

# *Tests AI & Data Science*

*For Hybrid Intelligence Spain*



# INTRODUCCIÓN

- El objetivo de estas pruebas es el de evaluar las capacidades, conocimientos y habilidades del candidato/a en materia de inteligencia artificial y ciencia de datos.
- Todas las pruebas deberán ser resueltas utilizando Python como lenguaje de programación, en particular con una [versión estable](#).
- El candidato/a deberá poder garantizar la reproducibilidad del análisis realizado en cada prueba mediante la definición de dependencias a través de fichero *requirements/poetry*.
- El candidato/a deberá seguir las reglas de estilo Python PEP8 así como utiliza analizadores de código para asegurar la calidad de este (código muerto, tipado de funciones, optimización de *imports*, complejidad ciclomática, etc.)
- Los jupyter notebooks deberán estar ordenados y documentados indicando claramente el orden de ejecución y la estructura de estos mediante celdas *markdown* con títulos y subtítulos.
- Se recomienda exportar el jupyter notebook con el resultado embebido a formato html.
- La estructura de los entregables deberá ser la de un repositorio de código tipo cookie cutter con un fichero *readme.md* indicando la función de cada módulo/script/fichero del entregable.
- Se valorará de forma positiva la entrega de una presentación con el planteamiento del caso, resultados y conclusiones obtenidos o un informe en Latex/beamer (en ambos casos solo entregar el pdf generado).
- Se valorará de forma positiva la creación de documentación con Mkdocs.
- Las entregas se realizarán via email mediante un archivo comprimido (uno por cada prueba) cuyo nombre siga la nomenclatura: **XYZ-TestDataScience-W.zip**, donde XYZ son las iniciales del candidato/a y W es el número de prueba.
- Alternativamente se podrá realizar la entrega mediante un repositorio publico de Github aplicando la misma nomenclatura.



# TEST 1 – DATA SCIENCE - CLASIFICACIÓN

- Escoger un dataset abierto de clasificación que no sea el de Iris (referenciar fuente)
- En un Jupyter notebook realizar análisis exploratorio (EDA), incluyendo limpieza de datos, transformaciones, agregaciones y visualizaciones que se consideren oportunas.
- Seleccionar, entrenar y testear el/los modelo/s que se consideren apropiados.
- Justificar el modelo elegido en función métricas de rendimiento.
- Elaborar conclusiones del ejercicio llevado acabo.
- Preparar entregable con todos los ficheros necesarios para reproducir el análisis y poner en producción el modelo entrenado (integración DevOps).



# TEST 2 – TIME SERIES - REGRESSION

- Escoger un dataset abierto de series temporales multivariante que no sea financiero y donde la variable a predecir no tenga estacionalidad (referenciar fuente).
- Realizar comprobación de que la serie temporal es no- estacional
- Seleccionar, entrenar y testear el/los modelo/s de machine learning/series temporales que permitan:
  - Predecir 100 periodos de tiempos en el futuro con la particularidad que a partir del instante donde se inicia la predicción no se dispone de los valores de las variables regresoras
- Representar gráficamente los resultados y evaluar la bondad del modelo utilizando los KPIs adecuados
- Elaborar conclusiones del ejercicio llevado acabo.
- Preparar entregable con todos los ficheros necesarios para reproducir el análisis y poner en producción el/los modelo/s entrenado/s (integración DevOps).



This presentation contains information that may be privileged or confidential and is the property of the Capgemini Group.

Copyright © 2024 Capgemini. All rights reserved.

## About Capgemini Engineering

Capgemini Engineering combines, under one brand, a unique set of strengths from across the Capgemini Group: the world leading engineering and R&D services of Altran – acquired by Capgemini in 2020 - and Capgemini's digital manufacturing expertise. With broad industry knowledge and cutting-edge technologies in digital and software, Capgemini Engineering supports the convergence of the physical and digital worlds. Combined with the capabilities of the rest of the Group, it helps clients to accelerate their journey towards Intelligent Industry. Capgemini Engineering has more than 52,000 engineer and scientist team members in over 30 countries across sectors including aeronautics, automotive, railways, communications, energy, life sciences, semiconductors, software & internet, space & defence, and consumer products.

Capgemini Engineering is an integral part of the Capgemini Group, a global leader in partnering with companies to transform and manage their business by harnessing the power of technology. The Group is guided every day by its purpose of unleashing human energy through technology for an inclusive and sustainable future. It is a responsible and diverse organization of 270,000 team members in nearly 50 countries. With its strong 50-year heritage and deep industry expertise, Capgemini is trusted by its clients to address the entire breadth of their business needs, from strategy and design to operations, fueled by the fast evolving and innovative world of cloud, data, AI, connectivity, software, digital engineering and platforms. The Group reported in 2020 global revenues of €16 billion.

Get the Future You Want | [www.capgemini.com/capgemini-engineering](https://www.capgemini.com/capgemini-engineering)