^N T_i} - \frac{\sum_{i=1}^N (1 - t_i) y_i \cdot \frac{\hat{e}(x_i)}{1 - \hat{e}(x_i)}}{\sum_{i=1}^N (1 - t_i) \cdot \frac{\hat{e}(x_i)}{1 - \hat{e}(x_i)}}$$

- *ATC*: Compare the mean of the control group with a weighted mean of the treatment group, where units with lower propensity scores (more likely to be a part of the control group) have more influence.

$$\widehat{ATC}^{IPW} = \frac{\sum_{i=1}^N t_i y_i \cdot \frac{1 - \hat{e}(x_i)}{\hat{e}(x_i)}}{\sum_{i=1}^N t_i \cdot \frac{1 - \hat{e}(x_i)}{\hat{e}(x_i)}} - \frac{\sum_{i=1}^N (1 - t_i) y_i}{\sum_{i=1}^N (1 - T_i)}$$

Propensity Score Matching (PSM) This method groups treated units with control units with similar propensity scores, effectively creating a matched sample where the distribution of observed baseline covariates is similar between treated and untreated subjects (Austin, 2011). Our implementation uses a nearest-neighbor approach with replacement, allowing multiple matches per unit. For each unit i , we find the n closest matches based on propensity scores. Let \mathcal{M}_i denote the set of indices of the n nearest neighbors for unit i in the opposite treatment group. We calculate weights w_{ij} for each match $j \in \mathcal{M}_i$ inversely proportional to the distance in propensity scores $w_{ij} = \frac{1/d_{ij}}{\sum_{k \in \mathcal{M}_i} 1/d_{ik}}$ where d_{ij} is the absolute difference in propensity scores between units i and j . The individual treatment effect for unit i is then calculated as

$$\tau_i = \begin{cases} y_i - \sum_{j \in \mathcal{M}_i} w_{ij} y_j & \text{if } t_i = 1 \\ \sum_{j \in \mathcal{M}_i} w_{ij} y_j - y_i & \text{if } t_i = 0 \end{cases}. \text{ The average effects are then calculated as follows:}$$

- *ATE*: Following Basu et al. (2023), the ATE estimator is $\widehat{ATE}^{PSM} = \frac{1}{N} \sum_{i=1}^N \tau_i$, where N is the total number of units. This represents the overall average treatment effect for the entire population, calculated by averaging the individual treatment effects over all units, regardless of whether they received the treatment.
- *ATT*: $\widehat{ATT}^{PSM} = \frac{1}{N_1} \sum_{i:t_i=1} \tau_i$, where N_1 is the number of treated units. This represents the average treatment effect for the treated group, calculated by averaging the individual treatment effects over all units that received the treatment.
- *ATC*: $\widehat{ATC}^{PSM} = \frac{1}{N_0} \sum_{i:t_i=0} \tau_i$, where N_0 is the number of control units. This represents the average treatment effect for the control group, calculated by averaging the individual treatment effects over all units that did not receive the treatment.

This method can be particularly effective in reducing bias due to confounding variables, especially when there is sufficient overlap in the propensity score distributions between the treated and control groups (Rosenbaum and Rubin, 1983).

Implementation Details The implementation uses the same scaled and encoded features as in the covariate adjustment methods, with a Random Forest classifier estimating the propensity scores. This choice was guided by classifier performance analysis for predicting the treatment assignment (T) given the covariates (X), as shown in Figure 3b. The Random Forest classifier achieved the highest AUC score, demonstrating its effectiveness in estimating propensity scores for this dataset. The IPW method applies these propensity scores to weigh the observations. To prevent issues with very small denominators, the implementation includes safeguards that clip propensity scores to a range of $[10^{-5}, 1 - 10^{-5}]$. The Propensity Score Matching implementation adopts a nearest-neighbor approach with replacement, allowing multiple matches per unit. Using 11 nearest neighbors per unit balances bias reduction and variance. The matching process incorporates weights inversely proportional to the propensity score distances, assigning greater importance to closer matches. This strategy leverages more information from the data than one-to-one matching, potentially yielding more precise estimates while maintaining effective bias reduction from confounding variables.

