



Fabio Scielzo Ortiz

Científico de datos. Estadístico.

► Madrid, España.

Skills

Estadística y Machine Learning



Python



R



Spark



SQL y MongoDB



Matlab



Visual Studio Code



GitHub



LaTeX



HTML



Español

Nativo



Inglés

B2



Sobre mí

Soy una persona pro-activa, trabajadora, curiosa y constante. Me apasiona la Ciencia de Datos, el Machine Learning, la Estadística y la Programación. Me mantengo en continuo proceso de aprendizaje sobre estas disciplinas, siempre intentando mejorar mis habilidades técnicas.

Mi mayor motivación es poder utilizar mis conocimientos técnicos sobre Ciencia de Datos para poder ayudar y aportar valor en la resolución de problemas reales que tengan un impacto en la sociedad.

Tengo una web personal donde muestro algunos de los **proyectos de Ciencia de Datos** que he desarrollado como parte de mi aprendizaje: **Fabioscielzoortiz.com**

Trayectoria Laboral

Técnico de investigación

Universidad Carlos III de Madrid

09/2023 - Today

Técnico de investigación en el proyecto **Valuation of care needs in dependent or vulnerable people using statistical techniques based on distances and depths** (DDD4D - Distance & Depth Datascience for Dependency) codirigido por Aurea Grané Chavez e Irene Albarran Lozano. Mis tareas como técnico de investigación asociado consisten en:

- Implementación y adaptación de algoritmos de aprendizaje estadístico, basados en distancias, desarrollados por el equipo, mediante software adecuado para grandes volúmenes de datos.
- Mejora y ampliación del paquete de Python **PyDistances**.

Becario de proyectos

UC3M-Santander Big Data Institute (IBiDat)

09/2022 - 05/2023

Mis tareas como becario de proyectos en IBiDat fueron:

- Limpieza y depuración de bases de datos con Azure Data Studio, SQL y Python.
- Proyecto de monitorización de algoritmos de Machine Learning con el Banco Santander.
- Proyecto de predicción de Series Temporales y estimación de venta base con El Corte Inglés.

Beca de colaboración

Ministerio de Educación

11/2022 - 06/2023

Obtuve la beca de colaboración concedida por el Ministerio de Educación, la cual está orientada a la participación activa en el desarrollo de un proyecto de investigación asociado a un departamento universitario. En mi caso trabajé con el Departamento de Estadística de la Universidad Carlos III de Madrid. Principales tareas realizadas:

- Construcción de distancias estadísticas robustas aptas para datos multivariantes de tipo mixto.
- Aplicaciones de dichas distancias al problema de Clustering.
- Elaboración de TFG: Scielzo, F. (2023) Clustering robusto y visualización de datos multivariantes heterogéneos. Trabajo Fin de Grado. Grado en Estadística y Empresa. Universidad Carlos III de Madrid.
- Desarrollo del paquete de Python **PyDistances**.

Contacto

Web Personal

✉ fabioscielzo98@gmail.com

🐙 GitHub

in LinkedIn

Certificados

Cambridge Linguaskill English Test

Mejor expediente Grado en Estadística y Empresa (promoción 2019-2023)

Póster Congreso Nacional de Estadística (SEIO 2023)

Título oficial del Grado en Estadística y Empresa

Computational Statistics with Python

Python para Data Science

PySpark para Data Science

R para Data Science

Profesor de programación en Python

Academia Montero Espinosa

07/2023 - Today

Encargado de diseño y docencia del curso online *Introducción a la programación en Python*, en colaboración Academia Universitaria Montero Espinosa.

Educación

Máster en Big Data Analytics

Universidad Carlos III de Madrid

09/2023 - 07/2024

Máster universitario en Inglés que proporciona una excelente formación para realizar análisis estadísticos y computacionales complejos de grandes volúmenes de datos, con el objetivo de obtener resultados aplicables en identificación de patrones, predicción, previsión, simulación y optimización. Calificaciones a destacar (hasta ahora):

- Cinco (5) **sobresalientes**: Optimization for large scale data (10), High performance computing for Big Data: (9.6), Statistics for Data Analysis (9.5), Machine Learning I (9.2), Predictive modeling (9).
- Dos (2) **notables altos**: Back-end for Big Data (8.2), Technological fundamentals of Big Data (8).
- Nota media: 8.25

Grado en Estadística y Empresa

Universidad Carlos III de Madrid

09/2019 - 07/2023

Este grado universitario está dividido en un bloque de **Ciencia de Datos** (70%) y otro de **Empresa** (30%). Este grado me ha otorgado conocimientos y habilidades relacionados con el mundo de la Ciencia de Datos, concretamente, he adquirido fuertes destrezas en Estadística, Machine Learning y programación. He sido el **mejor expediente del Grado en mi promoción (2019-2023)**, destacando por los siguientes méritos:

- Doce (12) **Matrículas de Honor**: Aprendizaje Automático, Aprendizaje Estadístico, Análisis de Muestras, Análisis Multivariante, Métodos de Regresión, Métodos Avanzados de Regresión, Métodos Bayesianos, Optimización, Probabilidad II, Técnicas de Inferencia Estadística I, Técnicas de Remuestreo, Trabajo Final de Grado.
- Ocho (8) **sobresalientes**: Cálculo I, Economía Financiera, Hojas de Cálculo, Prácticas en IBiDat, Programación I, Programación II, Principios de Economía, Visual Thinking.
- Cinco (5) **notables altos** en las siguientes asignaturas: Álgebra Lineal, Ampliación de Matemáticas, Bases de Datos, Marketing Digital, Métodos Económicos, Procesos Estocásticos.
- Nota media del bloque de Ciencia de Datos de 9, y media total de 8.5.

Proyectos y Publicaciones

Póster en Congreso Nacional de Estadística

11/2023

Sociedad de Estadística e Investigación Operativa (SEIO)

Scielzo, F. and Grané. A. (2023) Robust proposals for clustering with multivariate heterogenous data. En Actas del XL Congreso Nacional de Estadística e Investigación Operativa (SEIO 2023). Presentación de póster.

PyDistances

06/2023

Desarrollo de paquete de Python

PyDistances es paquete de Python que permite calcular distancias estadísticas, las cuales pueden ser usadas en diferentes campos de la estadística y el aprendizaje automático, como clustering, regresión, clasificación, reducción de la dimensión, imputación de valores faltantes, detección de valores atípicos, etc. La contribución más valiosa del paquete es que permite utilizar distancias adecuadas para datos multivariantes heterogéneos, incluso con valores atípicos (versión robusta). Este paquete fue inicialmente desarrollado en mi TFG, y esta siendo ampliado y mejorado en mi posición como Técnico de Investigación, mejoras que serán plasmadas en mi TFM.

Fabioscielzoortiz.com

2024

Página web personal

En mi web personal pueden encontrarse muchos más proyectos de Ciencia de Datos que he desarrollado como parte de mi continuo proceso de aprendizaje.

Madrid, 10th February 2024

Fabio Scielzo Ortiz