

PROGRAMA DE CURSO

Código	Nombre		
IN7781	Taller de Análisis de Datos y Aprendizaje de Máquinas		
Nombre en Inglés			
<i>Data Analysis and Machine Learning Workshop</i>			
SCT	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
6	3	1,5	1,5
Requisitos		Carácter del Curso	
--		Electivo MBE (Área TI)	
Competencias a las que tributa el curso			
CE2	Concebir y diseñar soluciones que crean valor para resolver problemas que surgen en las organizaciones, utilizando los conocimientos provenientes de la gestión de operaciones, tecnologías de información y comunicaciones, finanzas, economía y marketing.		
CE4	Emplear y aplicar los conocimientos de las distintas disciplinas constitutivas de la ingeniería industrial: gestión de operaciones, tecnologías de información y comunicaciones, finanzas, economía y marketing, en las respectivas áreas funcionales de las organizaciones.		
CE6	Crear oportunidades de negocio a través de un emprendimiento.		
CG1	Comunicación profesional y académica. Comunicar en español de forma estratégica, clara y eficaz, tanto en modalidad oral como escrita, puntos de vista, propuestas de proyectos y resultados de investigación fundamentados, en situaciones de comunicación compleja, en ambientes sociales, académicos y profesionales.		
CG6	Innovación. Concebir ideas viables y novedosas para resolver problemas o necesidades, materializadas en productos, servicios o en mejoras a procesos, considerando el contexto sociocultural, económico y los beneficios para el usuario.		
CG7	Emprendimiento. Identificar oportunidades aprovechables para crear e introducir nuevos servicios o productos con valor económico y social, a partir de la toma de decisiones en un contexto complejo de incertidumbre, demostrando compromiso e iniciativa en su quehacer.		

Propósito del curso

El curso *Taller de Análisis de Datos y Aprendizaje de Máquinas* tiene como propósito que el estudiante utilice herramientas y metodologías de Ciencia de Datos aplicada a un problema o necesidad en la industria (que cumpla con los requisitos para ser abordado bajo este enfoque) a fin de identificarlo, comprenderlo y proponer un objetivo en base a una posible solución (hipótesis), seleccionar y/o transformar variables, modelarlo y evaluar los resultados obtenidos.

Este curso está orientado a tener un acercamiento teórico y práctico a los distintos conceptos y herramientas utilizados a lo largo de las etapas principales de un proyecto de Ciencia de Datos.

Se espera que el alumno ya este familiarizado con cursos de Estadística básicos. Adicional a esto, se requiere un nivel intermedio/avanzado en lenguaje de programación Python 3.

Resultados de Aprendizaje	Competencia a la que tributa (CE-CG)
RA1: Identifica un problema de Ciencia de Datos, aplicando metodologías de análisis (KDD o CRISP - DM), a fin de proponer una solución a un problema/necesidad específica en una organización, seleccionar los datos propios del problema y las variables que los caractericen y detectar patrones subyacentes en estos.	CE2-CG6
RA2: Utiliza herramientas que habilitan el análisis con sus respectivas metodologías, pre-procesando datos, aplicando algoritmos complejos y visualizando resultados, en el contexto de su aplicación a un problema/necesidad.	CE4
RA3: Modela un problema de Ciencia de Datos, considerando una hipótesis inicial, variables, uno o más algoritmos y sus <i>hiperparámetros</i> , métricas de desempeño adecuadas y metodologías de visualización, a fin de proponer mejoras a las necesidades de una empresa a partir de las conclusiones obtenidas, cuyos resultados explica de manera clara y coherente.	CE4-CG1
RA4: Determina la coherencia entre los resultados aportados por el modelo y la propuesta de solución a un problema de negocio, contrastando dichos resultados, a fin de evaluar y determinar puntos de mejora a la propuesta.	CE4-CG7

Metodología Docente	Evaluación General
Cátedras y auxiliares cada semana	Tareas (1 por unidad)

Unidades Temáticas

Número	RA al que tributa	Nombre de la Unidad	Duración en Semanas
1	RA1	Fundamentos de Ciencia de Datos e Identificación del Problema/Necesidad	2
Contenidos			Referencias a la Bibliografía
1.1. Conceptos básicos. 1.2. Procesos básicos: KDD y CRISP DM. 1.3. Proceso General 1.4. Posibles Problemas y Fuentes de Datos. 1.5. Planteamiento del problema a abordar. 1.6. Estableciendo una posible solución. 1.7. Objetivo y alcances del proyecto. 1.8. Medición de Impacto.			[1], [10]

Número	RA al que tributa	Nombre de la Unidad	Duración en Semanas
2	RA1, RA2	Análisis Exploratorio de Datos	2
Contenidos			Referencias a la Bibliografía
2.1. Datos y Atributos 2.2. Conjuntos de datos, características y tipos 2.3. Análisis Exploratorio de Datos 2.4. Métricas de Tendencia Central 2.5. Métricas de Dispersión y Gráficos Descriptivos 2.6. Métricas de relación entre variables 2.7. Teoría del Muestreo e Inferencia Estadística 2.8. Teoría Estadística de la Decisión 2.9. Visualización de relación entre variables 2.10. Métricas de similitud			[1], [10], [2], [3], [4]