3.2.3 Doubly-Robust Method

The doubly robust method fortuitously combines both outcome regression and propensity score weighting, providing a consistent estimator for the average treatment effect even if either the propensity score model or the outcome regression model is misspecified, as long as one of the two is correctly specified (Bang and Robins, 2005). The estimator first requires the estimation of the propensity score, $\hat{e}(x_i)$, and the outcome models, $\hat{\mu}_1(X_i)$ and $\hat{\mu}_0(X_i)$. These are then combined in the following way to estimate the average treatment effects:

- *ATE*: Compares the weighted and augmented averages of the observed outcomes in the treated and control groups:

$$\widehat{ATE}^{DR} = \frac{1}{N} \sum_i \left(\frac{t_i(y_i - \hat{\mu}_1(x_i))}{\hat{e}(x_i)} + \hat{\mu}_1(x_i) \right) - \frac{1}{N} \sum_i \left(\frac{(1 - t_i)(y_i - \hat{\mu}_0(x_i))}{1 - \hat{e}(x_i)} + \hat{\mu}_0(x_i) \right)$$

- *ATT*: Following Tao and Fu (2019), the doubly robust estimator for the ATT compares the observed outcomes in the treated group with a weighted and augmented estimate of the control group's outcomes:

$$\widehat{ATT}^{DR} = \frac{\sum_i \left[t_i y_i - \frac{t_i - \hat{e}(x_i)}{1 - \hat{e}(x_i)} \hat{\mu}_0(x_i) \right]}{\sum_i t_i}$$

- *ATC*: Following Tao and Fu (2019), the doubly robust estimator for the ATC compares the observed outcomes in the control group and a weighted and augmented estimate of the treatment group's outcomes:

$$\widehat{ATC}^{DR} = \frac{\sum_i \left[\left\{ \frac{1 - \hat{e}(x_i)}{\hat{e}(x_i)} t_i y_i - \frac{t_i - \hat{e}(x_i)}{\hat{e}(x_i)} \hat{\mu}_1(x_i) \right\} - (1 - t_i) y_i \right]}{\sum_i 1 - t_i}$$

The doubly robust estimator uses "augmentation" terms, combining outcome modeling and propensity score weighting. For example, in the ATE estimator, the terms $(y_i - \hat{\mu}_1(x_i))$ and $(y_i - \hat{\mu}_0(x_i))$ represent the difference between observed outcomes and predicted outcomes. These augmentation terms act as "error corrections". If the outcome model is misspecified, the propensity score weighting helps correct the bias, and if the propensity score model is misspecified, the outcome model helps correct the bias. This dual correction mechanism provides the estimator's 'double robustness' property.

Implementation Details We leveraged the strengths of both the outcome regression and propensity score models identified in our previous analyses. We used Gradient Boosting to model the potential outcomes and Random Forest for the propensity score model. This decision is supported by Figure 3 and is consistent with the implementation in the previous methods.

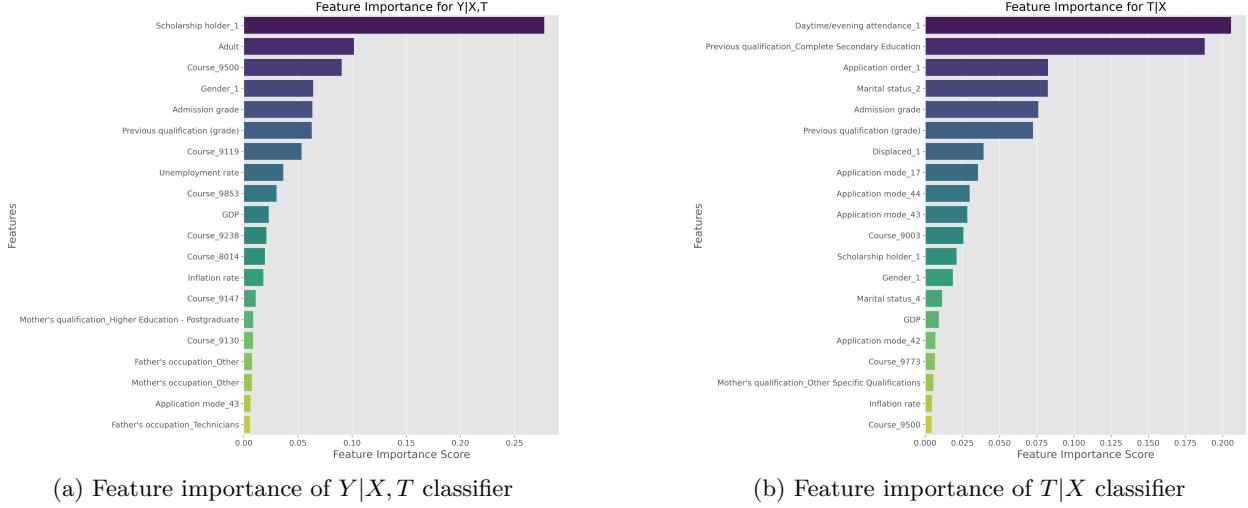
4 Results

In this section, we present our findings regarding the causal effect of adult enrollment on graduation probability, examining the relationships between variables through feature importance analysis and the estimated average treatment effects through various methodologies.

