



Cocinando datos con Python para Investigación reproducible

2019

Descripción de la Actividad de Formación Continua

Nombre	Cocinando datos con Python para Investigación reproducible	
Tipo	Curso de Perfeccionamiento	
Duración en horas	20	
Área de conocimiento del director	Estadística e Investigación Operativa	
Rama de conocimiento	Estadística e Investigación Operativa	
Correo electrónico de contacto	gestion.cyborg@umh.es	
Web	https://cyborg.edu.umh.es/cursos/	
Modalidad		
Presencial	Semipresencial	A distancia
100%	0%	0%
Tutorización virtual		
Tutorización virtual	Elaboración material docente virtual	Presencialidad
Lugar de impartición	Centro Experimental CYBORG. Edificio Severo Ochoa. Universidad Miguel Hernández Campus de Sant Joan d'Alacant	
Plataforma educativa	https://colab.research.google.com	
Fechas inicio y fin de impartición	15 al 19 de julio en horario de Mañana	
Nº mínimo estudiantes	10	
Fecha inicio y fin de matriculación	1 de mayo hasta 30 de junio	
Pago en web <i>(indicar sí o no)</i>	SI	

Información general

La distinción entre un informe científico reproducible y la práctica común actual es como la diferencia entre una lista parcial de ingredientes y una receta. Para hornear una buena barra de pan, no es suficiente saber que contiene harina. Ni siquiera es suficiente saber que contiene harina, agua, sal y levadura. La marca de harina se puede omitir de la receta, al igual que el día de la semana en que se hornea el pan.

Pero la proporción de ingredientes, las operaciones, su tiempo y la temperatura del horno no pueden ser obviadas.

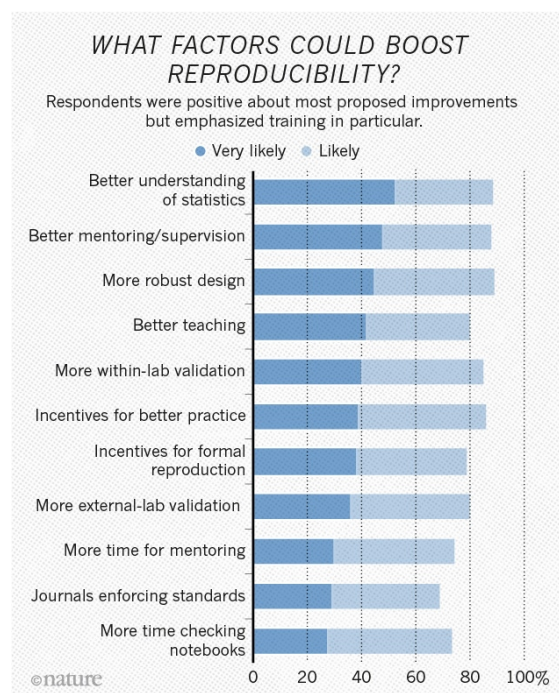
Para la reproducibilidad, una "receta científica", debería hacer una barra de pan científica similar. Si seguimos la receta, pero no obtenemos el mismo resultado, o bien el resultado es sensible a pequeños detalles que no se pueden controlar o la receta no fue lo suficientemente precisa.

Más del 70% de los investigadores han intentado y fracasado en reproducir los experimentos de otro científico, y más de la mitad no han podido reproducir sus propios experimentos.

Esas son algunas de las cifras reveladoras que surgieron de la encuesta de Nature de 1.576 investigadores que respondieron un breve cuestionario en línea sobre la reproducibilidad en la investigación. (Nature 533,452–454 (26 May 2016) <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>)

Respecto a los factores que podrían aumentar la reproductibilidad, casi el 90% - más de 1,000 personas – marcaron "Mejor comprensión de la estadística", "Mejor tutela y supervisión" y "Diseño experimental más robusto".

Los Jupyter Notebooks y el Lenguaje Python nos proporcionan un entorno interactivo web de código abierto que nos permite crear y compartir documentos que contienen en un mismo documento, el acceso a los ficheros de datos, la creación y modificación de variables, el proceso de depuración de los datos, la generación de estadísticas, gráficos, test y modelos estadísticos, mediante secuencias de código de programación que están incluidos y que pueden ser supervisados rápidamente por los investigadores principales al incluir texto para explicar y comentar como se está resolviendo los objetivos planteados en el proyecto, así como incluir la discusión tras la visualización de los resultados.



Al compartir un Jupyter Notebook a otro grupo de investigación estamos aumentando la reproducibilidad de la investigación, ya que en pocos minutos otro grupo puede comprobar si consigue los mismos resultados sobre su conjunto de datos.

Así, los Jupyter Notebooks se están implantando en entornos académicos (UC Berkeley, Stanford, UW, NYU, Cal Poly, etc.), de investigación en el sector público (NASA JPL, KBase), y también en el sector privado (IBM, Facebook, Microsoft, Bloomberg, JP Morgan, WhatsApp, Quantopian, GraphLab, Enthought, etc.). Al tratarse de una arquitectura de módulos abiertos, son ampliamente utilizados para crear todo tipo de soluciones y servicios, tanto comerciales como sin ánimo de lucro.

Este curso introduce en el uso de Python para el análisis de datos estadísticos, bajo la premisa que los investigadores no han utilizado ningún lenguaje de programación.

Programa

1. La Cocina de Python

Aprender cómo usar Jupyter Notebooks sobre Azure de Microsoft para crear documentos que combinan código de programación, texto, imágenes, fórmulas, etc.

2. Los Utensilios de la Cocina de Python (Calculadora)

Conocer las expresiones y los operadores básicos, la precedencia entre operadores y comprender las variables y el proceso de asignación de valores.

3. Los Procesos de Cocción en Python (Control de flujo)

Conocer el tipo de datos booleano, los operadores de comparación y aprender a usar las sentencias de control de flujo de un programa.

4. Las Cacerolas de Python (Tipos de datos)

Conocer los tipos de datos numéricos y cadenas de texto, la función de inspección de tipos, la conversión entre tipos de datos.

5. Las Recetas de Python (Funciones)

Conocer cómo se usan y construyen las funciones definidas por el propio usuario y comprender el alcance de la variable usadas en su definición y en el resultado obtenido.

6. Los Libros de Recetas de Python (Librerías de funciones)

Conocer la librería de funciones estándar, como importarla y usar sus módulos.

7. La Despensa de Python (Estructuras de datos)

Conocer y usar las estructuras de datos como Cadenas, listas, tuplas, diccionarios y conjuntos.

8. El libro de recetas NumPy: Manipulación de arrays

Aprender a usar la librería NumPy para manipular vectores y matrices y conocer las diferencias entre funciones vectorizadas y no vectorizadas.

9. El libro de recetas Pandas: Manipulación de Datos

Aprender a usar la librería Pandas para cargar y procesar datos para el análisis estadístico.

10. El libro de recetas Matplotlib: Visualización de Datos

Aprender a usar la librería Matplotlib para elegir los gráficos apropiados para una o dos variables basados en los tipos de datos disponibles.

