

Práctica de programación

Beca Skill Center Robotics, Data & AI (Turtles)

Esta práctica trata de evaluar tus conocimientos de programación, así como tu experiencia y facilidad para afrontar problemas complejos. Está compuesta por dos partes.

Para realizar la práctica usaremos Python, en concreto, se deberá programar en la interfaz web de código abierto Jupyter Notebook.

EJERCICIO Nº 1

Son muchas las ocasiones en las que de manera accidental los pescadores se encuentran tortugas en sus redes o varadas en la costa, lo que supone un peligro para ellas ya que son susceptibles de ser atacadas por depredadores. Acciona, es una empresa comprometida con el medio ambiente y es por ello que en 1997, en nuestra división de Acciona Agua se fundó la **Acciona Turtle Initiative** que ofrece un beneficio a los habitantes de los pueblos más cercanos a nuestras desaladoras a cambio de entregar estas tortugas varadas a la desaladora, para que pueda llevar un registro de estas y devolverla al mar.

Esto nos ayuda a controlar el impacto de nuestras plantas de desalación y a mantener los ecosistemas en los que operamos.

Por tanto, la primera parte de la práctica trata de usar modelos de Machine Learning para predecir el número de tortugas capturadas **por semana y por zona** para los años 2019 y 2020 a partir de los datos disponibles. Se propone un ejemplo de csv para entregar los datos, no obstante, se deberá entregar el notebook como pone en la nota final.

Los archivos disponibles en el Ejercicio1 DataScientist.zip son los siguientes:

E1_Dataset_tortugas.csv - es el conjunto de datos que se utilizara para entrenar su modelo. Proporciona el detalle de cada rescate de tortugas marinas. Este dataset tiene información desde 1998 hasta 2018.

E1_Definicion_variables.csv - es la lista completa de variables encontradas en el archivo Dataset_tortugas.csv y sus explicaciones.

E1_Muestra_de_presentacion.csv - es un ejemplo de cómo debería ser su archivo de envío. El orden de las filas no importa, pero los IDs deben estar todos presentes y correctamente formateados. Los ID tienen la siguiente forma como "CaptureSite"<Número de ID del sitio de captura><año><número de semana>, por ejemplo, CaptureSite_0_201901 es el sitio de captura con número de ID 0 en la primera semana de 2019 (1-6 de enero de 2019).

E1_Categoria_Sitio_Captura.csv - es la categoría del sitio de captura. Las categorías están definidas por sitios que están geográficamente cerca unos de otros. El tipo es una característica que enmascara el sitio de captura.

EJERCICIO Nº 2 (Opcional)

En este ejercicio, hay que procesar el dataset del archivo E2_Dataset_Boro. En él encontrarás los datos de funcionamiento de una de nuestras desaladoras durante parte del año de 2019 y 2020. Esta desaladora se corresponde con uno de los "CaptureSite" del ejercicio 1 y tiene una variable destacada, el Boro (tag "BORON"). Este elemento químico tiene una gran influencia en el comportamiento de las tortugas.

Asumiendo que tu predicción del ejercicio 1 es correcta y has ya tienes los datos de captura de tortugas durante los años del segundo dataset, ¿podrías decirnos a qué CaptureSite se corresponde esta desaladora?

NOTA IMPORTANTE:

Todos los ejercicios deben contener una explicación del porqué de los modelos usados además de explicar los resultados obtenidos en base a las predicciones finales.

Para entregarla, puedes enviarnos el código de la manera que prefieras, repositorio, email, Torrent... (por favor incluye instrucciones para que nosotros podamos probarlo), además, incluye un vídeo demostrando el funcionamiento de lo implementado.