

DATA SCIENCE

Data Science

DS-BUE-8/Clase 02



Hoja de ruta de hoy

19:00	Repaso y puesta en común de la tarea
19:50	Jupyter notebook y python
20:30	Break
20:40	Hands on: hola python
21:25	Puesta en común
21:55	Cierre y próximo encuentro



¿Comentarios?





PROGRESO /
RECURSOS



COMUNICACIÓN



¿Qué es Data Science?



¿Qué es Data Science?



¿Qué es Data Science?

Recogida de la información



No siempre se tiene la información inmediatamente disponible en una base de datos (u hoja de cálculo) y **su obtención es parte del proceso**. La recogida de la información puede realizarse mediante colocación de **sensores, scrapeo de la web, peticiones a APIs, formularios...** o de cualquier forma que la creatividad nos permita.

Preprocesado



Una vez se tiene la información que se considera necesaria para resolver nuestro modelo, se le debe aplicar un **preprocesado**. Es decir, sin perder información de valor, dar vueltas a la información que ya tenemos para **prepararla para la siguiente fase**. El objetivo es representar la información en un formato que consiga reducir el coste de cómputo y optimizar los resultados de los algoritmos.



¿Qué es Data Science?

Entrenamiento o machine learning



Entrenar el modelo significa **alimentar algoritmos de machine learning** con nuestros datos.

Los algoritmos de machine learning (o aprendizaje automático) son capaces de **predecir y clasificar información** nueva, a raíz de haber sido entrenados con información pasada.

Testing



Por lo general, no existe un algoritmo mejor que otro, algunos rinden mejor en unos casos y otros, en otros casos. Y la manera de hallar la mejor solución es **probándolos uno a uno**, con diferentes configuraciones, hasta encontrar el mejor para nuestro caso. Se sigue una metodología experimental **prueba-error** porque, a priori, es muy difícil adivinar qué tipo de algoritmo y configuración va a tener mejores resultados. Solo los mejores profesionales en machine learning pueden tener una vaga idea de qué tipo de algoritmos pueden funcionar mejor con un set de datos determinado. Pero aún así **tendrían que probar varias opciones y testearlas**.



¿Qué es Data Science?

Visualización e interpretación de datos



Finalmente, para poder comunicarlo a terceros, se tiene que representar de alguna forma el conocimiento obtenido. Se tiene que hallar la manera más **limpia e intuitiva de visualizar los resultados** y, con ayuda de diferentes softwares, crear estas visualizaciones.



Conocimientos ¿previos?

- Probabilidad y Estadística
- Programación (cualquier lenguaje)
- Uso de la terminal (Windows, Mac, Ubuntu)
- [Recomendación] Entender algo de inglés



Conocimientos ¿previos?

- Probabilidad y Estadística
- **Programación (cualquier lenguaje)**
- Uso de la terminal (Windows, Mac, Ubuntu)
- [Recomendación] Entender algo de inglés



Herramientas





- Alto nivel, fácil de usar
 - Rápido y eficientes
 - Gran comunidad online
 - Librerías para prácticamente todo lo que uno quiera hacer
-
- **Importante:** Python 3!
 - **Instalación recomendada:** Anaconda o Miniconda.

```
[1]: print('Hola Mundo')
```

```
Hola Mundo
```

```
[2]: for i in range(5):  
      print(i)
```

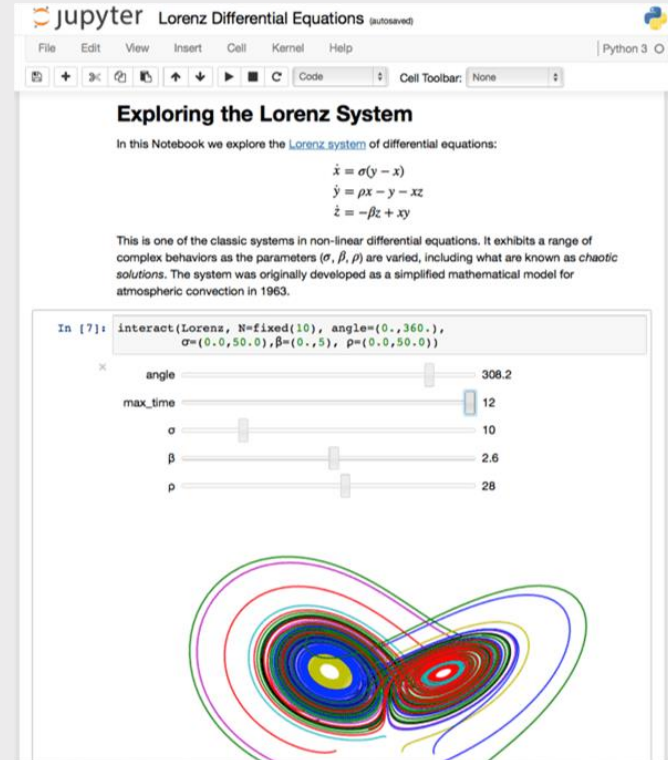
```
0  
1  
2  
3  
4
```





Entorno de trabajo

<https://jupyter.org/>



Y la más importante



|

Buscar con Google

Me siento con suerte

Ofrecido por Google en: [Español \(Latinoamérica\)](#)



Instalación de Python

¿Cómo les fue?
¿lo googlearon?





