**1. Introducción**

Este proyecto se basa en un conjunto de datos provenientes de una institución de educación superior que incluye información sobre los estudiantes matriculados en diferentes carreras de pregrado. El objetivo es predecir la deserción escolar y el éxito académico de los estudiantes utilizando datos recopilados en el momento de la inscripción, como su historial académico, factores demográficos y socioeconómicos, así como su desempeño académico durante los dos primeros semestres. El problema se formula como una tarea de clasificación en tres categorías, con un fuerte desequilibrio en una de las clases.

**2. Objetivo General**

Identificar la relación que existe entre diferentes factores y la deserción escolar

**3. Objetivos Específicos**

* Limpiar y procesar los datos proporcionados sobre los estudiantes (perfil académico, demográfico y socioeconómico).
* Convertir las variables numéricas en su respectivo significado
* Determinar la relación entre diversos factores y la deserción escolar.
* Construir un modelo de clasificación para predecir la probabilidad de deserción, continuación o éxito académico.

**4. Planteamiento del Problema**

La deserción escolar es un problema crítico que enfrentan las instituciones de educación superior. Factores como el entorno socioeconómico, el rendimiento académico temprano y las características demográficas pueden influir significativamente en la probabilidad de que un estudiante abandone su programa de estudios. Este proyecto tiene como objetivo construir un modelo predictivo que identifique a los estudiantes en riesgo de deserción con base en los datos disponibles al momento de la inscripción y su rendimiento académico en los primeros dos semestres.

**5. Posibles Preguntas de Investigación**

* ¿Existen diferencias en la tasa de deserción entre las distintas carreras?
* ¿Existen diferencias en las tasas de deserción entre estudiantes desplazados y no desplazados?
* ¿Existe una correlación significativa entre el nivel educativo de los padres y la probabilidad de deserción de los estudiantes?
* ¿Hay diferencias notables en las tasas de deserción entre estudiantes de diferentes géneros?
* ¿Hasta qué punto podemos establecer una relación entre el número de deudas financieras de los estudiantes y la tasa de deserción en aquellos que las tienen?

**6. Posible Solución**

La solución consiste en implementar un proceso de limpieza y procesamiento de datos utilizando Python y pandas. Se abordarán tareas como la eliminación de valores nulos, la transformación de datos numéricos en sus respectivos significados dentro de las variables y el análisis preliminar de la relación entre variables clave. El procesamiento de estos datos permitirá que los mismos estén listos para futuras etapas de análisis más detalladas.

**7. Consideraciones Futuras**

Una vez que los datos hayan sido procesados y limpiados, será posible realizar un análisis más profundo, que nos permita desarrollar un modelo de clasificación para predecir la probabilidad de deserción, continuación o éxito académico. El análisis de los factores que más influyen en la deserción escolar, nos permitirá diseñar intervenciones tempranas que podrían mitigar este problema.

**8. Conclusión**

Este proyecto busca aplicar técnicas de ciencia de datos para predecir la deserción y el éxito académico de estudiantes universitarios, utilizando datos sobre factores socioeconómicos y académicos tempranos. La construcción de un modelo predictivo efectivo permitirá a las instituciones educativas intervenir de manera temprana y desarrollar estrategias para reducir la deserción, optimizando así el éxito académico de sus estudiantes.

**9. Entregables**

* Código ejecutable en Python (Google Colab o Jupyter Notebook) que realice el procesamiento y modelado de los datos.
* Documento técnico detallando los pasos seguidos en el procesamiento y análisis de los datos.
* Presentación con los resultados del análisis y las recomendaciones basadas en los hallazgos del modelo predictivo.