

ASOCIACIÓN INDUSTRIAL PARA EL IMPULSO DE LA ECONOMÍA DEL
DATO Y DE LA INTELIGENCIA ARTIFICIAL – INDESIA

Guía rápida de uso de la plataforma Datatón Indesia

01 | **Acceso a la plataforma**

02 | **Directrices previas**

03 | **Databricks**

04 | **Azure Machine Learning**

05 | **Servicios cognitivos**

06 | **Recursos compartidos**

07 | **Configuración de GIT**

01

Acceso a la plataforma

Plataforma Azure

- A. Acceso Azure:** Para poder visualizar los recursos proporcionados, deberemos dirigirnos al portal de Azure, en nuestro perfil haremos clic en la opción “Switch directory” y habrá que elegir el directorio de Indesia tal y como se muestra en las imágenes.



The screenshot shows the Azure portal user interface. At the top, a blue header bar displays the user's email 'marta.valentina@indesi...' and the organization 'ASOCIACIÓN INDUSTRIAL PARA ...'. Below this, a user profile card for 'marta.valentina@i...' is visible, with links for 'View account' and 'Switch directory'. A blue arrow points from the 'Switch directory' link to a table of directories. The table has columns for 'Directory name', 'Domain', and 'Directory ID'. It lists one directory: 'Asociación industrial para el im...' with domain 'indesia.org' and ID '3886555b-d6c7-4f51-8a4b-22...'. The 'Current' status is indicated by a green checkmark.

Directory name ↑↓	Domain ↑↓	Directory ID ↑↓
☆ Asociación industrial para el im...	indesia.org	3886555b-d6c7-4f51-8a4b-22...

Endpoints a los workspaces

- A. Databricks:** [https://databricks-\[nombre_equipo\].pro.datatonindesia.es/redirect](https://databricks-[nombre_equipo].pro.datatonindesia.es/redirect)
- B. Azure Machine Learning:** <https://ml.azure.com/>

02 | Directrices previas

Entrega de código

A. Todo el código deberá quedar subido al repositorio dentro de databricks para que pueda ser evaluado.

Recursos compartidos

- A. Endpoints.** Los endpoints a los recursos comunes (OpenAI y AI Search) se proporcionarán en un documento vía Correo.
- B. Recursos disponibles.** Recursos de OpenAI disponibles tendremos gpt-4-turbo, gpt-4-vision, gpt-4-o, gpt-35-turbo-instruct, gpt-35-turbo-16k, gpt-35-turbo y text-embedding-3-large. Además, también se cuenta con gpt-4-32k, gpt-4 bajo demanda.
- C. Recursos adicionales.** Para el uso de GPT-4 o cualquier otro requisito necesario para vuestro proyecto, por favor comentárselo al equipo de Indesia.

03 | Databricks

Ingesta de datos

- A. La subida de los datasets**, se hará directamente en databricks. En el volumen creado dentro del catálogo (identificado con el nombre del equipo) bajo el esquema Raw

Catálogo

Teclee para filtrar

> hive_metastore

> main

> prtepr

> information_schema

> raw

prtepr

> samples

> system

Catálogos > prtepr > raw >

prtepr.raw.prtepr

Propietario: a945ef16-8c48-43c5-b4f2-9d36f479a0be

Comentario: Añadir comentario

Explorar Detalles Permisos

/Volumes/prtepr/raw/prtepr

Nombre

Tamaño

Último cambio

No hay contenido en el volumen

03 | Databricks

Recursos de databricks

A. Librerías precargadas:

"pyod", "joblib", "matplotlib", "numpy", "numba", "scipy", "scikit_learn", "six", "statsmodels", "combo",
"keras", "tensorflow", "torch", "torchvision", "matplotlib", "Suod", "xgboost", "pillow", "hyperopt", "mlflow", "azureml", "azureml-
mlflow", "azureml-core", "azure-ai-ml", "bamboolib", "jinja2==3.0.3", "itsdangerous==2.0.1", "lightgbm", "prophet",
"dask", "plotly", "folium", "pygeos", "rtree", "osmnx", "networkx", "shapely", "tqdm", "geopandas", "geopy", "scikit-learn"

04 | Azure Machine Learning

Ingesta de datos

- A. Opción 1.** Depositar los archivos en el Storage Account denominados con la siguiente nomenclatura "indplatformpaml<nombre-equipo>", directamente en el portal de Azure, dentro del blob [nombre-equipo]
- B. Opción 2.**
- I. Descargar Azure Storage Explorer
 - II. Hacer login con la cuenta proporcionada para el datatón
 - III. Seleccionar la suscripción, una vez hecho eso, se podrá ver el Storage Account "indplatformpamlstesteam1".
 - IV. Dentro del blob [nombre_equipo] Arrastrar los ficheros.

05

Servicios cognitivos

Acceso

- A. **Para poder acceder a los servicios cognitivos**, deberemos dirigirnos al Resource Group de nuestro equipo, `indplatformprg[nombre-equipo]`. Allí nos encontraremos un recurso llamado AI-multiservice de tipo “Azure AI services multi-service account”, el cual podremos ver su endpoint y clave en la sección “Keys and Endpoint”. Una vez los tengamos, ya podremos hacer uso del recurso el cual integra los servicios cognitivos.

Recursos disponibles

- A. Decision, Language, Speech, Vision, Document Intelligence y Metrics Advisor.

06

Recursos compartidos

Azure Open AI

A. Implementaciones. El nombre de las implementaciones es igual a los nombres de los modelos.

Azure AI Search

A. Nomenclatura. Para crear los index, se ruega que por favor se haga uso de esta nomenclatura para duplicados y fallos.

I. Ejemplo:

[nombre-índice]-[Nombre-equipo]
Ejemplo: ejemploíndice-team1

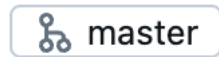
07 | Configuración de GIT

Credenciales

- A. En el documento llamado KEYS Y ENDPOINTS aparecerá el personal access token que se utilizará para la configuración de GIT

Como configurar las credenciales

- A. Presionando el botón que se ve en la imagen dentro del entorno de databricks, accederemos al menú de configuración de GIT.



- B. En el menú de configuración, seleccionaremos como proveedor “Azure DevOps Services (Personal Access Token)”, rellenaremos los campos con nuestro email y el personal access token y presionaremos el botón Save. Una vez hecho eso, ya estaríamos listos para guardar nuestro trabajo realizado en el repositorio que está configurado

Git provider ⓘ

Azure DevOps Services (Personal acces... ▼

Git provider username or email ⓘ

Token

To generate a Azure DevOps personal access token, follow the [Azure DevOps](#)

Token with repo read/write permissions

Save

INDESIA

ASOCIACIÓN INDUSTRIAL PARA EL IMPULSO DE LA
ECONOMÍA DEL DATO Y LA INTELIGENCIA ARTIFICIAL