IST 707 Final Project - Student Educational Outcomes

Christoper Murphy, Padmaja Kurumaddali, La Monte Yarroll

Septemner 07, 2024

# Introduction

# Analysis and Models

## About the data

### Dataset Summary

The following dataset was located on Kaggle at the following page:

<https://www.kaggle.com/code/mdismielhossenabir/student-s-dropout-and-academic-success>

The original source at UCI is here:

<https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success>

This dataset contains information on students’ marital status, application details, previous qualifications, family background, admission grades, financial status, and academic performance. The target variable indicates whether a student dropped out or achieved academic success. The data spans multiple semesters and includes economic indicators such as unemployment rate, inflation rate, and GDP. Researchers can leverage this dataset to develop models for predicting dropout and academic success, contributing to the understanding of factors influencing educational outcomes.

### Data Retrieval

Data structure structure when read in from the .CSV file:

## 'data.frame': 4424 obs. of 37 variables:  
## $ Marital.status : int 1 1 1 1 2 2 1 1 1 1 ...  
## $ Application.mode : int 17 15 1 17 39 39 1 18 1 1 ...  
## $ Application.order : int 5 1 5 2 1 1 1 4 3 1 ...  
## $ Course : int 171 9254 9070 9773 8014 9991 9500 9254 9238 9238 ...  
## $ Daytime.evening.attendance. : int 1 1 1 1 0 0 1 1 1 1 ...  
## $ Previous.qualification : int 1 1 1 1 1 19 1 1 1 1 ...  
## $ Previous.qualification..grade. : num 122 160 122 122 100 ...  
## $ Nacionality : int 1 1 1 1 1 1 1 1 62 1 ...  
## $ Mother.s.qualification : int 19 1 37 38 37 37 19 37 1 1 ...  
## $ Father.s.qualification : int 12 3 37 37 38 37 38 37 1 19 ...  
## $ Mother.s.occupation : int 5 3 9 5 9 9 7 9 9 4 ...  
## $ Father.s.occupation : int 9 3 9 3 9 7 10 9 9 7 ...  
## $ Admission.grade : num 127 142 125 120 142 ...  
## $ Displaced : int 1 1 1 1 0 0 1 1 0 1 ...  
## $ Educational.special.needs : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Debtor : int 0 0 0 0 0 1 0 0 0 1 ...  
## $ Tuition.fees.up.to.date : int 1 0 0 1 1 1 1 0 1 0 ...  
## $ Gender : int 1 1 1 0 0 1 0 1 0 0 ...  
## $ Scholarship.holder : int 0 0 0 0 0 0 1 0 1 0 ...  
## $ Age.at.enrollment : int 20 19 19 20 45 50 18 22 21 18 ...  
## $ International : int 0 0 0 0 0 0 0 0 1 0 ...  
## $ Curricular.units.1st.sem..credited. : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Curricular.units.1st.sem..enrolled. : int 0 6 6 6 6 5 7 5 6 6 ...  
## $ Curricular.units.1st.sem..evaluations. : int 0 6 0 8 9 10 9 5 8 9 ...  
## $ Curricular.units.1st.sem..approved. : int 0 6 0 6 5 5 7 0 6 5 ...  
## $ Curricular.units.1st.sem..grade. : num 0 14 0 13.4 12.3 ...  
## $ Curricular.units.1st.sem..without.evaluations.: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Curricular.units.2nd.sem..credited. : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Curricular.units.2nd.sem..enrolled. : int 0 6 6 6 6 5 8 5 6 6 ...  
## $ Curricular.units.2nd.sem..evaluations. : int 0 6 0 10 6 17 8 5 7 14 ...  
## $ Curricular.units.2nd.sem..approved. : int 0 6 0 5 6 5 8 0 6 2 ...  
## $ Curricular.units.2nd.sem..grade. : num 0 13.7 0 12.4 13 ...  
## $ Curricular.units.2nd.sem..without.evaluations.: int 0 0 0 0 0 5 0 0 0 0 ...  
## $ Unemployment.rate : num 10.8 13.9 10.8 9.4 13.9 16.2 15.5 15.5 16.2 8.9 ...  
## $ Inflation.rate : num 1.4 -0.3 1.4 -0.8 -0.3 0.3 2.8 2.8 0.3 1.4 ...  
## $ GDP : num 1.74 0.79 1.74 -3.12 0.79 -0.92 -4.06 -4.06 -0.92 3.51 ...  
## $ Target : chr "Dropout" "Graduate" "Dropout" "Graduate" ...

