



Programa de Asignatura

I. IDENTIFICACIÓN				
Carrera o programa: Ingeniería Civil en Computación e Informática				
Unidad responsable: Escuela de Ingeniería				
Nombre de la asignatura: Introducción a Data Science				
Código: ECIN-00512				
Semestre en la malla¹: 5				
Créditos SCT - Chile: 5				
Ciclo de Formación	Básico		Profesional	X
Tipo de Asignatura	Obligatoria	X	Electiva	
Clasificación de área de conocimiento²				
Área: Ingeniería y Tecnología		Sub área: Ingeniería Informática		
Requisitos:				
Pre-requisitos:		Requisito para:		
<ul style="list-style-type: none">• ECIN-00336 Estadística• ECIN-00407 Estructuras de Datos				

II. ORGANIZACIÓN SEMESTRAL							
Horas Dedicación Semanal (Cronológicas)	Docencia Directa		3	Trabajo Autónomo	5	Total	8
Detalle Horas Directas	Cátedra	Ayudantía	Laboratorio	Taller	Terreno	Exp. Clínica	Supervisión
	1.5		1.5				

¹Este campo

²Clasificación del curso de acuerdo a la OCDE



III. APOORTE AL PERFIL DE EGRESO

La asignatura contribuye al dominio 1 del perfil de egreso, “Conocimiento científico y disciplinario”. Además, contribuye al dominio 2 “Habilidades y Actitudes Personales y Profesionales”. También contribuye al dominio 4 “Habilidades para la Práctica de la Ingeniería”. Al finalizar la asignatura las y los estudiantes serán capaces de aplicar técnicas de ciencia de datos y de sus disciplinas asociadas (estadística y aprendizaje automático) para modelar y resolver problemas aplicados en diversos contextos.

IV. HABILIDADES PERFIL DE EGRESO (RELACIÓN)

1.3 Aplicar conocimientos, métodos y herramientas de la especialidad para resolver problemas complejos de Ingeniería de Software, Plataformas y Gestión de Tecnologías.

2.1 Identificación, formulación, modelación y resolución de problemas complejos de ingeniería considerando las interacciones y la dinámica de las variables.

2.2 Aplicación del método científico para diseñar, conducir y realizar investigación en ingeniería.

4.3 Concebir las soluciones TIC requeridas en las organizaciones haciendo uso eficiente de los recursos como personas, información, y procesos. Estas soluciones consideran las arquitecturas TI junto a sus modelos de servicios y modelos operativos; los sistemas de software; y las plataformas de cómputo y comunicaciones junto a sus servicios asociados.

4.4 Diseñar las soluciones TIC para la industria intensiva en procesamiento de la información. Estas soluciones consideran las arquitecturas TI junto a sus modelos de servicios y modelos operativos; los sistemas de software; y las plataformas de cómputo y comunicaciones junto a sus servicios asociados.

4.5 Implementar las soluciones TIC. Estas soluciones consideran las arquitecturas TI junto a sus modelos de servicios y modelos operativos; los sistemas de software; y las plataformas de cómputo y comunicaciones junto a sus servicios asociados.

4.6 Operar arquitecturas TI junto a sus modelos de servicios y modelos operativos; los sistemas de software; y las plataformas de cómputo y comunicaciones junto a sus servicios asociados.



CG6 Trabajo en equipo: Capacidad de desarrollar labores tipo grupal, facilitando el despliegue de las habilidades de sus integrantes, promoviendo el compromiso y un clima respetuoso, que permita facilitar la sinergia con personas de diferentes disciplinas y/o culturas, con el propósito de alcanzar metas colectivas. El egresado UCN, en la interacción del trabajo colaborativo, respeta la diversidad de pensamiento de opinión, de expresión y de conciencia que se genera en la relación con otros, en consecuencia, con el valor de la libertad.

V. RESULTADOS DE APRENDIZAJE

1. Identificar los conceptos fundamentales de Data Science y su relación con otras disciplinas como matemática/estadística y aprendizaje automático.
2. Implementar métodos de obtención de datos en distintos dominios de aplicación
3. Distinguir las características de los datos a través de técnicas de visualización de datos
4. Diseñar modelos básicos para el análisis de datos en diversos contextos de aplicación
5. Aplicar modelos básicos de análisis de datos teniendo en consideración el dominio que está siendo analizado
6. Interpretar los resultados obtenidos mediante modelos de análisis de datos
7. Contribuir a la generación de acciones desde el rol que ejerce dentro del equipo, favoreciendo el aprendizaje colectivo.

