



CD900 Proyecto Final de Graduación

Descripción

En este curso el estudiante desarrollará un proyecto de principio a fin, siguiendo las buenas prácticas de gestión de proyectos en Ciencia de Datos según la metodología CRISP-DM. Para el proyecto el estudiante utilizará datos de la empresa en donde trabaja o datos en los cuales está particularmente interesado(a).

Objetivos

En este curso el estudiante será capaz de:

1. Utilizar adecuadamente la metodología CRISP-DM para el desarrollo de proyectos en Ciencia de Datos.
2. Distinguir entre un proyecto de segmentación, clasificación, regresión y series de tiempo.
3. Elegir, para un problema concreto, qué técnicas de Ciencia de Datos son las más apropiadas.
4. Determinar la herramienta de software más adecuada para el enfrentar el problema planteado.
5. Generar los modelos y patrones elegidos utilizando una herramienta o paquete de Ciencia de Datos.
6. Evaluar la calidad de un modelo, utilizando técnicas sencillas de evaluación, validación cruzada o remuestreo.

Metodología

Basado en la teoría y en la aplicación directa de los conceptos aprendidos. Para esto se dispondrán de las siguientes herramientas:

1. Una video conferencia semanal, las cuales quedarán grabadas en Zoom, para que los alumnos la puedan acceder en cualquier momento.

2. Trabajos prácticos semanales.
3. Foros para plantear dudas al tutor y compañeros.
4. Aula virtual en Moodle.

Luego de este curso el estudiante será capaz de:

Desarrollar proyectos de Ciencia de Datos de principio a fin, es decir, iniciando con la limpieza de datos hasta la generación de conocimiento derivados de los datos de la organización.

Contenido

1. Fundamentos de la metodología para la implementación de proyectos en Ciencia de Datos CRISP-DM.
2. Estudio comparativo de diferentes tipos problemas en Ciencia de Datos.
3. Presentación del anteproyecto.
4. Presentación y defensa del proyecto final.

Evaluación

El curso se evalúa con 4 tareas, una por semana, cada tarea tiene un valor de 25 puntos. La nota mínima de aprobación es de 70.

Bibliografía

[1] José Alberto Gallardo Arancibia. "Metodología para la definición de requisitos en proyectos de data mining (er-dm)". <http://purl.org/dc/dcmitype/Text>. Universidad Politécnica de Madrid, 2009. url: <https://dialnet.unirioja.es/servlet/tesis?codigo=20961> (visited on 10/19/2022).

[2] Ana Azevedo and Manuel Filipe Santos. "KDD, SEMMA and CRISP-DM: a parallel overview". In: IADS - DM (2008). Accepted: 2012-06-14T09:51:14Z. url: <https://recipp.ipp.pt/handle/10400.22/136> (visited on 10/19/2022).

[3] CRISP-DM 1.0: Step-by-step Data Mining Guide. Google-Books-ID: po7FtgAACAAJ.SPSS, 2000. book.

[4] Graham J. Williams and Simeon J. Simoff. Data Mining: Theory, Methodology, Techniques, and Applications. Google-Books-ID: 44z0BwAAQBAJ. Springer, Jan. 22, 2006. 341 pp. isbn: 978-3-540-32548-2.