#### Initial Data Sample

A sample set of records:

## Marital.status Application.mode Application.order Course  
## 1 1 17 5 171  
## 2 1 15 1 9254  
## 3 1 1 5 9070  
## 4 1 17 2 9773  
## 5 2 39 1 8014  
## Daytime.evening.attendance. Previous.qualification  
## 1 1 1  
## 2 1 1  
## 3 1 1  
## 4 1 1  
## 5 0 1  
## Previous.qualification..grade. Nacionality Mother.s.qualification  
## 1 122 1 19  
## 2 160 1 1  
## 3 122 1 37  
## 4 122 1 38  
## 5 100 1 37  
## Father.s.qualification Mother.s.occupation Father.s.occupation  
## 1 12 5 9  
## 2 3 3 3  
## 3 37 9 9  
## 4 37 5 3  
## 5 38 9 9  
## Admission.grade Displaced Educational.special.needs Debtor  
## 1 127.3 1 0 0  
## 2 142.5 1 0 0  
## 3 124.8 1 0 0  
## 4 119.6 1 0 0  
## 5 141.5 0 0 0  
## Tuition.fees.up.to.date Gender Scholarship.holder Age.at.enrollment  
## 1 1 1 0 20  
## 2 0 1 0 19  
## 3 0 1 0 19  
## 4 1 0 0 20  
## 5 1 0 0 45  
## International Curricular.units.1st.sem..credited.  
## 1 0 0  
## 2 0 0  
## 3 0 0  
## 4 0 0  
## 5 0 0  
## Curricular.units.1st.sem..enrolled. Curricular.units.1st.sem..evaluations.  
## 1 0 0  
## 2 6 6  
## 3 6 0  
## 4 6 8  
## 5 6 9  
## Curricular.units.1st.sem..approved. Curricular.units.1st.sem..grade.  
## 1 0 0.00000  
## 2 6 14.00000  
## 3 0 0.00000  
## 4 6 13.42857  
## 5 5 12.33333  
## Curricular.units.1st.sem..without.evaluations.  
## 1 0  
## 2 0  
## 3 0  
## 4 0  
## 5 0  
## Curricular.units.2nd.sem..credited. Curricular.units.2nd.sem..enrolled.  
## 1 0 0  
## 2 0 6  
## 3 0 6  
## 4 0 6  
## 5 0 6  
## Curricular.units.2nd.sem..evaluations. Curricular.units.2nd.sem..approved.  
## 1 0 0  
## 2 6 6  
## 3 0 0  
## 4 10 5  
## 5 6 6  
## Curricular.units.2nd.sem..grade.  
## 1 0.00000  
## 2 13.66667  
## 3 0.00000  
## 4 12.40000  
## 5 13.00000  
## Curricular.units.2nd.sem..without.evaluations. Unemployment.rate  
## 1 0 10.8  
## 2 0 13.9  
## 3 0 10.8  
## 4 0 9.4  
## 5 0 13.9  
## Inflation.rate GDP Target  
## 1 1.4 1.74 Dropout  
## 2 -0.3 0.79 Graduate  
## 3 1.4 1.74 Dropout  
## 4 -0.8 -3.12 Graduate  
## 5 -0.3 0.79 Graduate

