

## SEMANA 3 – GUÍA DE TRABAJO PRÁCTICO

### Objetivos

- ✍ Familiarizarse con la herramienta MLflow para el manejo y administración de los experimentos realizados con diferentes estimadores y parámetros.
- ✍ Entrenar un modelo propio

### Prerrequisitos

- ✍ Conocimientos intermedios de programación en Python y habilidades de uso de notebooks. Esto es, uso de librerías, lectura y seguimiento de código ejemplo, desarrollo de código propio, depuración de programas, generación de resultados.
- ✍ Saber descargar e instalar programas en su máquina local personal.
- ✍ Saber seguir rutas de localización de archivos y carpetas.
- ✍ Saber descargar documentos y subir entregables en la infraestructura de Coursera.

### Metodología

- ✍ Se realiza de manera **Grupal**.
- ✍ Deben utilizarse la infraestructura, datos y *notebooks* entregados para el desarrollo de los retos propuestos.
- ✍ Las instrucciones de cada reto se encuentran descritas en el enunciado a continuación.
- ✍ No se reciben trabajos desarrollados en otras infraestructuras ni con otros conjuntos de datos.
- ✍ Se entrega en la infraestructura de Coursera, en el enlace previsto para tal fin.

### Enunciado

Los trabajos prácticos del curso se realizan en una infraestructura virtual, que debe ser instalada en su máquina personal. Esto le permite trabajar en un entorno computacional que ya está configurado, de manera que su esfuerzo y tiempo se concentren en el desarrollo de los retos. A continuación, encuentra la descripción de cada una de las actividades que debe lograr.

**IMPORTANTE:** para esta semana usted cuenta con datos procesados como ejemplo para seguir la guía de esta semana, sin embargo, es recomendable que utilice sus datos procesados de la semana 2 ya que estos datos a manera de ejemplo no tienen una buena limpieza y es parte de su trabajo haberlo realizado durante la semana 2.

#### [1] Inicio de actividad

- ✍ Para esta actividad solo necesita en la carpeta s3 de su instancia.
- ✍ Para correr el contenedor corra el siguiente comando: “- docker-compose up” En la terminal encontrará una URL de este estilo: “<http://<ip-de-su-maquina>:8888/lab?token=0f5dabdcdf9fd9941ade8ce267c9a61327c0ed4db82c1a0c>” Esta URL le da acceso al libro de jupyter para que pueda realizar la actividad. En caso de dudas referenciar la guía en la sección “guías de Tecnología”.

## [2] ETL

¿Qué es un MLFLOW?

MLflow es una plataforma open source que nos ayuda a manejar el ciclo de vida de los modelos en inteligencia artificial (ML lifecycle); para esto contamos con experimentos, reproducibilidad, despliegues y un registro central de los modelos entrenados que podemos comparar.

Recordemos que el ciclo de vida de los modelos en inteligencia artificial está compuesto en cuatro pasos.

- Preparación de los datos
- Construcción del modelo
- Análisis del modelo
- Despliegue del modelo

En esta actividad **debe utilizar los datos procesados en la actividad ETL de la semana 2**, con estos datos deberá entrenar su modelo y generar el informe con el uso de la herramienta de MLFlow. El contenedor Docker cuenta con unos datos que ya se encuentran mapeados los puntos más cercanos (la estación de clima y el punto de interés) a la estación de bicicleta inicial, los cuales se utilizan como ejemplo para el uso de la herramienta, sin embargo, están lejos de ser óptimos o estar completamente limpios.

Usted debe entrenar su propio modelo y realizar un informe del proceso con MLFlow. si desea trabajar con otras librerías para el entrenamiento como *Tensorflow* o *Dask* puede hacerlo con total libertad.

Es **IMPORTANTE** que en el informe compare el resultado de los modelos entrenados, justifique cuál fue el mejor resultado y por qué.

En caso de tener dudas sobre los ejemplos, los *notebooks* o la manera de subir sus resultados a Coursera, utilice el foro del curso previsto en Slack.



Ubicación del notebook de trabajo: /home/ivanarturo9620/s3

Manera de nombrar los archivos de resultados: usuario correo uniandes. por ejemplo, "jp.gonzales10\_informe\_s2.pdf" o jp.gonzales10\_informe\_s2.zip"

## Entrega

Realice la entrega del informe en PDF utilizando el enlace previsto en Coursera y entregue el jupyter notebook con sus procedimientos en un archivo zip.

**Fecha y hora límite de entrega:** **domingo** de la **semana 3** del curso a las **10 PM hora COLOMBIA**

-  Es la única forma válida de entrega.
-  Asegúrese de entregar el documento correspondiente.

👉 No se reciben entregas tardías o incompletas a excepción de tener una excusa válida.