4.1 Feature Importance Analysis

The feature importance analysis was conducted using different approaches for the two classification tasks, as shown in Figure 4. We employed a Random Forest classifier to predict the treatment assignment ($T|X$) and analyzed feature importance using the built-in importance scores based on mean decrease in impurity (Gini importance). We used a Gradient Boosting classifier to predict the outcome given covariates and treatment ($Y|X, T$). We examined its feature importance metrics based on the accumulated reduction in the loss function across all trees.

Figure 4: Important features of classifiers



For the treatment prediction ($T|X$), as illustrated in Figure 4b, the most informative features are "Daytime/evening attendance" and "Previous qualification_Complete Secondary Education", with importance scores of approximately 0.205 and 0.187 respectively. This suggests these variables strongly correlate with whether a student enrolls as an adult. The application order, marital status, and admission grades also show substantial predictive power for the treatment assignment.

For the outcome prediction ($Y|X, T$), shown in Figure 4a, scholarship holder status emerges as the most informative feature with an importance score of about 0.277. Notably, the treatment variable ("Adult") appears as the second most informative predictor, with an importance score around 0.101, indicating its strong correlation with graduation outcomes. Course-specific variables (particularly Course_9500, which corresponds to Nursing) and gender also demonstrate substantial predictive power. Academic performance indicators such as admission grade and previous qualification grades are also among the top predictive features.

4.2 Average Effects Analysis

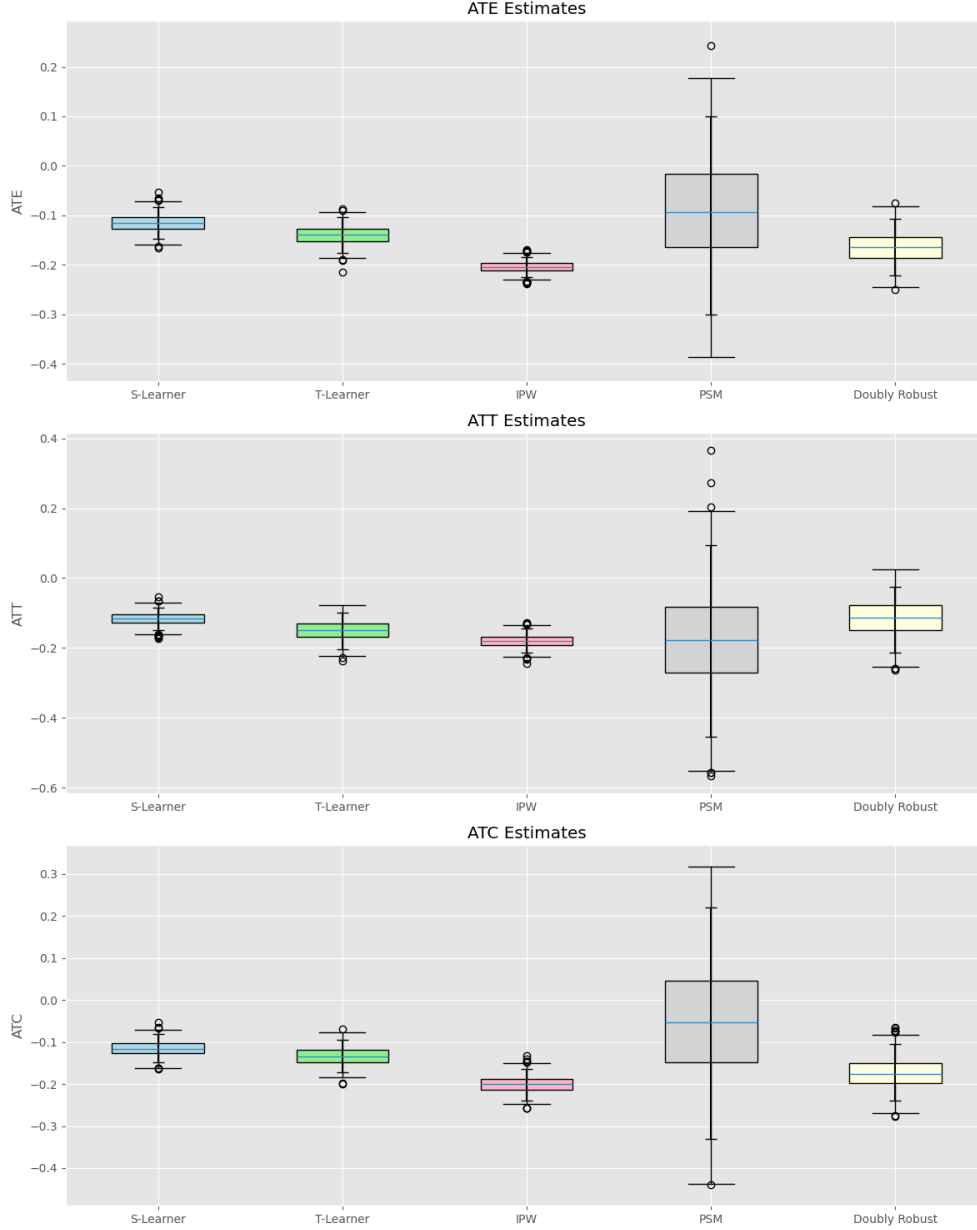
The box plots in Figure 5 provide a visualization of the distribution of estimated treatment effects across different methods. Each box represents the interquartile range (IQR) containing 50% of the estimates, with the horizontal line within each box showing the median estimate. The lower and upper edges of each box represent the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points that are not considered outliers (within 1.5 times the IQR), while individual points beyond the whiskers represent outlier estimates.

