

CURSO DE FORMACIÓN EN LÍNEA PARA EL PDI

Título del curso

INTRODUCCIÓN PRÁCTICA A TÉCNICAS DE BIG DATA PARA INVESTIGACIÓN: HERRAMIENTA VISUAL ORANGE PARA ANÁLISIS Y PREDICCIÓN DE INFORMACIÓN

Coordinador

¿Participa como formador? Sí

Nombre y apellidos:

Pablo de la Torre Moreno

Correo electrónico:

pablo.delatorre@uca.es

Centro:

Escuela Superior de Ingeniería

Universidad / Empresa:

Universidad de Cádiz

Categoría profesional:

Profesor Titular de Universidad

Área de conocimiento:

Ingeniería Informática

Receptores de la formación

Perfil del PDI al que recomienda esta formación:

Personal Docente e Investigador.

Requisitos previos que deben cumplir los receptores para poder acceder a esta formación:

Manejo básico de equipos informáticos.

Planificación

Duración del curso en días naturales:

25 (Comienza el día 16 de enero y finaliza el día 9 de febrero de 2024)

Horas estimadas de trabajo:

25

Beneficios para la docencia, para la investigación y/o para la gestión

El aprendizaje computacional y análisis de *big data* desempeñan un papel cada vez más relevante en la investigación gracias a sus capacidades para analizar la información y extraer modelos complejos. Considerando la facilidad de estos sistemas para descubrir patrones ocultos, realizar predicciones con alta precisión y, con carácter general, obtener información relevante, la exploración no automatizada de los datos empieza a resultar obsoleta, así como más costosa en tiempo y recursos.

En este contexto, se plantea un proyecto de formación que cubra, en primer lugar, los conceptos necesarios para entender cómo funcionan estas técnicas y su aplicación a la investigación y a la docencia; en segundo lugar, el aprendizaje de una herramienta orientada a este objetivo y utilizada en el ámbito científico. Específicamente, Orange Data Mining es el resultado de un proyecto *open source* para visualización, transformación, análisis, clasificación y aprendizaje automático de conjuntos de datos.

Este curso pretende aportar conocimiento y nociones básicas prácticas sobre *big data* y análisis de datos en general, técnicas de clasificación y análisis predictivo, que se pondrán en práctica mediante el aprendizaje de la herramienta Orange mencionada.

Para dar universalidad al desarrollo de cada curso, evitando usar un conjunto de datos demasiado especializado en una materia concreta, los ejemplos se realizarán sobre análisis y detección de géneros cinematográficos, aprendiendo cómo se puede emplear el sistema para, por ejemplo, clasificar automáticamente películas según imágenes o información proporcionada sobre éstas, o realizar predicciones por factores como la edad de los espectadores o el beneficio en taquilla, entre otros. Las técnicas aprendidas con estos ejemplos son extrapolables a otros dominios de información. Al terminar el ciclo formativo, los participantes habrán adquirido habilidades fundamentales en el uso del aprendizaje computacional con objeto de visualizar, interpretar y, en definitiva, aprovechar al máximo los conjuntos de datos de estudio.

Tutorías

Herramientas que empleará para asesorar a los estudiantes:

Foros y mensajería interna. El formador se compromete a responder a las dudas de los alumnos en un plazo no superior a 48 horas, inclusive en fines de semana.

Competencias que se desarrollan en el curso

Técnicas de investigación.
Innovación.

Objetivos

Generales:

Adaptar los estudios analíticos a las nuevas tecnologías.
Mejorar la eficiencia y uso de recursos en la práctica de la investigación.
Aprender a utilizar una herramienta orientada a la exploración de datos.
Estimar la información significativa mediante el apoyo de técnicas visuales.

Específicos:

Familiarizarse con conceptos relativos a *big data* y aprendizaje automático.
Identificar los mejores modelos analíticos de acuerdo a la información de interés.
Evaluar y aplicar diferentes clasificadores/predictores de variables categóricas.
Evaluar y aplicar diferentes clasificadores/predictores de variables numéricas.

Contenidos

Tema 1. Introducción al *big data* y primeros pasos con Orange.
Tema 2. Introducción al *machine learning* para análisis predictivo.
Tema 3. Técnicas de clasificación y demostración práctica.
Tema 4. Técnicas de regresión y demostración práctica.

Formato de los contenidos

Páginas web, vídeos en YouTube, datos de ejemplo, y archivos de referencia.

Formato de las tareas propuestas

Cuestionarios auto-evaluables.
Imágenes en formatos nativos y estándares.

Metodología

Aprendizaje on-line.
Aprendizaje basado en proyectos.

Referencias

- Orange (<https://orangedatamining.com/>).
- J. Demšar, et al. Orange: data mining toolbox in Python. *The Journal of machine Learning research* 14.1 (2013): 2349-2353.
- Z-H. Zhou. *Machine learning*. Springer Nature, 2021.
- C-W. Tsai, et al. Big data analytics: a survey. *Journal of Big data* 2.1 (2015): 1-32.
- H. Hassani, et al. Text mining in big data analytics. *Big Data and Cognitive Computing* 4.1 (2020): 1.
- D. Keim, et al. Big-data visualization. *IEEE computer graphics and applications* 33.4 (2013): 20-21.
- M. Felt. Social media and the social sciences: How researchers employ Big Data analytics. *Big data & society* 3.1 (2016).