

Título de la tarea: Utilización de librerías de Ciencia de Datos con Python

Unidad: 4

Curso de Especialización y módulo: Inteligencia Artificial y Big Data - Programación en Inteligencia Artificial.

Curso académico: 2022-2023

¿Qué contenidos o resultados de aprendizaje trabajaremos?

Resultados de aprendizaje

- ✓ **RA1.** Caracteriza lenguajes de programación valorando su idoneidad en el desarrollo de inteligencia artificial.
- ✓ **RA2.** Desarrolla aplicaciones de Inteligencia artificial utilizando entornos de modelado.

Contenidos

- 1.- Pandas.
- 2.- Matplotlib.
- 3.- Scikit-learn.
- 4.- Tensorflow y Keras.
- 5.- Pytorch.

Siguiente »

1.- Descripción de la tarea.



Caso práctico

Lorena ha estado leyendo bastante documentación para identificar las librerías que contienen funciones adecuadas al tipo de solución que necesitan en Pick&Deliver.

Muchos miembros de la comunidad comparten a diario distintos desarrollos con modelos de aprendizaje automático, y Lorena ha ido detectando los paquetes que se suelen utilizar en estos desarrollos. Para coger algo de práctica, se propone probar las funciones que le parecen más interesantes de las librerías más famosas.

Para poderse concentrar en aplicarlas bien, decide recurrir a un conjunto de datos sencillo y muy contrastado, que permite detectar un patrón muy concreto entre los datos y así comprobar que los algoritmos están funcionando.

"El dataset de las especies de la flor iris es justo lo que necesito ahora", piensa Lorena "voy a probar las librerías Pandas, Matplotlib y Scikit-learn con estos datos, a ver qué tal se me da".



[@casfatesvano](#) (CC BY-SA)

¿Qué te pedimos que hagas?

✓ Apartado 1: Explora los datos con Pandas

- ◆ Inicia un nuevo notebook, preferiblemente en Google Colab. Para guiarte en el proceso, puedes utilizar este cuaderno-guía con los fragmentos de código indicados en las celdas de texto, pero tendrás que escribir el código en la celda de código correspondiente y ejecutarlo.
- ◆ Importa las librerías **Numpy** y **Pandas**.
- ◆ Importa la función `load_iris` de la biblioteca de datasets de **Scikit-learn** y echa un vistazo rápido a sus principales elementos '`data`', '`target`' y '`target_names`'.
- ◆ Utiliza la clase `DataFrame` de Pandas para crear el dataset `df` y añade la columna "Species" a partir de la secuencia de datos `target`. Utiliza la función `head` para ver los primeros registros del nuevo dataset.
- ◆ Utiliza la función `describe` para ver los principales valores estadísticos del dataset.

✓ Apartado 2: Visualiza los datos con Pyplot.

- Utiliza el paquete **Pyplot** para hacer representaciones gráficas de los datos. Importa **Pyplot** de la librería **Matplotlib** y crea una figura tipo "dispersión de puntos" (Scatter plot) con la variable *sepal length (cm)* en el eje x y la variable *sepal width (cm)* en el eje y.
- Crea otra figura distinguiendo con el color azul la especie "Setosa", con el color verde la especie "Versicolor" y con color rojo la especie "Virginica".
- Analiza cómo se distribuyen los casos si, en vez de representar según las variables de las dimensiones de los sépalos, utilizas las variables basadas en las dimensiones de los pétalos. Representa los datos en una figura tipo scatter, utilizando en el eje x la variable "petal length (cm)" y en el eje y la variable "petal width". De nuevo, distingue las especies con tres colores: color azul la especie "Setosa", con color verde la especie "Versicolor" y con color rojo la especie "Virginica".

✓ Apartado 3: Entrena modelos de aprendizaje automático con Scikit-learn.

- Importa los módulos de la librería **Scikit-learn** que nos permiten hacer un modelo de **regresión lineal**, un modelo de **máquina de vectores soporte**, un modelo de tipo los **K vecinos más cercanos (KNN)**, y un modelo de tipo **árbol de decisión**.
- Genera el conjunto de datos X con las variables de entrada a los modelos, y el conjunto y de las etiquetas o variable de salida del modelo, eligiendo para éste último, la variable "Species".
- Utiliza la función *train_test_split* para separar los datos en el conjunto train y test según el ejemplo.
- Crea un modelo de Regresión Logística. Utiliza la función *fit* para entrenarlo y utiliza la función *predict* sobre los datos de test para medir la precisión del modelo. Muestra el valor de dicha precisión con *print*.
- Crea un modelo de SVC o Máquinas de Vectores de Soporte, entrénalo y calcula la precisión utilizando los datos de test. Muestra la precisión.
- Crea un modelo de KNN o K vecinos más cercanos, entrénalo y calcula la precisión utilizando los datos de test. Muestra la precisión.
- Crea un modelo de árbol de decisión, entrénalo y calcula la precisión utilizando los datos de test. Muestra la precisión.
- Compara los valores de precisión que has ido consiguiendo en los diferentes modelos. ¿Cuál sería el mejor para este dataset?