For each method, black error bars with caps represent the 95% confidence intervals. The middle point of these error bars indicates the mean estimate, while the upper and lower caps show the bounds of the confidence interval. The methods are distinguished by color coding: S-Learner (light blue), T-Learner (light green), IPW (light pink), PSM (light grey), and Doubly Robust (light yellow).

The vertical axis quantifies the magnitude of the effect, where negative values indicate that being an adult student reduces the probability of graduating, positive values suggest an increase in graduation probability, and zero represents no effect.

The box plots in Figure 5 reveal consistent patterns across different estimation methods, with some notable variations in both magnitude and uncertainty. For the Average Treatment Effect (ATE), both the S-Learner and T-Learner estimate an effect around -0.12 with relatively narrow confidence intervals. The IPW and Doubly Robust estimators also align closely, showing slightly more negative estimates around -0.17 . While the PSM method shows higher variance, its median estimate remains negative. Importantly, confidence intervals across all methods exclude zero, suggesting a robust negative effect.

Figure 5: Box plots of average effects



The patterns observed in the ATE estimates largely persist in both the ATT and ATC estimates, though with slightly different magnitudes. For both measures, the S-Learner and T-Learner continue to show strong agreement, with estimates around -0.11 for ATT and -0.12 for ATC. The IPW and Doubly Robust methods estimate slightly more negative effects in both cases. The PSM method consistently shows the highest variance among all methods, with confidence intervals barely including zero in the case of ATC. However, the overall negative trend remains consistent across all methods and measures.

The consistency of negative effects across all three measures (ATE, ATT, and ATC) and most estimation methods provides strong evidence for a negative relationship between adult enrollment and graduation probability. The similarity in magnitude across ATE, ATT, and ATC suggests this relationship is relatively uniform across different sub-populations. While the PSM method shows higher variance, its general alignment with the negative trend supports the robustness of these findings.

5 Discussion

This study investigated the causal effect of enrolling as an adult student (age 21 or older) on the probability of graduating from academic programs within the allotted time. Through rigorous analysis using multiple estimation methods and a comprehensive dataset from Portuguese higher education institutions, we found consistent evidence of a negative relationship between adult enrollment and graduation probability.

The results across different estimation methods (S-Learner, T-Learner, IPW, PSM, and Doubly Robust) consistently showed negative effects, with average treatment effects (ATE) ranging from approximately -0.12 to -0.17 . Those estimates suggest that, on average, adult students are 12–17 percentage points less likely to graduate within the allotted time than their younger counterparts. The consistency of these findings across different methodologies strengthens our confidence in this conclusion.

Notably, the similarity between the Average Treatment Effect on the Treated (ATT) and the Average Treatment Effect on the Controls (ATC) suggests that this negative effect is relatively uniform across the population. This implies that the challenges adult students face in completing their degrees within the expected time frame are not significantly different between those who currently choose to enroll as adults and those who don't.

Our analysis of feature importance showed that scholarship status, course type, and academic performance indicators are strong predictors of graduation outcomes. The treatment variable (adult status) emerged as the second most crucial predictor of graduation outcomes, further supporting its significance in academic success. Additionally, evening attendance and previous qualifications are strongly associated with adult enrollment, suggesting these factors play essential roles in the ability or decision to pursue higher education as an adult.

