

Fecha: 11 de noviembre de 2022

Lineamientos Generales - Proyecto Final del Diplomado en Introducción a la Ciencia de Datos con Python

Este documento contiene la información general para la elaboración del proyecto final de este diplomado, el cual deberá entregarse hasta el 30/12/2022 a través de la plataforma empleada en el curso (<https://ead.pol.una.py/politecnica/course/view.php?id=7121>). El objetivo principal de este proyecto es la aplicación de los conocimientos adquiridos a lo largo de los cuatro módulos del diplomado.

Los puntos principales son los siguientes:

- A. Los equipos pueden tener hasta 2 integrantes.
- B. Cada equipo deberá completar un formulario (hasta el domingo 27/11/2022) con los siguientes datos:
 - a. Área de interés del proyecto final.
 - b. Integrantes del equipo.
 - c. Disponibilidad previa de un dataset de interés para el equipo. El mismo debe ser de acceso público o tener la autorización de la fuente para su uso. Se propone el siguiente buscador de Google para la búsqueda de datasets de interés: <https://datasetsearch.research.google.com/>. También podría buscarse en Kaggle: <https://www.kaggle.com/datasets>.
 - d. El dataset seleccionado deberá tener al menos 1000 registros y 3 variables.
- C. En relación al punto anterior, en el caso de que los participantes no cuenten con un dataset previamente, se determinará el mismo con los profesores del curso, con base en los intereses expuestos por los participantes.
- D. Se deberá presentar un documento del anteproyecto con: el título del proyecto, objetivos, datos principales, metodología, etc. Este anteproyecto no debe contener más de dos páginas. La fecha límite de presentación es el 02/12/2022.
- E. Se tendrá al menos una reunión intermedia (en modalidad virtual) entre los integrantes del equipo y un profesor del diplomado, para revisar los avances y responder consultas.
- F. Para la fecha de culminación del proyecto se deberá entregar un reporte que contendrá, como mínimo, la siguiente información:
 - a. Breve introducción al problema que quiere ser resuelto o información que quiera determinarse. Plantear un conjunto de preguntas que relacionen las variables.
 - b. Descripción del dataset utilizado: variables, tipos de datos, número de registros, existencia de datos faltantes.
 - c. Tipo de problema planteado: regresión, clasificación o agrupamiento.
 - d. Metodología empleada para resolver el problema.
 - e. Métricas de desempeño utilizadas. Uso (o no) de aprendizaje supervisado. Uso (o no) de validación cruzada.
 - f. Descripción de los resultados obtenidos. Identificación de sobre ajuste en dichos resultados. Respuesta a las preguntas planteadas en la identificación del problema.

- G. En relación al punto anterior, también deberá entregarse un video explicativo (de duración de entre 5 a 10 minutos) sobre los puntos más relevantes del proyecto realizado.