



Syllabus Fundamentos de Ciencias de Datos (TICS-314)

Unidad académica	Pregrado		
Carrera o programa	Ingeniería Civil Informática		
Año	2021	Semestre	2
Profesor	Gonzalo Anriquez G	Email	gonzalo.anriquez@edu.uai.cl
		Horario de atención	Agendar por email.
Ayudante	Juan Carlos Urrutia	Email	juurrutia@alumnos.uai.cl
Créditos SCT-Chile	6	Total horas	180
Horas de Docencia Directa		Horas de Trabajo Autónomo	
Cátedra	Laboratorio	Ayudantía	
45	0	22,5	112,5
Tipo de Asignatura	Curso obligatorio de Ingeniería Civil Informática		
Área	Informática		
Pre-requisitos	Probabilidad y Estadística (MAT106)		
Descripción de la Asignatura	<p>En la actualidad, las empresas colectan y administran una gran cantidad de datos, y la forma de procesarlos está logrando ventajas competitivas y estratégicas en la toma de decisiones. Esto impacta directamente tareas tales como el razonamiento bajo incertidumbre, identificación de patrones de comportamiento, predicción de fenómenos, identificación de factores más determinantes en un negocio, detección de tendencias, segmentación de grupos, etc.</p> <p>Para abordar las tareas previas, se requiere realizar un proceso de análisis de los datos, el cual en su primera etapa corresponde al análisis exploratorio de datos. Este análisis busca entender y maximizar el conocimiento de los datos, visualizar la estructura de los datos (distribuciones), identificar variables importantes dentro de los datos, detectar anomalías y outliers, verificar suposiciones iniciales, y más importante, poder generar hipótesis desde los datos.</p>		
Competencias del Egresado	(1) Aplicar conocimiento de matemáticas, ciencia e ingeniería. (2) Identificar, formular y resolver problemas de ingeniería. (3) Usar técnicas, habilidades y herramientas modernas de ingeniería necesarias para la práctica ingenieril.		
Resultados de Aprendizaje	(1) Comprender y aplicar los objetos y conceptos de análisis exploratorio de datos que serán abordados durante el curso. (2) Dominar las técnicas y los principales tratamientos que se utilizan en estas áreas para el análisis de los datos. (3) analizar datos utilizando los métodos del análisis exploratorio de datos.		
Estrategias de Enseñanza y Aprendizaje			
El curso considera clases lectivas (cátedra), prácticas y ayudantías de apoyo.			
Las cátedras están orientadas a discutir las diferentes ideas y conceptos del curso, incorporando en lo posible la participación de los alumnos a través de preguntas, permitiendo interacciones conducentes a una comprensión efectiva. Se utilizarán ejemplos ilustrativos de los temas y se aplicará los códigos correspondientes para la generación de los análisis respectivos.			

Las clases prácticas buscan que el estudiante adquiera habilidades de análisis de datos en base a ejercicios y ambientes controlados. Con esto el estudiante trabajará en grupo en base a ejercicios dirigidos y con mentoría del profesor.

La ayudantía se centra en el desarrollo de habilidades para el tratamiento de los problemas del curso y en sus aplicaciones concretas.

Procedimientos de Evaluación de Aprendizajes

El curso será evaluado a través de dos actividades principales: (a) Trabajos Grupales (TG), donde grupos de 4 alumnos deberán trabajar en torno a un proyecto que se les asigne; y (b) Tareas (T), correspondiente a trabajo personal. Se consideran dos trabajos grupales (TG1 y TG2) y cuatro tareas (T1, T2, T3 y T4), cuyo promedio corresponderá a T.

El objetivo principal de los trabajos grupales es de certificar el nivel de aprendizaje del estudiante. Mientras que las tareas son para reforzar la materia y prepararlos para las entregas finales.

La nota final (NF) está definida por $NF = 0,35 \cdot TG1 + 0,35 \cdot TG2 + 0,30 \cdot T$.