✓ Apartado 4: Entrena modelos de aprendizaje automático con pocas variables.

- Imagina que no has podido tener todas las variables, y que solo has conseguido los valores de las medidas de los sépalos, y con esos datos debes entrenar un modelo que acierte con el tipo de especie de flor de iris. Para ello, crea un nuevo dataset que tenga solo las columnas de las dimensiones de los sépalos y la de la especie.
- Separa los datos en *X_sepalo* para las variables de entrada e *y_sepalo* para la variable de salida.
- Separa los datos en un conjunto *X_train_s*, *X_test_s*, *y_train_s*, *y_test_s*, para entrenamiento y test del modelo.
- Crea, entrena y mide la precisión de un modelo de Regresión Logística. Muestra la precisión. ¿Es muy diferente al mismo modelo del apartado anterior?.
- Crea, entrena y mide la precisión de un modelo de Máquinas de Vectores Soporte. Muestra la precisión ¿Es muy diferente al mismo tipo de modelo del apartado anterior?.
- Crea, entrena y mide la precisión de un modelo de K vecinos más cercanos. Muestra la precisión ¿Es muy diferente al mismo tipo de modelo del apartado anterior?.
- Crea, entrena y mide la precisión de un modelo de Árbol de decisión. Muestra la precisión ¿Es muy diferente al mismo tipo de modelo del apartado anterior?.
- Compara los valores de precisión que has ido consiguiendo en los diferentes modelos. ¿Cuál sería el mejor para este dataset?.

NOTA IMPORTANTE

Para todos los apartados es necesario incluir las capturas de pantalla de los principales pasos realizados en el documento de la entrega. En dicha presentación,

incluir enlace abierto al notebook con el código para que cualquier persona con el enlace pueda acceder a él.

[« Anterior](#)

[Siguiete »](#)

2.- Información de interés.



Recursos necesarios y recomendaciones

Recursos necesarios

- ✓ Ordenador personal con, al menos, 4 Gigabytes de memoria RAM
- ✓ Conexión a Internet.
- ✓ Navegador web.
- ✓ [Cuaderno-guía con anotaciones de código sugeridas.](#)

Recomendaciones

- ✓ Antes de abordar la tarea:
 - lee con detenimiento la unidad, consulta los enlaces para saber más, examina el material proporcionado por el profesor y aclara las dudas que te surjan con él.
 - Realiza el examen online de la unidad, y consulta nuevamente las dudas que te surjan. Solo cuando lo tengas todo claro, debes abordar la realización de la tarea.
- ✓ No olvides elaborar el documento explicativo y adjuntar el archivo del notebook que se haya generado.



Indicaciones de entrega

Una vez realizada la tarea, el envío se realizará a través de la plataforma. El archivo se nombrará siguiendo las siguientes pautas:

Apellido1_Apellido2_Nombre_PIA_Tarea04

« Anterior Siguiente »

3.- Evaluación de la tarea.

Criterios de evaluación implicados

Criterios de evaluación RA1

- ✓ a) Se ha identificado la estructura de un programa informático.
- ✓ b) Se han valorado características en los lenguajes de programación adecuadas al tipo de aplicaciones a implementar.
- ✓ d) Se han valorado características de los lenguajes de programación para el desarrollo de Inteligencia Artificial.

Criterios de evaluación RA2

- ✓ b) Se han caracterizado entornos de modelo de aplicaciones de inteligencia artificial.
- ✓ c) Se ha definido el modelo que se quiere implementar según el problema planteado.
- ✓ d) Se ha implementado la aplicación de Inteligencia Artificial.
- ✓ e) Se han evaluado los resultados obtenidos.

¿Cómo valoramos y puntuamos tu tarea?

Rúbrica de la tarea	
Apartado 1: Se crea un notebook en Colab, con su título y las celdas de código indicadas de la librería Pandas	2,5 puntos
Apartado 2: Se incluyen las celdas de código indicadas de la librería Matplotlib/Pyplot	2,5 puntos
Apartado 3: Se incluyen las celdas de código necesarias para, al menos, crear y entrenar un modelo de aprendizaje automático con Scikit-learn.	1,5 puntos
Apartado 3: Se incluyen todas las celdas de código para entrenar varios modelos de aprendizaje automático basados en diferentes tipos de algoritmos. Se comparan los índices de precisión de éstos.	2 puntos
Apartado 4: Se modifica el dataset de trabajo a solo dos variables (dimensiones sépalos) y se entrenan varios modelos de aprendizaje automático disponibles en Scikit-learn. Se comparan los índices de precisión con los del apartado anterior.	1,5 puntos
Redacción clara y correcta, sin errores ortográficos	Se resta 0,1 puntos por cada error ortográfico o expresiones incorrectas.

NOTA IMPORTANTE

Aquellos apartados/subapartados en los que las capturas de pantalla no sean claras o no tengan como fondo de pantalla la plataforma con tu usuario mostrando claramente la foto de tu perfil, no serán corregidos.

« Anterior