VI. ÁREAS TEMÁTICAS

1. Introducción
 - 1.1 Definición de Data Science
 - 1.2 Metodologías
 - 1.3 Relación con Estadística y Machine Learning
 - 1.4 Aplicaciones
2. Obtención de datos
 - 2.1 Data scraping
 - 2.2 Repaso álgebra lineal
 - 2.3 Grafos y procesamiento de redes
 - 2.4 Procesamiento de lenguaje natural
3. Exploración y Visualización de Datos
 - 3.1 Visualización de distribución de valores de una variable



- 3.2 Visualización de relaciones entre pares de variables
- 3.3 Visualización multidimensional de datos
- 4. Modelamiento estadístico y aprendizaje automático
 - 4.1 Estadística descriptiva básica
 - 4.2 Regresión lineal y clasificación lineal (SVM y regresión logística)
 - 4.3 Modelamiento no lineal y regularización
 - 4.4 Métricas para regresión y clasificación
 - 4.5 Evaluación de modelos
- 5. Otras técnicas de modelamiento
 - 5.1 Aprendizaje supervisado (árboles de decisión, K-NN y redes neuronales poco profundas)
 - 5.2 Aprendizaje no supervisado (K-Means, PCA y reglas de asociación)

VII. ORIENTACIONES METODOLÓGICAS

1. La metodología a desarrollar en esta asignatura debe favorecer la interacción entre las y los estudiantes a través de trabajos prácticos colaborativos que permitan la solución a problemas específicos contextualizados a la asignatura.
 - Se sugiere el uso de clases expositivas y participativas con método combinado, es decir, clases expositivas con alternancia de trabajos en grupo de corta duración para responder preguntas.
 - Se sugiere la utilización de la metodología activa de análisis de casos para desarrollar experiencias que permitan incorporar los elementos teórico-prácticos asociados a los resultados de aprendizaje de la asignatura.
2. Las experiencias de cátedra/laboratorio/taller deben ser realizadas por medio de la utilización de software moderno aplicable a la asignatura.
3. Se recomienda que las y los estudiantes realicen presentaciones periódicas sobre el trabajo realizado que incluya: contextualización, desarrollo y conclusiones.
4. Actividades prácticas recomendadas: cápsulas teóricas, reuniones de trabajo, taller de trabajo en equipo y liderazgo, presentaciones e informes escritos de avance en español, revisión del estado del arte asociado al problema, lluvia de ideas, análisis de alternativas y descripción detallada de la solución.



VII. ORIENTACIONES Y CRITERIOS PARA LA EVALUACIÓN

1. Se recomienda la aplicación de una evaluación diagnóstica al inicio de la asignatura.
2. La asignatura podría contemplar dos instancias de evaluación de los resultados de aprendizaje: cátedra y taller/laboratorio.
 - En el caso de existir, ambas debieran ser aprobadas por separado: el porcentaje de cada una de ellas deberá ser de 60% para cátedra y 40% para taller/laboratorio.
 - En el caso que la asignatura tenga actividades de taller/laboratorio, éstas deben ser realizadas en grupos de estudiantes y se recomienda la elaboración por parte de los estudiantes de un informe sobre la actividad desarrollada.
3. Se evaluará el conocimiento conceptual y procedimental mediante el desarrollo de al menos dos pruebas sumativas de carácter presencial.
 - Se recomienda además la aplicación de una evaluación mediante la entrega de un trabajo desarrollado en las horas indirectas asociadas a la asignatura.
 - Se recomienda que las y los estudiantes realicen una o más presentaciones de los trabajos realizados, la evaluación de la misma debe ser por medio de la aplicación de una rúbrica.
4. Se recomienda realizar evaluaciones de carácter formativo. Esto permite al docente introducir correcciones, añadir alternativas y reforzar los aspectos para ayudar al estudiantado en el logro de sus habilidades.
5. La asistencia y condiciones de aprobación de la asignatura debe ser acorde a la aplicación del Reglamento de Docencia de Pregrado.



IX. RECURSOS BIBLIOGRÁFICOS

Bibliografía Mínima

- Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, Wes McKinney, 3rd Edition, O'Reilly Media, 2022
- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 3rd Edition, Aurélien Géron, O'Reilly Media, 2022
- DATA SCIENCE with KNIME: Data Exploration, Machine Learning and Visualization using CO-DELESS Visual Programming, Nickholas Anting, NASPSOFT Intelligence Academy, 2022

Bibliografía Complementaria

- An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani, 7 ma edición, 2017, Springer, disponible en <http://www-bcf.usc.edu/gareth/ISL/>
- Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems), Ian H. Witten, Eibe Frank, et al., 2016
- Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, Foster Provost & Tom Fawcett, 2013, O'Reilly Media.
- The Elements of Statistical Learning: data mining, inference, and prediction, Trevor Hastie, Robert Tibshirani, Jerome Friedman, 2da edición, 2017, Springer, disponible en <https://hastie.su.domains/ElemStatLearn/>
- Fundamentals of Data Visualization, Claus Wilke, 1ra edición, 2020, O'Reilly, disponible en <https://clauswilke.com/dataviz/index.html>