

# The effect of age of enrollment on the probability of graduating in courses - Confounders analysis

Gur Keinan 213635899, Yarden Adi 212585848

August 13, 2024

In this work we analyze the effect of the age of enrollment on the probability of graduating in courses. We use the data from the Kaggle competition Predicting Graduation Outcome in Courses, and was presented in ?. This dataset contains information about students' demographics, academic performance, and other relevant features. The target variable is the graduation outcome, which can be one of three categories: dropout, enrolled, or graduated. The kaggle competition was about predicting the graduation outcome based on the available features.

We start from the confounders that may affect the probability of graduating in courses. Following ?, we identify the following confounders groups:

- **Pre-academic performance:** The student's academic performance before the course. May include high school grades, admission test results, and previous courses grades.
- **Student demographics:** The student's demographic information. May include gender, age, race/ethnicity, socioeconomic status, and family background.
- **Student environment:** The student's environment during the course. May include class type, semester duration, and type of program.
- **Psychological factors:** The student's psychological factors. May include student interest, behavior of study, stress, anxiety, time of preoccupation, self-regulation, and motivation.

In addition, we think that Academic progression and Macroeconomic indicators are also important confounders that may affect the probability of graduating in courses.

As for confounders that may affect the age of enrollment, we identify the following groups:

- **Pre-academic performance:** The student's academic performance before the course. May include high school grades, admission test results, and previous courses grades.
- **Student demographics:** The student's demographic information. May include financial status, family background, and cultural background.
- **Student environment:** The student's environment during the course. Can include special programs ("Atuda"), day/night classes, basic / advanced courses.
- **Psychological factors:** The student's psychological factors. May include student interest, behavior of study, stress, anxiety, time of preoccupation, self-regulation, and motivation.

- **External factors:** Factors that are outside the student's control but may significantly influence their age of enrollment. These can include socio-political changes, economic conditions, and shifts in education policy. For instance, economic recessions might delay enrollment as students may need to work before continuing their studies, or educational reforms might alter the standard age of enrollment.

After a thorough analysis of the data (while using Features information), we categorize the confounders as follows:

- **Pre-academic performance:**

1. Previous qualification: The level of the student's previous academic qualification, such as secondary education or higher education degrees.
2. Previous qualification (grade): The grade associated with the student's previous qualification, providing a quantitative measure of past academic performance.
3. Admission grade: The student's grade upon admission, which could be an indicator of their academic capabilities at the start of the course.
4. Scholarship holder: Whether the student is a recipient of a scholarship, which might correlate with their academic merit or financial need.

- **Student demographics:**

1. Gender: The gender of the student.
2. Marital status: The marital status of the student, which may impact their availability and focus.
3. Nationality: The nationality of the student, which could be a proxy for cultural and language differences that might influence academic performance.
4. Age at enrollment: The primary variable of interest, representing the student's age at the time of enrollment.
5. International: Whether the student is an international student, which may bring additional challenges such as cultural and language barriers.
6. Mother's qualification: The educational level of the student's mother, which may be a proxy for family support and educational environment.
7. Father's qualification: The educational level of the student's father, similar to the mother's qualification.
8. Mother's occupation: The occupation of the student's mother, which might influence the student's socioeconomic status and support.
9. Father's occupation: The occupation of the student's father, similar to the mother's occupation.

- **Student environment:**

1. Course: The specific course the student is enrolled in, which may have varying levels of difficulty.
2. Daytime/evening attendance: Whether the student attends classes during the day or in the evening, which might reflect their time availability and external responsibilities.

3. Displaced: Whether the student lives away from home during their studies, which could influence their support systems and stress levels.
  4. Educational special needs: Whether the student has any special educational needs that require additional support.
  5. Tuition fees up to date: Whether the student's tuition fees are up to date, which could indicate financial stability.
  6. Debtor: Whether the student has outstanding debts, which could cause stress and affect academic performance.
- **Psychological factors:** Those are quite complex to measure or to keep track of, one of our weak points.
  - **Academic progression:**
    1. Curricular units 1st sem (credited): The number of curricular units credited in the 1st semester.
    2. Curricular units 1st sem (enrolled): The number of curricular units the student enrolled in during the 1st semester.
    3. Curricular units 1st sem (evaluations): The number of evaluations the student underwent in the 1st semester.
    4. Curricular units 1st sem (approved): The number of curricular units the student passed in the 1st semester.
    5. Curricular units 1st sem (grade): The average grade of the student in the 1st semester.
    6. Curricular units 1st sem (without evaluations): The number of curricular units the student enrolled in but did not undergo evaluations for in the 1st semester.
    7. Curricular units 2nd sem (credited): The number of curricular units credited in the 2nd semester.
    8. Curricular units 2nd sem (enrolled): The number of curricular units the student enrolled in during the 2nd semester.
    9. Curricular units 2nd sem (evaluations): The number of evaluations the student underwent in the 2nd semester.
    10. Curricular units 2nd sem (approved): The number of curricular units the student passed in the 2nd semester.
    11. Curricular units 2nd sem (grade): The average grade of the student in the 2nd semester.
    12. Curricular units 2nd sem (without evaluations): The number of curricular units the student enrolled in but did not undergo evaluations for in the 2nd semester.
  - **Macroeconomic indicators:**
    1. Unemployment rate: The unemployment rate during the period of study, which might impact the student's stress levels and focus.
    2. Inflation rate: The inflation rate during the period of study, which could influence the student's financial stability.
    3. GDP: The Gross Domestic Product during the period of study, reflecting the overall economic environment that may affect the student's circumstances.

This dataset encompasses a wide variety of variables across all relevant categories, allowing for a comprehensive analysis of the factors that influence the probability of graduating in courses. We believe that the extensive coverage of these variables will help mitigate the impact of unknown confounders. Although some potential confounders may remain unobserved, we are confident that they are likely correlated with the variables we already possess, thereby reducing their potential bias in our analysis.