Several limitations should be considered when interpreting our results:

1. **Population Restriction:** Our analysis is based exclusively on Portuguese higher education institutions, which may limit the generalizability of our findings to other educational systems and cultural contexts. After data preprocessing, our sample consisted entirely of Portuguese nationals, restricting the population representation.
2. **Sample Size:** With 4,249 records after preprocessing, our sample size, while adequate for statistical analysis, may not capture the full diversity of student experiences and circumstances.
3. **Unmeasured Confounders:** Despite our comprehensive set of covariates, we lack data on important psychological factors such as motivation, stress levels, and study habits. These unmeasured confounders could potentially affect both the decision to enroll as an adult and academic success.
4. **SUTVA Assumption:** Our analysis assumes no interference between units, but the presence of adult students might affect the learning environment and outcomes of younger students, potentially violating this assumption.
5. **Post-Treatment Bias:** We had to exclude potentially informative variables about academic progression because they were measured after treatment assignment, which might have limited our ability to understand the mechanisms through which age affects graduation probability.
6. **Time-Varying Factors:** Our data represents a specific time period in Portuguese higher education, and the relationships we observed might change over time due to evolving educational policies and societal changes.

Several promising directions for future research emerge from our study:

1. **Cross-Cultural Validation:** Extending this analysis to educational systems in other countries would help understand how cultural and institutional differences affect the relationship between age and academic success.
2. **Longitudinal Studies:** Following students over extended periods would allow investigation of long-term outcomes beyond graduation, including career progression and lifetime earnings.

3. **Thorough Investigation:** Future studies could focus on understanding adult students' specific challenges, potentially through mixed-methods research incorporating qualitative data collection.
4. **Alternative Outcomes:** Expanding the analysis to consider other measures of academic success beyond graduation within the allotted time, such as final GPA or employment outcomes.
5. **Psychological Factors:** Including measures of psychological variables like motivation, stress, and study habits could provide a more complete understanding of the age-success relationship.

References

- Younathan Abdia, KB Kulasekera, Somnath Datta, Maxwell Boakye, and Maiying Kong. Propensity scores based methods for estimating average treatment effect and average treatment effect among treated: a comparative study. *Biometrical Journal*, 59(5):967–985, 2017.
- Eyman Alyahyan and Dilek Düşteğör. Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1):3, 2020.
- Peter C Austin. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3):399–424, 2011.
- Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.
- Anirban Basu, Aig Unuigbo, and Cristina Masseria. Understanding differences between what alternate propensity score methods estimate. *Journal of Managed Care & Specialty Pharmacy*, 29(4):391–399, 2023.
- Sander Greenland and James M Robins. Identifiability, exchangeability, and epidemiological confounding. *International journal of epidemiology*, 15(3):413–419, 1986.
- Jinyong Hahn. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, pages 315–331, 1998.
- Miguel A Hernán and James M Robins. Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health*, 60(7):578–586, 2006.
- Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952.
- Michael Lechner. *Identification and estimation of causal effects of multiple treatments under the conditional independence assumption*. Springer, 2001.
- OECD. *Review of Inclusive Education in Portugal*. 2022. doi: <https://doi.org/https://doi.org/10.1787/a9c95902-en>. URL <https://www.oecd-ilibrary.org/content/publication/a9c95902-en>.
- Valentim Realinho, Jorge Machado, Luís Baptista, and Mónica V. Martins. Predicting student dropout and academic success. *Data*, 7(11), 2022. ISSN 2306-5729. doi: 10.3390/data7110146. URL <https://www.mdpi.com/2306-5729/7/11/146>.
- Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.
- Donald B Rubin. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74(366a):318–328, 1979.
- Juliana Schulz, Erica EM Moodie, and Susan M Shortreed. No unmeasured confounding: Known unknowns or... not? *American Journal of Epidemiology*, 192(9):1604–1605, 2023.
- Yebin Tao and Haoda Fu. Doubly robust estimation of the weighted average treatment effect for a target population. *Statistics in medicine*, 38(3):315–325, 2019.

Shandong Zhao, David A van Dyk, and Kosuke Imai. Propensity score-based methods for causal inference in observational studies with non-binary treatments. *Statistical methods in medical research*, 29(3):709–727, 2020.

A Comprehensive Description of all the Features

The following pages contain a detailed description of all features in the dataset, including their types and possible values, as provided by the original dataset authors.

Dataset “Predict students' dropout and academic success”

Citation Requests/Acknowledgements

If you use this dataset in experiments for a scientific publication, please kindly cite our paper and/or the dataset available in Zenodo.

