

PRIMERA ENTREGA DE PROYECTO

POR  
GERALDINE MICHELLE OTERO CRIOLLO  
KEVIN STEVEN HERNÁNDEZ LÓPEZ  
SANTIAGO HERNAN MAFLA LONDOÑO

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

## 1. DESCRIPCIÓN DEL PROBLEMA PREDICTIVO A RESOLVER.

En el proyecto que se va a realizar para dar resolución a un problema seleccionamos la predicción de la deserción y éxito académico de los estudiantes de una universidad. El cual tiene múltiples variables y parámetros como lo vienen siendo factores socioeconómicos, relaciones interpersonales, cursos matriculados, cantidad de cursos, entre otros. Todo esto con el fin de dar con un resultado que nos provea con suficiente información como para tomar acciones y decisiones importantes sobre el problema de estudio.

## 2. DATASET UTILIZADO.

Para realizar este modelo usamos un dataset de Kaggle de los estudiantes que tiene 35 columnas, el cual tenía como principal objetivo revisar el impacto económico y social de los estudiantes en su desempeño.

-> [Predict students' dropout and academic success | Kaggle](https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention?resource=download)  
<https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention?resource=download>

Este archivo contiene datos de una institución de educación superior con diversas variables relacionadas con los estudiantes de pregrado, incluidos datos demográficos, factores socioeconómicos y rendimiento académico, para investigar el impacto de estos factores en la deserción de los estudiantes y el éxito académico.

- Marital status: The marital status of the student.
- Application mode: The method of application used by the student.
- Application order: The order in which the student applied.
- 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.
- Age at enrollment: The age of the student at the time of enrollment. (Numerical)
- International: Whether the student is an international student. (Categorical)

- 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.

### 3. MÉTRICA DE DESEMPEÑO REQUERIDA.

La métrica de desempeño a usar para el proyecto será específicamente el log loss (Multiclass Logarithmic Loss), debido a que nos provee con una mayor precisión en cuanto al indicador de confianza de los resultados, permitiendo así tener una mejor y más exacta visión de la predicción hecha.

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}),$$

### 4. CRITERIO DEL DESEMPEÑO DESEABLE EN PRODUCCIÓN.

Según el estudio realizado y los datos recolectados se espera que la cantidad de estudiantes graduados sea considerablemente mayor a la cantidad de estudiantes desertores, preferiblemente en una escala  $\frac{1}{3}$ , en el cual por cada estudiante desertor existan 3 estudiantes graduados.

### 5. BIBLIOGRAFÍA

<https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention?resource=download>