

Programa



CURSO:INTRODUCCIÓN A LA CIENCIA DE DATOS

TRADUCCIÓN:INTRODUCTION TO DATA SCIENCE

SIGLA:IMT2200

CRÉDITOS:10 UC

MÓDULOS:03

CARÁCTER:MÍNIMO

TIPO:CÁTEDRA

CALIFICACIÓN:ESTÁNDAR

DISCIPLINA:CIENCIA DE DATOS

PALABRAS CLAVE:CIENCIA DE DATOS, RECOLECCIÓN, ANÁLISIS DE DATOS

NIVEL FORMATIVO:PREGRADO

I.DESCRIPCIÓN DEL CURSO

Las organizaciones utilizan sus datos para apoyar la toma de decisiones, y para desarrollar productos y servicios intensivos en datos. El conjunto de competencias requeridas para apoyar estas funciones se ha agrupado bajo el término Ciencia de Datos. En este curso los estudiantes analizarán la importancia de este campo y su crecimiento exponencial, describiendo 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.

II.RESULTADOS DE APRENDIZAJE

- 1.Describir lo que es la ciencia de datos, y su importancia para la ciencia, sociedad y negocios.
- 2.Identificar los conjuntos de habilidades necesarios para ser un científico de datos.
- 3.Identificar problemas éticos y de privacidad que emergen en ciencia de datos.
- 4.Explicar las etapas y tareas que forman parte del ciclo de vida de un proyecto de ciencia de datos.
- 5.Reconocer distintos tipos y formatos de datos estructurados y no estructurados.
- 6.Desarrollar el proceso de extracción, transformación y carga de datos para un proyecto sencillo de ciencia de datos.
- 7.Identificar los fundamentos matemáticos y estadísticos para el análisis exploratorio y modelamiento estadístico en un proyecto de ciencia de datos.
- 8.Utilizar técnicas de análisis exploratorio y modelamiento apropiados para problemas sencillos de ciencia de datos.
- 9.Aplicar algoritmos básicos de aprendizaje de máquina para el análisis descriptivo y predictivo en proyectos sencillos de ciencia de datos.
- 10.Comunicar efectivamente los resultados de un proyecto de ciencia de datos.

III.CONTENIDOS

1.Introducción

- 1.1.¿Qué es la ciencia de datos? Definiciones y aplicaciones.
- 1.2.Herramientas fundamentales para la ciencia de datos.
- 1.3.Tipos de datos.
- 1.4.Ética y gobernanza de datos.

2.Extracción, transformación y carga de datos.

- 2.1.Fuentes y tipos de datos.
- 2.2.Limpieza y preparación de datos.
- 2.3.Combinación y estructuración de conjuntos de datos.

3.Fundamentos matemáticos y estadísticos para el análisis de datos.

- 3.1.Vectores y matrices.
- 3.2.Distribuciones de probabilidad.
- 3.3.Hipótesis e inferencia estadística.

4.Análisis Exploratorio de Datos.

- 4.1.Análisis gráfico.
- 4.2.Análisis cuantitativo.
- 4.3.Estadísticas y correlaciones.

5.Modelamiento estadístico.

- 5.1.Regresión lineal simple.
- 5.2.Regresión múltiple.
- 5.3.Estimación de parámetros.

6.Introducción a algoritmos de aprendizaje de máquinas.

- 6.1.Aprendizaje supervisado y no supervisado.
- 6.2.Algoritmos básicos de aprendizaje de máquina.
- 6.3.Visualización y comunicación de datos y resultados.

IV. ESTRATEGIAS METODOLÓGICAS

-Clases expositivas.

-Ayudantías.

-Talleres prácticos.

-Tareas.

-Proyecto.

-Expertos invitados.

-Análisis de casos.

-Nota: en este curso se utilizará el lenguaje de programación Python.

V. ESTRATEGIAS EVALUATIVAS

-Controles: 20%

-Tareas: 50%

-Proyecto: 30%

VI. BIBLIOGRAFÍA

Mínima

Cathy O'Neil and Rachel Schutt. "Doing Data Science, Straight Talk from the Frontline". O'Reilly Media, 2013.

Wes McKinney. "Python for data analysis: Data wrangling with Pandas, NumPy, and IPython". O'Reilly Media, 2da edición, 2017.

Complementaria

Joel Grus. "Data Science from Scratch: First Principles with Python". O'Reilly, 2ª edición, 2019.

Foster Provost and Tom Fawcett. "Data Science for Business: What you need to know about data mining and data-analytic thinking". O'Reilly Media, 2013.

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE,
INSTITUTO DE INGENIERÍA MATEMÁTICA Y COMPUTACIONAL / SEPTIEMBRE 2020