#### Check for Missing Values

## Marital.status   
## 0   
## Application.mode   
## 0   
## Application.order   
## 0   
## Course   
## 0   
## Daytime.evening.attendance.   
## 0   
## Previous.qualification   
## 0   
## Previous.qualification..grade.   
## 0   
## Nacionality   
## 0   
## Mother.s.qualification   
## 0   
## Father.s.qualification   
## 0   
## Mother.s.occupation   
## 0   
## Father.s.occupation   
## 0   
## Admission.grade   
## 0   
## Displaced   
## 0   
## Educational.special.needs   
## 0   
## Debtor   
## 0   
## Tuition.fees.up.to.date   
## 0   
## Gender   
## 0   
## Scholarship.holder   
## 0   
## Age.at.enrollment   
## 0   
## International   
## 0   
## Curricular.units.1st.sem..credited.   
## 0   
## Curricular.units.1st.sem..enrolled.   
## 0   
## Curricular.units.1st.sem..evaluations.   
## 0   
## Curricular.units.1st.sem..approved.   
## 0   
## Curricular.units.1st.sem..grade.   
## 0   
## Curricular.units.1st.sem..without.evaluations.   
## 0   
## Curricular.units.2nd.sem..credited.   
## 0   
## Curricular.units.2nd.sem..enrolled.   
## 0   
## Curricular.units.2nd.sem..evaluations.   
## 0   
## Curricular.units.2nd.sem..approved.   
## 0   
## Curricular.units.2nd.sem..grade.   
## 0   
## Curricular.units.2nd.sem..without.evaluations.   
## 0   
## Unemployment.rate   
## 0   
## Inflation.rate   
## 0   
## GDP   
## 0   
## Target   
## 0

### Data Attribute Summary

The Kaggle data card can also be found on the following page <https://www.kaggle.com/code/mdismielhossenabir/student-s-dropout-and-academic-success/input>

### Attribute Summary:

* Marital Status: The marital status of the student
* Application mode: The mode through which the student applied for admission
* Application order: The order in which the student applied for admission.
* Course: The academic course or program the student is enrolled in.
* Daytime/evening attendance: The attendance mode of the student (e.g., Daytime, Evening).
* Previous qualification: The highest academic qualification the student attained before the current enrollment.
* Previous qualification (grade): The grade or score achieved in the previous qualification.
* Nacionality: The nationality of the student.
* Mother’s qualification: The highest academic qualification of the student’s mother.
* Father’s qualification: The highest academic qualification of the student’s father.
* Mother’s occupation: The occupation of the student’s mother.
* Father’s occupation: The occupation of the student’s father.
* Admission grade: The grade or score obtained during the admission process.
* Displaced: Indicates whether the student has been displaced from their previous location.
* Educational special needs: Flags if the student has any educational special needs.
* Debtor: Indicates if the student is a debtor.
* Tuition fees up to date: Indicates whether the student’s tuition fees are up to date.
* Gender: The gender of the student.
* Scholarship holder: Flags if the student is a holder of a scholarship.
* Age at enrollment: The age of the student at the time of enrollment.
* International: Indicates if the student is an international student.
* Curricular units 1st sem (credited): The number of curricular units credited in the first semester.
* Curricular units 1st sem (enrolled): The number of curricular units enrolled in the first semester.
* Curricular units 1st sem (evaluations): The number of evaluations conducted in the first semester.
* Curricular units 1st sem (approved): The number of curricular units approved in the first semester.
* Curricular units 1st sem (grade): The overall grade for curricular units in the first semester.
* Curricular units 1st sem (without evaluations): Curricular units in the first semester without evaluations.
* Curricular units 2nd sem (credited): The number of curricular units credited in the second semester.
* Curricular units 2nd sem (enrolled): The number of curricular units enrolled in the second semester.
* Curricular units 2nd sem (evaluations): The number of evaluations conducted in the second semester.
* Curricular units 2nd sem (approved): The number of curricular units approved in the second semester.
* Curricular units 2nd sem (grade): The overall grade for curricular units in the second semester.
* Curricular units 2nd sem (without evaluations): Curricular units in the second semester without evaluations.
* Unemployment rate: The unemployment rate at the time of data collection.
* Inflation rate: The inflation rate at the time of data collection.
* GDP: The Gross Domestic Product at the time of data collection.
* Target: The target variable indicating student outcome (e.g., Dropout, Academic Success).