1. V. Realinho, J. Machado, L. Baptista, M. V. Martins. (2021). “Predict students' dropout and academic success” (1.0) [Data set]. Zenodo. DOI: [10.5281/zenodo.5777340](https://doi.org/10.5281/zenodo.5777340)
2. M. V. Martins, D. Tolledo, J. Machado, L. M. T. Baptista, V. Realinho. (2021) "Early prediction of student's performance in higher education: a case study" Trends and Applications in Information Systems and Technologies, vol.1, in Advances in Intelligent Systems and Computing series. Springer. DOI: [10.1007/978-3-030-72657-7_16](https://doi.org/10.1007/978-3-030-72657-7_16)

Features Information

Feature	Type	Description
Marital status	Numeric/discrete	1 – single 2 – married 3 – widower 4 – divorced 5 – facto union 6 – legally separated
Application mode	Numeric/discrete	1 - 1st phase - general contingent 2 - Ordinance No. 612/93 5 - 1st phase - special contingent (Azores Island) 7 - Holders of other higher courses 10 - Ordinance No. 854-B/99 15 - International student (bachelor) 16 - 1st phase - special contingent (Madeira Island) 17 - 2nd phase - general contingent 18 - 3rd phase - general contingent 26 - Ordinance No. 533-A/99, item b2) (Different Plan) 27 - Ordinance No. 533-A/99, item b3 (Other Institution) 39 - Over 23 years old 42 - Transfer 43 - Change of course 44 - Technological specialization diploma holders 51 - Change of institution/course 53 - Short cycle diploma holders 57 - Change of institution/course (International)
Application order	Numeric/discrete	Application order (between 0 - first choice; and 9 last choice)
Course	Numeric/discrete	33 - Biofuel Production Technologies 171 - Animation and Multimedia Design 8014 - Social Service (evening attendance) 9003 - Agronomy 9070 - Communication Design 9085 - Veterinary Nursing 9119 - Informatics Engineering 9130 - Equiculture 9147 - Management 9238 - Social Service 9254 - Tourism 9500 - Nursing 9556 - Oral Hygiene 9670 - Advertising and Marketing Management 9773 - Journalism and Communication 9853 - Basic Education 9991 - Management (evening attendance)
Daytime/evening attendance	Numeric/binary	1 – daytime 0 - evening

Previous qualification	Numeric/discrete	1 - Secondary education 2 - Higher education - bachelor's degree 3 - Higher education - degree 4 - Higher education - master's 5 - Higher education - doctorate 6 - Frequency of higher education 9 - 12th year of schooling - not completed 10 - 11th year of schooling - not completed 12 - Other - 11th year of schooling 14 - 10th year of schooling 15 - 10th year of schooling - not completed 19 - Basic education 3rd cycle (9th/10th/11th year) or equiv. 38 - Basic education 2nd cycle (6th/7th/8th year) or equiv. 39 - Technological specialization course 40 - Higher education - degree (1st cycle) 42 - Professional higher technical course 43 - Higher education - master (2nd cycle)
Previous qualification (grade)	Numeric/continuous	Grade of previous qualification (between 0 and 200)
Nationality	Numeric/discrete	1 - Portuguese 2 - German 6 - Spanish 11 - Italian 13 - Dutch 14 - English 17 - Lithuanian 21 - Angolan 22 - Cape Verdean 24 - Guinean 25 - Mozambican 26 - Santomean 32 - Turkish 41 - Brazilian 62 - Romanian 100 - Moldova (Republic of) 101 - Mexican 103 - Ukrainian 105 - Russian 108 - Cuban 109 - Colombian
Mother's qualification	Numeric/discrete	1 - Secondary Education - 12th Year of Schooling or Eq. 2 - Higher Education - Bachelor's Degree 3 - Higher Education - Degree 4 - Higher Education - Master's 5 - Higher Education - Doctorate 6 - Frequency of Higher Education 9 - 12th Year of Schooling - Not Completed 10 - 11th Year of Schooling - Not Completed 11 - 7th Year (Old) 12 - Other - 11th Year of Schooling 14 - 10th Year of Schooling 18 - General commerce course 19 - Basic Education 3rd Cycle (9th/10th/11th Year) or Equiv. 22 - Technical-professional course 26 - 7th year of schooling 27 - 2nd cycle of the general high school course 29 - 9th Year of Schooling - Not Completed 30 - 8th year of schooling 34 - Unknown 35 - Can't read or write 36 - Can read without having a 4th year of schooling 37 - Basic education 1st cycle (4th/5th year) or equiv. 38 - Basic Education 2nd Cycle (6th/7th/8th Year) or Equiv. 39 - Technological specialization course 40 - Higher education - degree (1st cycle) 41 - Specialized higher studies course 42 - Professional higher technical course 43 - Higher Education - Master (2nd cycle) 44 - Higher Education - Doctorate (3rd cycle)
Father's qualification	Numeric/discrete	1 - Secondary Education - 12th Year of Schooling or Eq. 2 - Higher Education - Bachelor's Degree