Número	RA al que tributa	Nombre de la Unidad	Duración en Semanas
3	RA2, RA3	Ingeniería de Atributos	2
Contenidos			Referencias a la Bibliografía
3.1. Calidad de los datos 3.2. Ingeniería de Atributos – Definición y Desafíos 3.3. Vector de Atributos y Espacio de Atributos 3.4. Ingeniería de Atributos – Flujo de Trabajo 3.5. Limpieza de Atributos 3.6. Valores perdidos 3.7. Outliers y Ruido 3.8. Valores Duplicados e Inconsistencia de Datos 3.9. Diseño de Atributos y Transformaciones 3.10. Agregación, Normalización, Discretización, Muestreo 3.11. Reducción de Atributos: PCA, LDA, Kernel PCA, MDS			[1], [2], [3], [4]

Número	RA al que tributa	Nombre de la Unidad	Duración en Semanas
4	RA2, RA3, RA4	Introducción a Aprendizaje de Máquinas y Evaluación de Resultados	1
Contenidos			Referencias a la Bibliografía
4.1. Aprendizaje y Generalización – Conceptos Fundamentales 4.2. Aprendizaje Supervisado 4.3. Aprendizaje No Supervisado 4.4. Marco general para Regresión 4.5. Marco general para Clasificación 4.6. Evaluación de modelos de Regresión 4.7. Evaluación y Validación de modelos de Clasificación			[1], [2], [3], [4], [5], [6], [7]

Número	RA al que tributa	Nombre de la Unidad	Duración en Semanas
5	RA2, RA3	Aprendizaje Supervisado	2
Contenidos			Referencias a la Bibliografía
5.1. Modelos de regresión 5.2. Modelos de clasificación 5.3. Árboles de Decisión 5.4. Clasificadores Bayesianos 5.5. Support Vector Machines 5.6. Métodos Ensamblados o Conjuntos			[1], [3], [4], [8], [9], [11], [12]

Número	RA al que tributa	Nombre de la Unidad	Duración en Semanas
6	RA2, RA3	Aprendizaje No Supervisado	2
Contenidos			Referencias a la Bibliografía
6.1. Análisis de Clustering 6.2. Clustering Particional 6.3. Clustering Jerárquico 6.4. Clustering Aglomerativo 6.5. Validación de Clusters			[1], [3], [4], [13]

Bibliografía General

Bibliografía obligatoria

- [1] Velásquez, J. D. & Hernández, V. (2024) Diapositivas de curso “IN4151 Ingeniería de la Información” Capítulo 3, curso obligatorio de Ingeniería Civil Industrial, U. de Chile.
- [2] Igual, L., & Seguí, S. (2017). Introduction to Data Science. Springer, Cham.
- [3] Tan, P.-N., Steinbach, M., Karpadne, A. and Kumar, V. (2019). Introduction to Data Mining, Second Edition, Global Edition. Pearson Education, Inc., London.
- [4] Han, Kamber, Pei (2012). Data Mining: Concepts and Techniques (3rd Edition)
- [5] Ian Goodfellow and Yoshua Bengio and Aaron Courville (2016). Deep Learning. MIT Press.
- [6] Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. O’Reilly Media, Inc.
- [7] Kubat, M. (2021). An introduction to Machine Learning, Third Edition. Cham, Switzerland: Springer International Publishing.
- [8] Dietterich, T. (2000). Ensemble Methods in Machine Learning.
- [9] Mengye Ren and Matthew MacKay (2018), Lecture slides from CSC411: Introduction to Machine Learning, University of Toronto.
- [10] Provost & Fawcett (2013). Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking.
- [11] Dorner, M. (2013) Lecture slides from Artificial Intelligence course.
<https://github.com/michaeldorf/DecisionTrees>
- [12] Abati, D. and Porrello, A. (2018) Support Vector Machines lecture slides.
https://github.com/ndrplz/machine_learning_lectures
- [13] Steorts, R. (2015) Hierarchical clustering lecture slides.
www2.stat.duke.edu/~rsc46/predict.html

Bibliografía complementaria

- [13] Witten, Frank, Hall, Pal (2017). Data Mining: Practical Machine Learning Tools and Techniques (4th Edition).
- [14] Flach (2012). Machine Learning: The Art and Science of Algorithms that Make Sense of Data.
- [16] Bruce, Bruce, Gedeck (2017). Practical Statistics for Data Scientists (2nd Edition).
- [17] George (2021). Practical Data Science with Python.

Vigencia desde:	2025
Elaborado por:	Víctor Hernández M.
Validado por:	Comité de docencia
Revisado por:	Área de Gestión Curricular