

Facultad de Ingeniería y Ciencias
Escuela de Informática y Telecomunicaciones

PROGRAMA DE ASIGNATURA
Data Science

I. Identificación

- Nombre : Data Science
- Código : CIT-20XX
- Créditos : 6
- Duración : Semestral
- Ubicación en plan de estudio: Semestre 9
- Requisitos : CIT-2204 Probabilidades y estadísticas,
CIT-2009 Bases de datos Avanzadas,
CIT-2013 Inteligencia Artificial
- Sesiones semanales : 2 cátedras, 1 ayudantía

II. Descripción del curso

Curso que tratará con la recuperación, preparación, análisis y visualización de datos. Se integran los conocimientos adquiridos en los cursos de ciencias básicas con aquellos de especialidad para formar al estudiante en la manipulación de datos para la obtención de información relevante.

III. Resultados de aprendizaje

Al finalizar el curso el/la estudiante será capaz de:

1. Diseñar e implementar soluciones a partir de herramientas especializadas, o bien integrando herramientas genéricas para recabar datos, proveer un framework de análisis de ellos y el posterior procesamiento del análisis resultante.
2. Implementar visualizaciones de datos para presentar estadísticos descriptivos.
3. Implementar técnicas de análisis de datos utilizando herramientas especializadas, para estudiar relaciones de causalidad, tendencias, entre otros.
4. Implementar un proyecto grupalmente, de aplicación de ciencia de datos, presentando y documentando los resultados obtenidos, en un contexto de aplicación de las TICs.

IV. Unidades Temáticas

- Introducción a la Ciencia de Datos
 - Fundamentos de matemáticas, probabilidades, estadística y programación
 - Fuentes de datos
- Familiarización con el uso de herramientas de Ciencia de Datos
 - Herramientas para obtención de datos

- Técnicas de limpieza de datos
- Análisis exploratorio y visualización
 - Análisis exploratorio univariado y multivariado
 - Estadística descriptiva
 - Herramientas para visualización de datos
- Inferencia y Regresión
 - Inferencia estadística – Espacio de eventos, distribuciones y test de hipótesis
 - Clasificación y regresión
- Aprendizaje Automático
 - Algoritmos de reducción de dimensionalidad y de selección de atributos
 - Data mining: algoritmos de clustering, cálculo de reglas de asociación, detección de anomalías
- Productos de Ciencias de Datos
 - Diseño y construcción de productos de Ciencias de Datos

V. Metodología

Se contemplan dos clases semanales de cátedra, mezclando –a lo largo del semestre– sesiones de carácter expositivo (basadas en presentaciones electrónicas, con apoyo adicional de pizarrón, y/o contenido audiovisual) con sesiones de trabajo donde se realizarán talleres interactivos con software afín. Esto será complementado con lectura de textos (artículos científicos y textos de actualidad sobre el tema) y material adicional, para la realización de tareas, laboratorios o trabajos de investigación, permitiendo desarrollar habilidades relacionadas con el autoaprendizaje continuo.

Se fomentará la evaluación del conocimiento adquirido, el nivel de comprensión, la capacidad de abstracción y modelamiento, análisis, síntesis, y aplicación. Esto será medido a través de trabajos prácticos, más dos pruebas solemnes y un examen final escrito en la hora y día que establezca la Dirección de la Escuela.

VI. Evaluación

Se contempla la realización de trabajos prácticos, dos pruebas solemnes de igual valor y un examen. La nota final (NF) del curso se calculará a partir de una nota de presentación (NP) y la nota del examen (NE). Asimismo, para el cálculo de la NP participan las notas de las pruebas solemnes (S1 y S2) y la nota de tareas (NT).

La nota del curso se calculará de la siguiente forma:

$$NF = NP \times 0.7 + NE \times 0.3, \text{ donde:}$$

$$NP = (0.25 \times S1 + 0.25 \times S2 + 0.2 \times NT) / 0.7$$

Será condición de aprobación del curso, además que NT \geq 4.0. El profesor podrá eximir del examen final con NP \geq 5.0. La inasistencia a una prueba solemne implicará su reemplazo con la NE.

VII. Bibliografía Básica

1. O'Neil, C., Schutt, R. *Doing Data Science: Straight Talk from the frontlines*. O'Reilly, 2013.
2. Tufte, E. *The Visual Display of Quantitative Information*, 2nd edition; Graphics Press, 2001.
3. Van der Plas, J. *Python Data Science Handbook: Essential tools for working with data*. O'Reilly, 2016.

PAUTAS ETICAS BASICAS

El aula es un espacio donde los intercambios buscan generar un clima que potencie el aprendizaje, basado en el respeto y el buen trato. Las diferencias, tanto entre estudiantes, como entre estudiante y docentes, deben abordarse desde este marco de respeto.

La universidad cuenta con dos reglamentos importantes de conocer:

- *Reglamento de Convivencia*
- *Normativa de Prevención y Sanción de Acciones de Discriminación, Violencia Sexual y/o de Género.*

Puedes consultar los reglamentos aquí: <https://www.udp.cl/universidad/reglamentos-y-politicas/>

El plagio es el uso de las ideas o trabajo de otra persona sin el adecuado consentimiento. El plagio puede ser intencional o no. El plagio intencional es el claro intento de hacer pasar el trabajo o ideas ajenas como el suyo propio para su beneficio. El plagio no intencional puede ocurrir si Ud. no conoce el mecanismo adecuado de referenciar la fuente de sus ideas e información. Si no está seguro de los métodos aceptados para referenciar, debería consultar con su profesor, tutor o personal de biblioteca.

El plagio comprobado es una actitud que puede resultar en severas sanciones disciplinarias y/o en la exclusión de la Universidad (Artículo 44, Reglamento del Estudiante de Pregrado).

Elaborado por: Martín Gutiérrez
Fecha revisión: Septiembre de 2019
Fecha vigencia: Marzo de 2020