STUDENT ACADEMIC ANALYSIS

Machine Learning I

ABSTRACT

The student academic analysis delves into a comprehensive assessment of academic performance. Through a systematic examination of these elements, this analysis aims to uncover meaningful patterns and correlations that can inform targeted interventions to ensure minimal dropouts while improving overall student progress leading to graduation.

Medha Gurvishkumar Trivedi

Ruby Ruby

Simranjit Simranjit

INTRODUCTION

This dataset provides a comprehensive view of students enrolled in various undergraduate degrees offered at a higher education institution. It contains information on demographics, social economics, and academic achievement that can be used to examine potential indicators of student dropout and academic success. This dataset consists of several disconnected databases that each contain pertinent data that was available at the time of enrolment, including the application manner, marital status, the course preference, and more. Additionally, by examining the number of curricular units credited/enrolled/evaluated/approved as well as their individual grades, this data can be utilized to estimate total student achievement at the conclusion of each semester. Finally, the region's unemployment rate, inflation rate, and GDP can assist us better understand how economic issues affect academic performance or student dropout rates. This potent analysis tool will give important insight into what drives students to continue their education or drop out for a variety of fields like agronomy, design, education, and nursing. management of the media, social services, or technologies

ABSTRACT

This dataset can be used to understand and predict student dropouts and academic outcomes. The data includes a variety of demographic, social-economic and academic performance factors related to the students enrolled in higher education institutions. The dataset provides valuable insights into the factors that affect student success and could be used to guide interventions and policies related to student retention.

ROWS AND ATTRIBUTE INFORMATION

1. There are 18 categorical attributes in the dataset.
2. Attributes There are 8 numerical attributes.
3. Overall, dataset consist of total 26 attributes
4. There are 4424 instances in Student Academic Analysis

CATEGORICAL ATTRIBUTES DESCRIPTION

Sr. No	Attribute	Description	Categories
1	Marital status	The marital status of the student.	1. Single 2. Married 3. Widower 4. Divorced 5. Facto Union 6. Legally Separated
2	Application Mode	The method of application used by the student	1. 1st phase General Contingent 2. Ordinance No 612/93 3. 1st Phase Special Contingent (Azores Island) 4. Holders of other higher courses 5. Ordinance No. 854-B/99

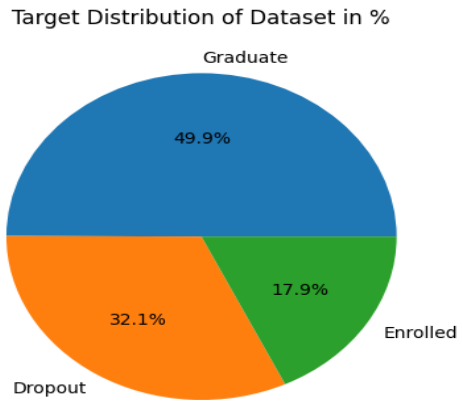
3	Application order	The order in which the student applied	0 (first choice) to 9 (last choice)
4	Course	The course taken by the student.	1. Biofuel Production Technologies 2. Animation and Multimedia Design 3. Social Service (evening attendance) 4. Agronomy 5. Communication Design
5	Daytime/evening attendance	Whether the student attends classes during the day or in the evening.	0 - Daytime 1 - Evening
6	Previous qualification	The qualification obtained by the student before enrolling in higher education.	1 Secondary education 2 Higher education - bachelor's degree 3 Higher education – degree
7	Nationality	The nationality of the student	1 - Portuguese 2 - German 3 - Spanish 4 - Italian 5 - Dutch
8	Mother's qualification	The qualification of the student's mother	1 - Secondary Education - 12th Year of Schooling 2 - Higher Education - Bachelor's Degree 3 - Higher Education - Degree 4 - Higher Education - Master's 5 - Higher Education - Doctorate
9	Father's qualification	The qualification of the student's father.	1 - Secondary Education - 12th Year of Schooling 2 - Higher Education - Bachelor's Degree 3 - Higher Education - Degree 4 - Higher Education - Master's 5 - Higher Education - Doctorate
10	Mother's occupation	The occupation of the student's mother.	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
11	Father's occupation	The occupation of the student's father.	0 - Student 1 - Representatives of the Legislative Power and Executive Bodies, Directors, Directors and Executive Managers

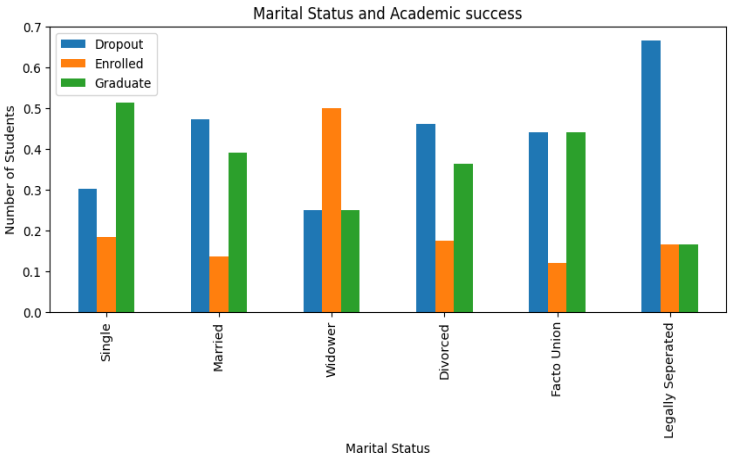
			2 - Specialists in Intellectual and Scientific Activities 3 - Intermediate Level Technicians and Professions 4 - Administrative staff
12	Displaced	Whether the student is a displaced person	0- No 1- Yes
13	Educational special needs	Whether the student has any special educational needs.	0 - No 1 - Yes
14	Tuition fees up to date	Whether the student's tuition fees are up to date	0 - No 1 - Yes
15	Gender	The gender of the student.	0 - Male 1 - Female
16	Scholarship holder	Whether the student is a scholarship holder.	0 - No 1 - Yes
17	International	Whether the student is an international student	0 - No 1 - Yes
18	Target	Status of student	1 - Enrolled 2 - Graduate 3 - Dropout