		3 - Higher Education - Degree 4 - Higher Education - Master's 5 - Higher Education - Doctorate 6 - Frequency of Higher Education 9 - 12th Year of Schooling - Not Completed 10 - 11th Year of Schooling - Not Completed 11 - 7th Year (Old) 12 - Other - 11th Year of Schooling 13 - 2nd year complementary high school course 14 - 10th Year of Schooling 18 - General commerce course 19 - Basic Education 3rd Cycle (9th/10th/11th Year) or Equiv. 20 - Complementary High School Course 22 - Technical-professional course 25 - Complementary High School Course - not concluded 26 - 7th year of schooling 27 - 2nd cycle of the general high school course 29 - 9th Year of Schooling - Not Completed 30 - 8th year of schooling 31 - General Course of Administration and Commerce 33 - Supplementary Accounting and Administration 34 - Unknown 35 - Can't read or write 36 - Can read without having a 4th year of schooling 37 - Basic education 1st cycle (4th/5th year) or equiv. 38 - Basic Education 2nd Cycle (6th/7th/8th Year) or Equiv. 39 - Technological specialization course 40 - Higher education - degree (1st cycle) 41 - Specialized higher studies course 42 - Professional higher technical course 43 - Higher Education - Master (2nd cycle) 44 - Higher Education - Doctorate (3rd cycle)
Mother's occupation	Numeric/discrete	0 - Student 1 - Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers 2 - Specialists in Intellectual and Scientific Activities 3 - Intermediate Level Technicians and Professions 4 - Administrative staff 5 - Personal Services, Security and Safety Workers and Sellers 6 - Farmers and Skilled Workers in Agriculture, Fisheries and Forestry 7 - Skilled Workers in Industry, Construction and Craftsmen 8 - Installation and Machine Operators and Assembly Workers 9 - Unskilled Workers 10 - Armed Forces Professions 90 - Other Situation 99 - (blank) 122 - Health professionals 123 - teachers 125 - Specialists in information and communication technologies (ICT) 131 - Intermediate level science and engineering technicians and professions 132 - Technicians and professionals, of intermediate level of health 134 - Intermediate level technicians from legal, social, sports, cultural and similar services 141 - Office workers, secretaries in general and data processing operators 143 - Data, accounting, statistical, financial services and registry-related operators 144 - Other administrative support staff 151 - personal service workers 152 - sellers 153 - Personal care workers and the like 171 - Skilled construction workers and the like, except electricians 173 - Skilled workers in printing, precision instrument manufacturing, jewelers, artisans and the like 175 - Workers in food processing, woodworking, clothing and other industries and crafts 191 - cleaning workers 192 - Unskilled workers in agriculture, animal production, fisheries and forestry 193 - Unskilled workers in extractive industry, construction, manufacturing and transport 194 - Meal preparation assistants

Father's occupation	Numeric/discrete	0 - Student 1 - Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers 2 - Specialists in Intellectual and Scientific Activities 3 - Intermediate Level Technicians and Professions 4 - Administrative staff 5 - Personal Services, Security and Safety Workers and Sellers 6 - Farmers and Skilled Workers in Agriculture, Fisheries and Forestry 7 - Skilled Workers in Industry, Construction and Craftsmen 8 - Installation and Machine Operators and Assembly Workers 9 - Unskilled Workers 10 - Armed Forces Professions 90 - Other Situation 99 - (blank) 101 - Armed Forces Officers 102 - Armed Forces Sergeants 103 - Other Armed Forces personnel 112 - Directors of administrative and commercial services 114 - Hotel, catering, trade and other services directors 121 - Specialists in the physical sciences, mathematics, engineering and related techniques 122 - Health professionals 123 - teachers 124 - Specialists in finance, accounting, administrative organization, public and commercial relations 131 - Intermediate level science and engineering technicians and professions 132 - Technicians and professionals, of intermediate level of health 134 - Intermediate level technicians from legal, social, sports, cultural and similar services 135 - Information and communication technology technicians 141 - Office workers, secretaries in general and data processing operators 143 - Data, accounting, statistical, financial services and registry-related operators 144 - Other administrative support staff 151 - personal service workers 152 - sellers 153 - Personal care workers and the like 154 - Protection and security services personnel 161 - Market-oriented farmers and skilled agricultural and animal production workers 163 - Farmers, livestock keepers, fishermen, hunters and gatherers, subsistence 171 - Skilled construction workers and the like, except electricians 172 - Skilled workers in metallurgy, metalworking and similar 174 - Skilled workers in electricity and electronics 175 - Workers in food processing, woodworking, clothing and other industries and crafts 181 - Fixed plant and machine operators 182 - assembly workers 183 - Vehicle drivers and mobile equipment operators 192 - Unskilled workers in agriculture, animal production, fisheries and forestry 193 - Unskilled workers in extractive industry, construction, manufacturing and transport 194 - Meal preparation assistants 195 - Street vendors (except food) and street service providers
Admission grade	Numeric/continous	Admission grade (between 0 and 200)
Displaced	Numeric/binary	1 – yes 0 – no
Educational special needs	Numeric/binary	1 – yes 0 – no
Debtor	Numeric/binary	1 – yes 0 – no
Tuition fees up to date	Numeric/binary	1 – yes 0 – no
Gender	Numeric/binary	1 – male 0 – female
Scholarship holder	Numeric/binary	1 – yes 0 – no
Age at enrollment	Numeric/discrete	Age of student at enrollment