### Data Attribute Analysis

The following section explores the nature of the attributes in the dataset.

#### Factors

#### Data with factors

summary(studentSuccessData)

## Marital.status Application.mode  
## Single :3919 1st phase - general contingent :1708   
## Married : 379 2nd phase - general contingent : 872   
## Widower : 4 Over 23 years old : 785   
## Divorced : 91 Change of course : 312   
## Facto Union : 25 Technological specialization diploma holders: 213   
## Legally Separated: 6 Holders of other higher courses : 139   
## (Other) : 395   
## Application.order Course   
## 1st :3026 Nursing : 766   
## 2nd : 547 Management : 380   
## 3rd : 309 Social Service : 355   
## 4th : 249 Veterinary Nursing : 337   
## 5th : 154 Journalism and Communication : 331   
## 6th : 137 Advertising and Marketing Management: 268   
## NA's: 2 (Other) :1987   
## Daytime.evening.attendance.  
## Min. :0.0000   
## 1st Qu.:1.0000   
## Median :1.0000   
## Mean :0.8908   
## 3rd Qu.:1.0000   
## Max. :1.0000   
##   
## Previous.qualification  
## Secondary education :3717   
## Technological specialization course : 219   
## Basic education 3rd cycle (9th/10th/11th year) or equiv.: 162   
## Higher education - degree : 126   
## Other - 11th year of schooling : 45   
## Higher education - degree (1st cycle) : 40   
## (Other) : 115   
## Previous.qualification..grade. Nacionality Mother.s.qualification  
## Min. : 95.0 Portuguese :4314 Min. : 1.00   
## 1st Qu.:125.0 Brazilian : 38 1st Qu.: 2.00   
## Median :133.1 Santomean : 14 Median :19.00   
## Mean :132.6 Spanish : 13 Mean :19.56   
## 3rd Qu.:140.0 Cape Verdean: 13 3rd Qu.:37.00   
## Max. :190.0 Guinean : 5 Max. :44.00   
## (Other) : 27   
## Father.s.qualification Mother.s.occupation Father.s.occupation Admission.grade  
## Min. : 1.00 Min. : 0.00 Min. : 0.00 Min. : 95.0   
## 1st Qu.: 3.00 1st Qu.: 4.00 1st Qu.: 4.00 1st Qu.:117.9   
## Median :19.00 Median : 5.00 Median : 7.00 Median :126.1   
## Mean :22.28 Mean : 10.96 Mean : 11.03 Mean :127.0   
## 3rd Qu.:37.00 3rd Qu.: 9.00 3rd Qu.: 9.00 3rd Qu.:134.8   
## Max. :44.00 Max. :194.00 Max. :195.00 Max. :190.0   
##   
## Displaced Educational.special.needs Debtor   
## Min. :0.0000 Min. :0.00000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000   
## Median :1.0000 Median :0.00000 Median :0.0000   
## Mean :0.5484 Mean :0.01153 Mean :0.1137   
## 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.00000 Max. :1.0000   
##   
## Tuition.fees.up.to.date Gender Scholarship.holder Age.at.enrollment  
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :17.00   
