

Syllabus Fundamentos de Ciencias de Datos (TICS-314)

Unidad académica	Pregrado		
Carrera o programa	Ingeniería Civil Informática		
Año	2020	Semestre	1
Profesor	Sebastián Moreno	Email	Sebastian.moreno@uai.cl
		Horario de atención	Solicitar por e-mail
Ayudante	Por definir	Email	Por definir
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 coleccionan y administran una gran cantidad de datos, y la forma de procesar los datos está logrando ventajas competitivas y estratégicas en la toma de decisiones en cualquier organización. 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

En la Cátedra las clases estarán orientadas a discutir las diferentes ideas y conceptos del curso incorporando en lo posible la participación de los alumnos a través de preguntas que permitan interacciones en clase 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.

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

Se realizarán dos pruebas de cátedra (P1 y P2) en horario especial programado para estos fines, siendo la segunda prueba en el periodo de exámenes. El objetivo principal de las pruebas es de certificar el nivel de aprendizaje del estudiante. Además se realizará un proyecto acumulativo para evaluar la parte práctica de la asignatura (PR).

En caso que el alumno tenga promedio de pruebas (PP) mayor a 3,5 y su nota de proyecto sea mayor a 3,5, entonces la nota final (NF) está definida por $NF = 0,3*P1 + 0,3*P2 + 0,4*PR = 0,6*PP + 0,4*PR$. Caso contrario, será la mínima nota entre el promedio de pruebas y la nota de proyecto.

$$NF = \begin{cases} 0,6*PP+0,4*PR & \text{si } PP \geq 3.5 \text{ y } PR \geq 3.5 \\ \min(PP, PR) & \text{en cualquier otro caso} \end{cases}$$

Unidades de la Asignatura (Contenidos)

1. Introducción
2. Preparación de los datos
 - 2.1. ¿Qué son los datos?
 - 2.2. Limpieza de datos
 - 2.3. 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. Reglas de asociación

Reglamento

Si el alumno faltó a una prueba de cátedra, la ausencia debe ser justificada ante pregrado, que autorizará a que la nota de la prueba sea reemplazada por la nota del examen recuperativo.

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 (pruebas de cátedra) 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 controles, pruebas o examen se considera una falta grave y será sancionada drásticamente.

Para asegurar buenas condiciones de desarrollo de clases, por su efecto altamente perjudicial para la concentración, no se admite el uso de teléfonos celulares, tabletas, computadores, etc. en las sesiones de cátedra. Por la misma razón, queda excluido el consumo de alimentos, salidas tempranas y llegadas tardías (más de 5 minutos).

Durante las clases y en las evaluaciones se recomienda el uso de un computador.

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

Recursos para el Aprendizaje (Bibliografía)

Cleveland, W. S. (1993), *Visualizing Data*, New Jersey: Hobart Press.

— (1994), *The Elements of Graphing Data*, Revised edition, New Jersey: Hobart Press.

Nussbaumer C. (2015), *Storytelling with data, a data visualization guide for business professionals*, Wiley.

Robbins, N. (2005), *Creating more Effective Graphs*, Wiley.

Wickham, H. (2016), *ggplot2: Elegant Graphics for Data Analysis*, 2nd Edition, Springer.

Winston, C. (2013), *R Graphics Cookbook*, O'REILLY.

--

Semana	Contenidos	Tópicos
1	Introducción al curso Preparación de datos: ¿Qué son los datos?	Tipos de variables y datos existentes
2	Preparación de datos: Limpieza de datos Preparación de datos: Reducción de datos y dimensionalidad	Práctica con código para limpiar y reducir la dimensionalidad de los datos.
3	Análisis exploratorio de datos: Resumen de datos Análisis exploratorio de datos: Visualización de datos	Introducción a una imagen y librería a usar
4	Análisis exploratorio de datos: Visualización de datos	Manipulación de los elementos existentes de un gráfico básico
5	Análisis exploratorio de datos: Visualización de datos	Manipulación de los elementos existentes de un gráfico básico
6	Análisis exploratorio de datos: Visualización de datos	Gráficos de distribuciones
7	Análisis exploratorio de datos: Visualización de datos	Gráficos de distribuciones
8	Prueba 1	
9	Análisis exploratorio de datos: Visualización de datos	Análisis de la regresión lineal
10	Análisis exploratorio de datos: Visualización de datos	Gráfico de dispersión y relaciones entre variables
11	Análisis exploratorio de datos: Visualización de datos	Multidimensionalidad
12	Análisis exploratorio de datos: Análisis de Outliers	Métodos estadísticos para la detección de outliers
13	Análisis exploratorio de datos: Análisis de Outliers	Métodos basado en distancia para la detección de outliers
14	Análisis exploratorio de datos: Técnicas cuantitativas	Introducción a los tests de hipótesis
15	Análisis exploratorio de datos: Técnicas cuantitativas	Tests estadísticos para analizar efectos en los datos
16	Reglas de asociación	Algoritmos y códigos para aplicar reglas de asociación.