 <b>FACULTAD DE INGENIERÍAS</b>	<b>SISTEMAS INTELIGENTES 2</b>  <b>Proyecto Final</b>	<b>FECHA ELABORACIÓN DEL DOCUMENTO:</b> 1/MAY/2023
		<b>VERSIÓN:</b> 1


Se le ha encomendado una de las dependencias de la Universidad a la cual le debe desarrollar un sistema informático el cual permita aplicar algoritmos de Machine learning y Deep learning.

Para esto se requiere cumplir con los siguientes requerimientos funcionales

- El sistema deberá permitir **cargar documentos de Excel desde un end point**
- Los archivos cargados deben ser almacenados en Mongo DB (Se recomienda utilizar Mongo Atlas) en formato JSON
- Una vez se cargue la información en la plataforma el sistema debe ofrecer los **siguientes endpoints para preprocesar el archivo**
  - Identificar los tipos de datos en las columnas (Texto, Numérico, Booleano)
  - Aplicar técnicas para tratamiento de datos faltantes
    - Descarte: Eliminar registros que contienen datos faltantes
      - Imputación por media
- La plataforma **debe poseer un endpoint que permita obtener las imágenes** de los histogramas de las **variables numéricas** y las matrices de correlación. La respuesta de este endpoint serán las rutas donde se encuentran alojadas las imágenes en el servidor

**Nota:** Luego de llamar el método preprocesar archivo, se debe crear un nuevo archivo en mongo.

- El sistema debe poseer un endpoint de “entrenamiento” que permita aplicar algoritmos de machine learning sobre un dataset que ya está preprocesado y cargado en la plataforma. Los algoritmos que como mínimo deben poder elegirse en el sistema son
  - Regresión Logística
  - KNN
  - Máquinas de soporte vectorial
  - Naive Bayes
  - Árboles de decisión
  - Regresión Lineal

 <b>FACULTAD DE INGENIERÍAS</b>	<b>SISTEMAS INTELIGENTES 2</b>  <b>Proyecto Final</b>	<b>FECHA ELABORACIÓN DEL DOCUMENTO:</b> 1/MAY/2023
		<b>VERSIÓN:</b> 1

**Nota:** El único algoritmo que servirá para problemas de regresión será la regresión lineal, de resto todos serán para problemas de clasificación

- Dentro del endpoint de “entrenamiento” de permitir la elección de los algoritmos
  - Hold Out (definiendo el porcentaje de pruebas) o Cross Validation (definiendo el número de folds)
  - Normalización: MinMax o Standar Scaler
- Una vez se entrenen los modelos se deben de almacenar en el sistema, posterior a esto se debe ofrecer un end point que permita listar **todos los modelos que están entrenados** para un conjunto de datos en particular y **el top 3** de los mejores
- Debe existir un endpoint que **permita visualizar las métricas que obtuvieron** los modelos en la fase de pruebas
- Debe existir un endpoint el cual permita realizar predicciones de los datos, este recibe los datos de entrada que requiere el modelo, el identificador del modelo que se quiere utilizar para la predicción y debe responder con la inferencia correspondiente.
- Se debe explorar una herramienta de visualización de datos de libre elección que permita visualizar 3 gráficos que sean de interés general para la dependencia. Debe por lo menos existir una gráfica de barras, circular (torta) y líneas. Estas graficas deben poder ser exportadas en un iframe de html el cual debe ser incrustado en un archivo html sencillo para visualización de los usuarios finales. Dentro de las herramientas que se pueden explorar son:
  - Power BI
  - Tableau
  - Chartblocks
- Cada uno de los endpoints deben de estar documentados con los formatos de peticiones y posibles resultados de error
- Se recomienda utilizar para construir las apis “Flask”