



Curso de Introducción a la Ciencia de Datos

Syllabus

Equipo Docente

Paula Aguirre

Doctora en Astrofísica, Pontificia Universidad Católica, 2012

Ingeniero Civil Mecánico y Licenciada en Astronomía UC. Doctor en Astrofísica UC.
Profesor Asistente del Instituto de Ingeniería Matemática y Computacional UC, e
Investigadora del Centro Nacional de Investigación para la Gestión Integrada de Desastres
Naturales (CIGIDEN).

Descripción del curso

Este curso introduce a los estudiantes al campo de la ciencia de datos a través de sus principios básicos y las principales técnicas y herramientas utilizadas. Los estudiantes aprenderán sobre recolección e integración de datos, análisis exploratorio de datos, análisis descriptivo y predictivo, y creación de productos de información. Los temas serán tratados en amplitud, más que en profundidad, haciendo énfasis en la integración y síntesis de conceptos, y evaluados en la aplicación a soluciones de problemas contextualizados.

Resultados de aprendizaje

- Describir lo que es ciencia de datos: sus procesos, aplicaciones y relevancia para distintos tipos de organizaciones.
- Implementar procesos simples de extracción, limpieza e integración de datos de diversas fuentes en lenguaje Python.
- Explicar la importancia del análisis exploratorio de datos en ciencia de datos y aplicar sus métodos estadísticos y gráficos a un estudio de caso.
- Aplicar algoritmos básicos de aprendizaje de máquina para el análisis descriptivo y predictivo.
- Analizar la importancia de una visualización y comunicación efectiva de los datos.

Estrategias metodológicas

- Clases expositivas en video.
- Cuadernos virtuales de programación (Jupyter Notebooks).
- Exposición y análisis de casos.
- Lecturas y reflexión individual.
- Cuestionarios en línea.

Estructura del Curso

Introducción a Ciencia de Datos y aprendizaje de máquinas.

- ¿Qué es la ciencia de datos? Usos y aplicaciones.
- El proceso de ciencia de datos.
- Herramientas de Python para ciencia de datos.

Obtención y transformación de datos.

- Tipos y formatos de datos.
- Fuentes y métodos de extracción de datos.
- Limpieza e integración de conjuntos de datos.

Análisis Exploratorio de Datos.

- Estadísticas descriptivas.
- Análisis gráfico univariado.
- Análisis gráfico multivariado.

Algoritmos básicos para análisis predictivo.

- Introducción a aprendizaje de máquinas: tipos de aprendizaje y algoritmos básicos.
- Algoritmos de aprendizaje supervisado.
- Algoritmos de aprendizaje no supervisado.
- Resumen y comunicación de resultados.

El curso tiene como requisitos de aprobación las siguientes instancias evaluativas:

Actividad	Evaluación
Tareas	45%
Controles	40%
Trabajo Final	15%

Información General

CURSO:	INTRODUCCIÓN A LA CIENCIA DE DATOS
TRADUCCIÓN:	<i>Introduction to data science</i>
SIGLA:	MDS3020
CRÉDITOS :	5 UC
HORAS:	90 horas de dedicación total (24 directas y 66 indirectas)
REQUISITOS:	INF3103
RESTRICCIONES:	(Programa MDS) o (Programa MAN)
CONECTOR:	0
CARÁCTER:	Mínimo
TIPO:	Cátedra, laboratorio
CALIFICACIÓN:	Estándar
PALABRAS CLAVE:	Computación, inteligencia, artificial, ciencia, datos, introducción, data, science
NIVEL FORMATIVO:	Magíster

Bibliografía

Mínima:

- Wes McKinney. "Python for data analysis: Data wrangling with Pandas, NumPy, and IPython". O'Reilly Media, 2017.
- Pedregosa, F., Michel, V., Varoquaux, G., et al. "Scikit-learn: Machine Learning in Python", Journal of Machine Learning Research, 12(85):2825–2830, 2011.

Complementaria:

- Grus, Joel. "Data Science from Scratch". O'Reilly Media. 2019.