Unidades de la Asignatura (Contenidos)

1. Introducción
2. Preparación de los datos
 - 2.1. ¿Qué son los datos?
 - 2.2. Manejo de datos tabulares
 - 2.3. Limpieza de datos
 - 2.4. Reducción de datos y dimensionalidad
3. Análisis exploratorio de los datos
 - 3.1. Resumen de datos
 - 3.2. Visualización de datos
 - 3.3. Detección de Outliers
 - 3.4. Técnicas cuantitativas
 - 3.5. Modelos de regresión
4. Introducción a técnicas de aprendizaje automático

Reglamento

Cualquier situación especial que no esté contenida en este reglamento deberá ser conocida y resuelta por la Secretaría de Pregrado. Si un alumno excede las ausencias a evaluaciones (entrega de informes) previstas en este reglamento, deberá presentar su situación debidamente justificada ante la Secretaría de Pregrado, donde se decidirá el procedimiento extraordinario a seguir.

El profesor estará disponible para responder preguntas de sus alumnos en el horario indicado al inicio de este Syllabus.

La copia o plagio en tareas, controles, pruebas o examen se considera una falta grave y será sancionada drásticamente.

Atención: debido a la pandemia del coronavirus, este syllabus puede ser modificado para adaptarse a las condiciones pertinentes.

Recursos para el Aprendizaje (Bibliografía)

- a) Cleveland, W. S. (1993), Visualizing Data, New Jersey: Hobart Press.
- b) ____ (1994), The Elements of Graphing Data, Revised edition, New Jersey: Hobart Press.
- c) Nussbaumer C. (2015), Storytelling with data, a data visualization guide for business professionals, Wiley.
- d) Robbins, N. (2005), Creating more Effective Graphs, Wiley.
- e) Wickham, H. (2016), ggplot2: Elegant Graphics for Data Analysis, 2nd Edition, Springer.
- f) Winston, C. (2013), R Graphics Cookbook, O'REILLY.

Semana	Lunes de la semana	Tópicos	Evaluación
1	02-08-2021	Introducción al curso	
		Preparación de datos: ¿Qué son los datos?	
2	09-08-2021	Manejo de datos tabulares con Pandas	
3	16-08-2021	Preparación de datos: Limpieza de datos	Entrega Tarea 1: Análisis de datos
		Análisis exploratorio de datos: Resumen de datos	
4	23-08-2021	Análisis exploratorio de datos: Visualización de datos	
		Manipulación de los elementos existentes de un gráfico básico	
5	30-08-2021	Manipulación de los elementos existentes de un gráfico básico	
6	06-09-2021	Estadística Básica para Data Science	Entrega Tarea 2: Graficos con código
		Distribuciones estadísticas	
7	13-09-2021	Trabajo de la prueba en grupos	
8	20-09-2021	Trabajo de la prueba en grupos	Entrega Trabajo Grupal 1, preguntas y respuestas
9	27-09-2021	Introducción a Machine Learning: regresión lineal	
10	04-10-2021	Introducción a Machine Learning: clasificación y medidas de desempeño	
11	11-10-2021	SEMANA DE PAUSA	
		Cierre Introducción a Machine Learning	Entrega Tarea 3: Regresión Lineal
12	18-10-2021	Multidimensionalidad	
		PCA	
13	25-10-2021	Ejemplo práctico PCA	
14	01-11-2021	Outliers	
15	08-11-2021	Flujo de un proyecto de <i>Data Science</i>	Entrega Tarea 4: PCA
16	15-11-2021	Trabajo de la prueba en grupos	
17	22-11-2021	Trabajo de la prueba en grupos	Entrega Trabajo Grupal 2, preguntas y respuestas
18	Período de Exámenes Finales Segundo Semestre (29 NOV- 13 DICIEMBRE; Derecho comienza 26 noviembre)		