



Ciclo formativo en Big Data y Ciencia del Dato para la Estadística Oficial: MINERÍA DE TEXTO CON R

Profesorado:

José Luis Pino Mejías
Fátima Palacios Rodríguez

Objetivos:

El propósito de la Minería de textos es procesar información no estructurada (textual) con el objetivo de:

- extraer índices descriptivos del texto, y
- permitir la aplicación de las técnicas de Minería de datos o de Aprendizaje estadístico.

La aplicación de las técnicas de Minería de texto en áreas como el marketing o la sociología se está generalizando. En la estadística pública aún es un campo emergente en el que se están evaluando las posibilidades del uso de Internet, las redes sociales y otros contenidos de lenguaje natural para la producción de estadísticas oficiales. El objetivo de este curso es suministrar las herramientas conceptuales y de software que permitan al personal técnico del Sistema Estadístico y Cartográfico de Andalucía iniciarse en la aplicación de las técnicas de Minería de texto.

Programa del curso:

- 1) R básico para Minería de texto
- 2) Extracción de textos. Ejemplo: Redes sociales
- 3) Depuración de textos
- 4) Análisis de frecuencia y asociación de palabras
- 5) Nubes de palabras
- 6) Clasificación y agrupación
- 7) Análisis de sentimientos
- 8) Análisis de Redes sociales

Observaciones:

Duración: 25 horas.

Fecha: 13, 14, 16, 19 y 21 de diciembre de 2016.

Horario: De 9:00 a 14:30 horas (descanso de 30 minutos incluido).

Prerrequisitos: Preferentemente los participantes deben poseer conocimientos prácticos en el campo del Aprendizaje estadístico equivalentes a los adquiridos en el curso "Introducción al Aprendizaje estadístico con aplicaciones en R", así como tener conocimientos elementales sobre técnicas y herramientas para extraer datos de la web (curso "Arquitectura y herramientas tecnológicas para el Big Data").