## 1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:19.00   
## Median :1.0000 Median :0.0000 Median :0.0000 Median :20.00   
## Mean :0.8807 Mean :0.3517 Mean :0.2484 Mean :23.27   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:25.00   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :70.00   
##   
## International Curricular.units.1st.sem..credited.  
## Min. :0.00000 Min. : 0.00   
## 1st Qu.:0.00000 1st Qu.: 0.00   
## Median :0.00000 Median : 0.00   
## Mean :0.02486 Mean : 0.71   
## 3rd Qu.:0.00000 3rd Qu.: 0.00   
## Max. :1.00000 Max. :20.00   
##   
## Curricular.units.1st.sem..enrolled. Curricular.units.1st.sem..evaluations.  
## Min. : 0.000 Min. : 0.000   
## 1st Qu.: 5.000 1st Qu.: 6.000   
## Median : 6.000 Median : 8.000   
## Mean : 6.271 Mean : 8.299   
## 3rd Qu.: 7.000 3rd Qu.:10.000   
## Max. :26.000 Max. :45.000   
##   
## Curricular.units.1st.sem..approved. Curricular.units.1st.sem..grade.  
## Min. : 0.000 Min. : 0.00   
## 1st Qu.: 3.000 1st Qu.:11.00   
## Median : 5.000 Median :12.29   
## Mean : 4.707 Mean :10.64   
## 3rd Qu.: 6.000 3rd Qu.:13.40   
## Max. :26.000 Max. :18.88   
##   
## Curricular.units.1st.sem..without.evaluations.  
## Min. : 0.0000   
## 1st Qu.: 0.0000   
## Median : 0.0000   
## Mean : 0.1377   
## 3rd Qu.: 0.0000   
## Max. :12.0000   
##   
## Curricular.units.2nd.sem..credited. Curricular.units.2nd.sem..enrolled.  
## Min. : 0.0000 Min. : 0.000   
## 1st Qu.: 0.0000 1st Qu.: 5.000   
## Median : 0.0000 Median : 6.000   
## Mean : 0.5418 Mean : 6.232   
## 3rd Qu.: 0.0000 3rd Qu.: 7.000   
## Max. :19.0000 Max. :23.000   
##   
## Curricular.units.2nd.sem..evaluations. Curricular.units.2nd.sem..approved.  
## Min. : 0.000 Min. : 0.000   
## 1st Qu.: 6.000 1st Qu.: 2.000   
## Median : 8.000 Median : 5.000   
## Mean : 8.063 Mean : 4.436   
## 3rd Qu.:10.000 3rd Qu.: 6.000   
## Max. :33.000 Max. :20.000   
##   
## Curricular.units.2nd.sem..grade.  
## Min. : 0.00   
## 1st Qu.:10.75   
## Median :12.20   
## Mean :10.23   
## 3rd Qu.:13.33   
## Max. :18.57   
##   
## Curricular.units.2nd.sem..without.evaluations. Unemployment.rate  
## Min. : 0.0000 Min. : 7.60   
## 1st Qu.: 0.0000 1st Qu.: 9.40   
## Median : 0.0000 Median :11.10   
## Mean : 0.1503 Mean :11.57   
## 3rd Qu.: 0.0000 3rd Qu.:13.90   
## Max. :12.0000 Max. :16.20   
##   
## Inflation.rate GDP Target   
## Min. :-0.800 Min. :-4.060000 Length:4424   
## 1st Qu.: 0.300 1st Qu.:-1.700000 Class :character   
## Median : 1.400 Median : 0.320000 Mode :character   
## Mean : 1.228 Mean : 0.001969   
## 3rd Qu.: 2.600 3rd Qu.: 1.790000   
## Max. : 3.700 Max. : 3.510000   
##

#### Continuous Attributes

These attributes are continuous variables: [“Unemployment.rate”, “Inflation.rate”, “GDP”]

#### The “XYZ” Attribute

### Exploratory Data Analysis

### Data Transformation

#### Step 1

#### Step 2

#### Step 3

## Models

### Model 1

#### Step 1

#### Step 2

#### Step 3

# Results

## Attribute Analysis

#### the “ABC” attribute

#### the “XYZ” attribute

## Model Results: Model 1

## Model Results: Model 2

# Conclusions

# Limitations of the Analysis and/or Available Data

# Bibliography

Realinho,Valentim, Vieira Martins,Mónica, Machado,Jorge, and Baptista,Luís. (2021). Predict Students’ Dropout and Academic Success. UCI Machine Learning Repository. <https://doi.org/10.24432/C5MC89>.