- ¿Cómo abrir un Jupyter notebook?
- Instalación de librerías
- Carga de librerías



Tipos de datos



Enteros

Son los números que usamos para contar, el 0 y los negativos

- -1
- 0
- 1
- 2

Floats

Son los números enteros, pero con coma.

En python se introducen usando puntos.

- 5.1
- -1.3
- 5.8
- 10.0

Strings

Cualquier carácter que podemos escribir con el teclado de la compu.

Se introducen entre comillas "" o ''.

- 'a'
- ''
- '&'
- 'Hola mundo'

Booleanos

Variables de verdad

Se introducen con una sentencia o escribiendolas

- True
- False
- 1 == 2
- 1==1



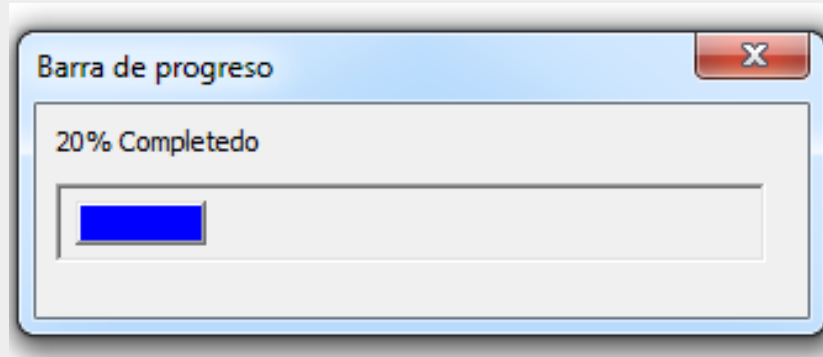


HANDS-ON
TRAINING

`clase_02_python.ipynb`



¿Cómo les fue?



En resumen...



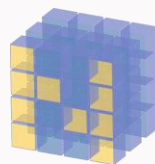
Tipos y manipulación de datos.

- Tipos de datos: enteros, floats, strings, booleanos.
- Operaciones básicas
- Listas
- Comandos: if, else, for.



Para el próximo encuentro...





NumPy

NumPy nos provee de:

- Estructuras de datos
- Operaciones eficientes sobre los datos

Es el corazón del ecosistema de Data Science.

www.numpy.org

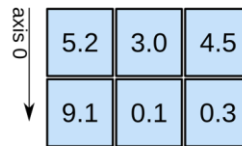
1D array



axis 0 →

shape: (4,)

2D array

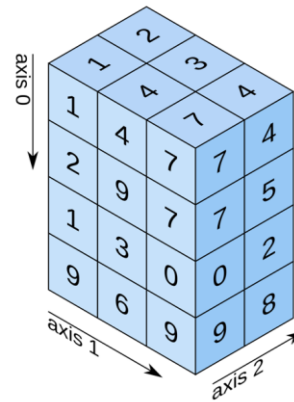


axis 0 ↓

axis 1 →

shape: (2, 3)

3D array



shape: (4, 3, 2)



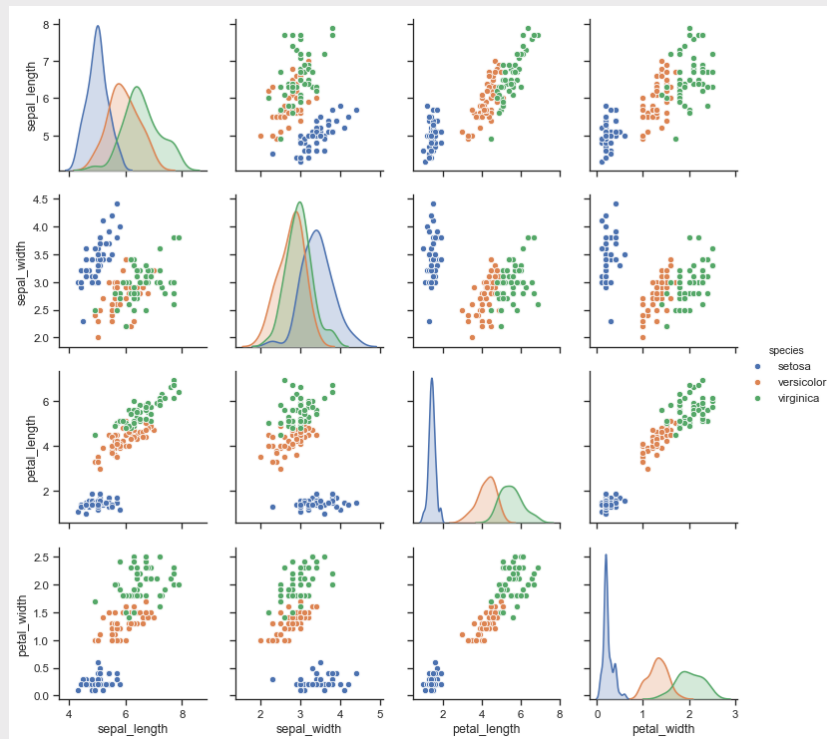


seaborn

¡Lindos Gráficos!

<https://matplotlib.org/>

<https://seaborn.pydata.org/>



Para la próxima...

- Ver los videos de la plataforma de Acámica de Numpy y Matplotlib
- Responder: ¿Qué es un diccionario?
- Tener completo el Hands on de Python

