



PROGRAMA DE ASIGNATURA

Nombre del curso o actividad	SISTEMAS DISTRIBUIDOS PARA BIG DATA
Nombre del programa	Doctorado en Ciencias de la Ingeniería mención Automática
Descripción	<i>Código: 11331, Nivel: 2 o 3, Créditos SCT-Chile: 10, Tipo: Electiva, Plan de Estudios: Resolución N°1118 del 03/04/20.</i> Este programa ofrece una formación integral en ciencia de datos, utilizando Python y Julia, para enseñar desde los fundamentos del aprendizaje automático hasta el manejo de grandes volúmenes de datos. En cuatro unidades, los estudiantes aprenderán a aplicar modelos estadísticos, optimizar algoritmos, y extraer conocimientos valiosos de los datos. Con un enfoque práctico, incluye análisis de casos, laboratorios y trabajo en equipo. La evaluación combina retroalimentación continua con proyectos prácticos, preparando a los participantes para resolver desafíos reales en ciencia de datos.
Objetivos	El objetivo de este programa es dotar a los participantes de conocimientos fundamentales y habilidades prácticas en Ciencia de Datos, utilizando los lenguajes de programación Python y Julia. A través de una estructura modular, el programa se enfoca en la comprensión y aplicación de técnicas de aprendizaje automático, desde sus conceptos básicos hasta su implementación en proyectos complejos. Los participantes aprenderán a manejar grandes volúmenes de datos, aplicar modelos estadísticos y algoritmos de aprendizaje para extraer conocimientos útiles, y desarrollar habilidades para la estimación y modelización probabilística. Además, se explorarán estrategias para la manipulación, limpieza y preparación de datos, así como para la selección y optimización de modelos. Este programa busca equilibrar la teoría con la práctica, preparando a los alumnos para enfrentar los retos del análisis de datos en diversos contextos, y promoviendo una comprensión profunda de cómo la ciencia de datos puede ser aplicada para resolver problemas reales de manera efectiva.
Contenidos	Unidad 1: El Proceso de Ciencia de Datos con PYTHON y JULIA • Aprendizaje automático • Principales categorías del aprendizaje automático • Etapas típicas de un proyecto de aprendizaje automático Unidad 2: Datos, Modelos, Parámetros, Estimación y Modelización Probabilística • Obtención de conocimientos adaptando un modelo a los datos • Maximización del valor de una función de coste • Manipulación, limpieza y preparación de los datos Unidad 3: Algoritmos de Aprendizaje Automático • Elección y diseño de características. • Validación cruzada, selección de modelos y optimización de sus hiper parámetros. • Dificultades asociadas al aprendizaje automático, inadaptación y sobre adaptación • Reducción de la dimensionalidad de los datos de entrenamiento Unidad 4: Manejo de Grandes Volúmenes de Datos en un solo Ordenador • Regresión lineal y polinómica • Regresión logística



	• K-vecinos más próximos • Máquinas de vectores soporte • Árboles de decisión • Estrategias de aprendizaje no supervisado, como agrupación, densidad y anomalías.
Modalidad de evaluación	Estrategias Metodológicas El curso se desarrolla utilizando como eje articulador el Análisis de casos reales y simulados relacionados con la aplicación a la Ingeniería a los departamentos que participan en el programa de doctorado y el desarrollo de soluciones simples a problemáticas desde la Telemática, considerando actividades en laboratorio y sala de clases, de manera individual y en equipos de trabajo. Evaluación Se consideran entregas y presentaciones de avance de las respectivas investigaciones, informes, exposiciones de los casos analizados y de las soluciones planteadas a los problemas presentados. Evaluaciones Sumativas de tipo teórico - práctico incluye pruebas escritas, trabajos de investigación en equipos, elaboración de informes, exposiciones y valoración de prototipo básico.
Resultados de aprendizajes esperados	Resultado de Aprendizaje General: Formular proyectos complejos con IoT/5G, aplicando Python y Julia para manejar los dispositivos IoT con comunicaciones móviles e inteligencia artificial, adaptando programas que estén en la red Internet sin plagio. Resultados de Aprendizaje Específicos: • Categorizar los conceptos fundamentales a través del manejo de objetos, jerarquías, procedimientos y estructuras, para el uso del lenguaje Python y del aprendizaje automático. • Adaptar los modelos de datos para extraer conocimiento de los datos, usando metodología Knowledge Discovery in Databases (KDD), para preparar los datos que permitan la maximización del valor. • Procesar datos estadísticos correlacionando los datos con el fin de reducir la dimensionalidad.
Bibliografía	Básica: • "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition" por Aurélien Géron, publicada en octubre de 2022 por O'Reilly Media, Inc., con ISBN: 9781098125974. • Wes McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, Editor: O'Reilly Media; ISBN-10: 1449319793, ISBN-13: 978-1449319793. • M. P. Deisenroth, A. A. Faisal, and C. S. Ong. "Mathematics for Machine Learning", ISBN-13: 978-1108455145, ISBN-10: 110845514X, 2019. Recomendada: • "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition" por Aurélien Géron, publicada en octubre de 2022 por O'Reilly Media, Inc., con ISBN: 9781098125974.