International	Numeric/binary	1 – yes 0 – no
Curricular units 1st sem (credited)	Numeric/discrete	Number of curricular units credited in the 1st semester
Curricular units 1st sem (enrolled)	Numeric/discrete	Number of curricular units enrolled in the 1st semester
Curricular units 1st sem (evaluations)	Numeric/discrete	Number of evaluations to curricular units in the 1st semester
Curricular units 1st sem (approved)	Numeric/discrete	Number of curricular units approved in the 1st semester
Curricular units 1st sem (grade)	Numeric/discrete	Grade average in the 1st semester (between 0 and 20)
Curricular units 1st sem (without evaluations)	Numeric/discrete	Number of curricular units without evaluations in the 1st semester
Curricular units 2nd sem (credited)	Numeric/discrete	Number of curricular units credited in the 2nd semester
Curricular units 2nd sem (enrolled)	Numeric/discrete	Number of curricular units enrolled in the 2nd semester
Curricular units 2nd sem (evaluations)	Numeric/discrete	Number of evaluations to curricular units in the 2nd semester
Curricular units 2nd sem (approved)	Numeric/discrete	Number of curricular units approved in the 2nd semester
Curricular units 2nd sem (grade)	Numeric/discrete	Grade average in the 2nd semester (between 0 and 20)
Curricular units 2nd sem (without evaluations)	Numeric/discrete	Number of curricular units without evaluations in the 1st semester
Unemployment rate	Numeric/continuous	Unemployment rate (%)
Inflation rate	Numeric/continuous	Inflation rate (%)
GDP	Numeric/continuous	GDP
Target	Categorical	Target. The problem is formulated as a three category classification task (dropout, enrolled, and graduate) at the end of the normal duration of the course.

The following table contains a partition of the features to the classes: demographic, socioeconomic, macroeconomic, academic data at enrollment, and academic data at the end of the first and second semesters.

Table 1: Attributes used grouped by class of attribute.

Class of Attribute	Attribute	Type
Demographic data	Marital status	Numeric/discrete
	Nationality	Numeric/discrete
	Displaced	Numeric/binary
	Gender	Numeric/binary
	Age at enrollment	Numeric/discrete
	International	Numeric/binary
Socioeconomic data	Mother's qualification	Numeric/discrete
	Father's qualification	Numeric/discrete
	Mother's occupation	Numeric/discrete
	Father's occupation	Numeric/discrete
	Educational special needs	Numeric/binary
	Debtor	Numeric/binary
	Tuition fees up to date	Numeric/binary
	Scholarship holder	Numeric/binary
Macroeconomic data	Unemployment rate	Numeric/continuous
	Inflation rate	Numeric/continuous
	GDP	Numeric/continuous
Academic data at enrollment	Application mode	Numeric/discrete
	Application order	Numeric/ordinal
	Course	Numeric/discrete
	Daytime/evening attendance	Numeric/binary
	Previous qualification	Numeric/discrete
Academic data at the end of 1st semester	Curricular units 1st sem (credited)	Numeric/discrete
	Curricular units 1st sem (enrolled)	Numeric/discrete
	Curricular units 1st sem (evaluations)	Numeric/discrete
	Curricular units 1st sem (approved)	Numeric/discrete
	Curricular units 1st sem (grade)	Numeric/continuous
	Curricular units 1st sem (without evaluations)	Numeric/discrete
Academic data at the end of 2nd semester	Curricular units 2nd sem (credited)	Numeric/discrete
	Curricular units 2nd sem (enrolled)	Numeric/discrete
	Curricular units 2nd sem (evaluations)	Numeric/discrete
	Curricular units 2nd sem (approved)	Numeric/discrete
	Curricular units 2nd sem (grade)	Numeric/continuous
	Curricular units 2nd sem (without evaluations)	Numeric/discrete
Target	Target	Categorical