NUMERICAL ATTRIBUTES DESCRIPTION

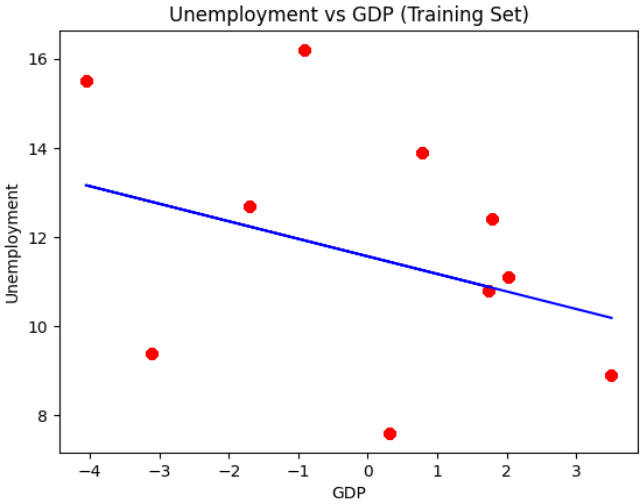
Sr. No	Attribute	Description
1	Age at enrolment	The age of the student at the time of enrolment
2	Curricular units 1st sem (credited)	The number of curricular units credited by the student in the first semester.
3	Curricular units 1st sem (enrolled)	The number of curricular units enrolled by the student in the first semester.
4	Curricular units 1st sem (evaluations)	The number of curricular units evaluated by the student in the first semester
5	Curricular units 1st sem (approved)	The number of curricular units approved by the student in the first semester.
6	Unemployment rate	The number of unemployment rate
7	Inflation rate	The number of Inflation rate

EXPLORATORY DATA ANALYSIS



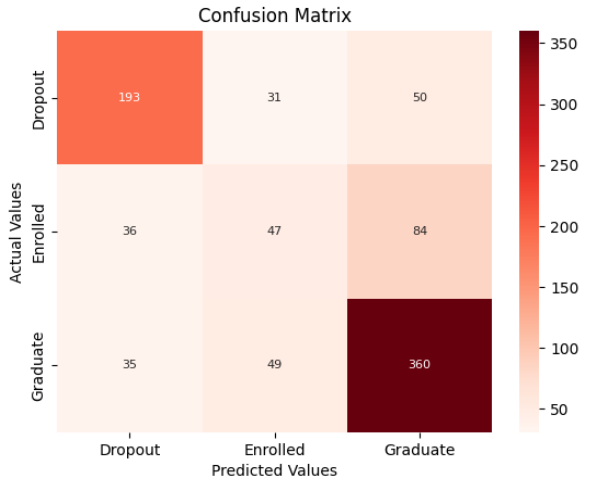


LINEAR REGRESSION: UNEMPLOYMENT & GDP



We aim to ascertain if unemployment and GDP play a role in influencing student dropouts. Linear regression presupposes that a linear connection exists between the independent variables (inputs) and the dependent variable (output). This implies that alterations in the independent variables correspond directly with changes in the dependent variable.

CONFUSION MATRIX



A confusion matrix offers a comprehensive and detailed overview of a classifier's predictions by contrasting them with the actual true labels in a dataset. We construct this matrix to gain insights into outcomes related to student outcomes, such as graduation and dropout situations. This approach provides us with a vivid portrayal of potential future predictions. This information becomes instrumental in making informed decisions that positively impact students' academic journeys, fostering their commitment to successfully complete their educational pursuits.

ANALYSIS

According to the pie chart, it is evident that the highest proportion belongs to graduates. The bar chart illustrates that unmarried students have a higher tendency to complete their education. In contrast, widowed individuals show more interest in pursuing further studies, leading to a significant enrollment rate. On the other hand, students who are legally separated are more prone to discontinuing their education.

According to linear regression analysis, there is a positive correlation between unemployment and GDP, meaning that when unemployment rises significantly, there is a notable decline in GDP.

The confusion matrix clearly highlights that the number of correctly predicted graduates is significant. However, we must not overlook the count of dropouts, as they remain at a substantial risk of escalation.

CONCLUSION

Customizing the curriculum, program structure, and fees based on marital status or adjusting enrollment strategies could enhance the likelihood of maximizing student graduation rates and minimizing dropouts.

Furthermore, promoting and motivating students to successfully complete their academic journey would be a prudent approach.

It is important to consider both unemployment rates and GDP as potential influences on student dropout rates when creating a program plan.

The confusion matrix highlights the importance of treating dropout predictions as true positives, emphasizing the need to address this issue seriously to ensure student's academic advancement.

REFERENCE

Dataset: Predict students' dropout and academic success | Kaggle