

SEGUNDA ENTREGA DE PROYECTO

POR
GERALDINE MICHELLE OTERO CRIOLLO
KEVIN STEVEN HERNÁNDEZ LÓPEZ

MATERIA
INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

PROFESOR
RAÚL RAMOS POLLÁN

UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERÍA
2023

MODELO DE PREDICCIÓN DE LA DESERCIÓN Y ÉXITO ACADÉMICO DE LOS ESTUDIANTES DE LA UNIVERSIDAD

Para la realización del proyecto se utilizó una base de datos de kaggle donde describe y da una visión integral de los estudiantes que están matriculados en un título universitario, con esta base de datos tenemos variables categóricas como:

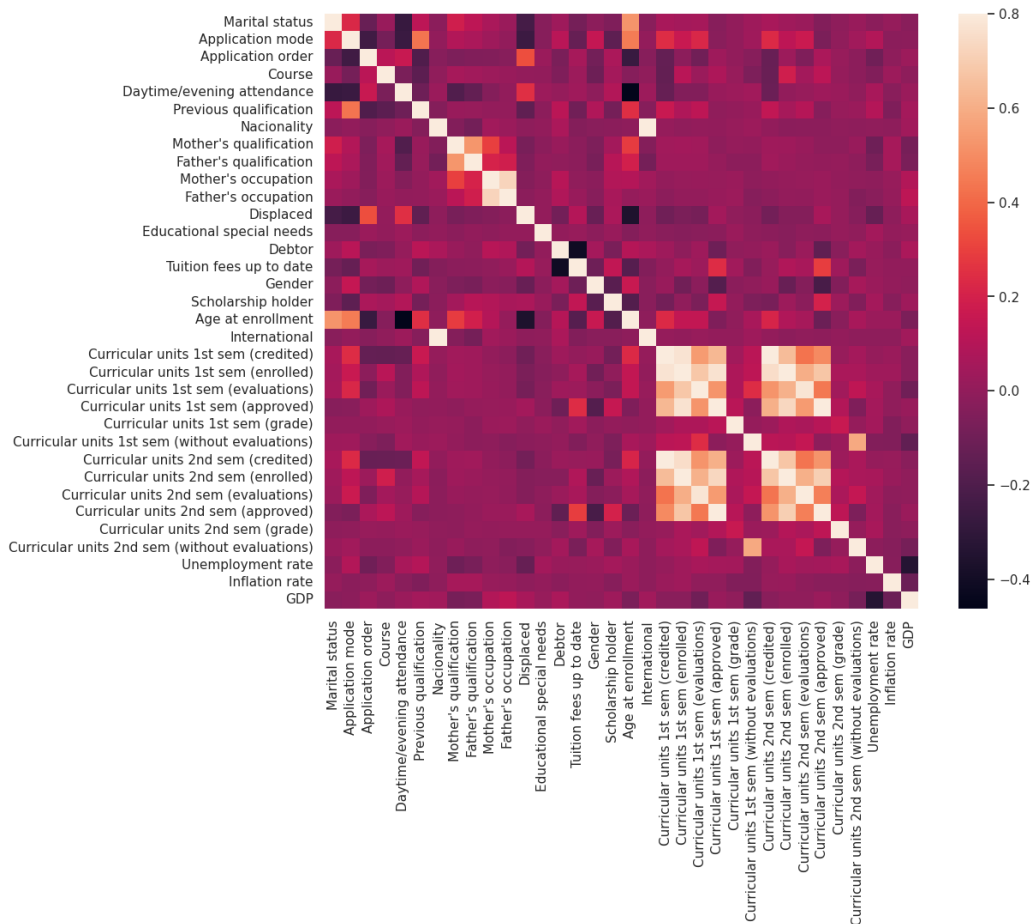
- Marital status: The marital status of the student.
- Application mode: The method of application used by the student.
- Course: The course taken by the student.
- Daytime/evening attendance: Whether the student attends classes during the day or in the evening.
- Previous qualification: The qualification obtained by the student before enrolling in Higher education.
- Nationality: The nationality of the student.
- Mother's qualification: The qualification of the student's mother.
- Father's qualification: The 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.
- Displaced: Whether the student is a displaced person.
- Educational special needs: Whether the student has any special educational needs.
- Debtor: Whether the student is a debtor.
- Tuition fees up to date: Whether the student's tuition fees are up to date.
- Gender: The gender of the student.
- Scholarship holder: Whether the student is a scholarship holder.
- International: Whether the student is an international student. (Categorical)

y unas variables numéricas como

- Application order: The order in which the student applied.
- Age at enrollment: The age of the student at the time of enrollment. (Numerical)
- Curricular units 1st sem (credited): The number of curricular units credited by the student in the first semester.
- Curricular units 1st sem (enrolled): The number of curricular units enrolled by the student in the first semester.
- Curricular units 1st sem (evaluations): The number of curricular units evaluated by the student in the first semester.
- Curricular units 1st sem (approved): The number of curricular units approved by the student in the first semester

Inicialmente se intenta entender la base de datos con la que se trabajará, teniendo en cuenta que las variables categóricas se les da un valor a cada una de sus categorías para facilitar la manipulación y la comprensión de estas por ejemplo: el estado civil del estudiante tiene 6 categorías siendo 1 soltero o único, 2 casado, 3 viudo, 4 divorciado y así sucesivamente.

Se realiza un diagrama de correlación que nos da un avistamiento superficial de cuales variables pueden tener a llegar una correlación alta para ayudar a entender el comportamiento de la variables, con lo cual se notan unas bastante significativas sobre todo con las variables numéricas, como también se nota una correlación alta entre la ocupación de la madre y el padre. También se obtiene la información de la base de datos como la media de las variables, su mínimo y su máximo.



Por otro lado obtuvimos las columnas en las cuales hacen falta datos, las cuales fueron:

- Age at enrollment 218
- Curricular units 1st sem (grade) 219
- Curricular units 2nd sem (grade) 217

Se realizó un gráfica en la que se relaciono los datos faltantes y se nota que están de forma aleatoria.

Una de las problemáticas que hemos tenido en lo que vamos realizando el trabajo es la forma de relacionar las variables numéricas con las categóricas como la variable objetivo. Otra problemática es el arreglo de los datos faltantes ya que al momento de llamar la librería (`from local.lib import mlutils`) presenta un error en la ejecución del código.