



Seminario del Programa de Doctorado “Formación en la Sociedad del Conocimiento” de la Universidad de Salamanca

Seminario Análisis de Big Data 2. Analítica predictiva sobre datos de encuestas: Análisis y tratamiento de datos estructurados en Python

Por Dr. Carlos Arcila Calderón

Profesor del Departamento de Sociología y Comunicación y miembro del
Observatorio de Contenidos Audiovisuales (OCA)
Universidad de Salamanca

Fechas y horario

2, 3 y 4 de julio de 2018, en horario e 10:00 a 14:00 horas (12 horas)
Aula 5, Facultad de Ciencias

Matricula gratuita. 20 plazas. Para inscripción escribir a carcila@usal.es

Contenidos del curso

1. Introducción a la analítica predictiva y a la ciencia de datos
2. Estructuras de datos y librerías básicas en Python
3. Ingesta y transformación de datos
4. Análisis estadístico de encuestas almacenadas en dataframes
5. Clasificadores de Aprendizaje automático sobre datos estructurados con Scikit-Learn
6. Pre-procesamiento del corpus de entrenamiento, reducción de la dimensionalidad y valores perdidos
7. Evaluación de los modelos y modificación de hiperparámetros

Requisitos para los asistentes

- (a) Se requiere haber asistido al seminario Análisis de Big Data 1 o tener conocimiento básico de algún lenguaje de programación (preferiblemente Python 3).
- (b) Se requiere conocimiento básico de estadística y de la técnica de “encuesta”
- (c) No se requiere ordenador personal ya que el seminario se llevará a cabo en un Laboratorio de informática

Lecturas recomendadas

- Arcila, C.; Barbosa, E. & Cabezuelo, F. (2016). Técnicas Big Data: Análisis de textos a gran escala para la investigación científica y periodística. *El Profesional de la Información*, 25 (4), 623-631.
<http://dx.doi.org/10.3145/epi.2016.jul.12>
- Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT Press.