

Programa del Curso

Clase 1





AGENDA DE HOY

01

Presentaciones

02

Programa del Curso

03

Introducción a Ciencia
de Datos

04

Herramientas Prácticas
de Programación

+ BREAK 20:00 (10-15')



01

Presentaciones

Equipo Docente

Profesora

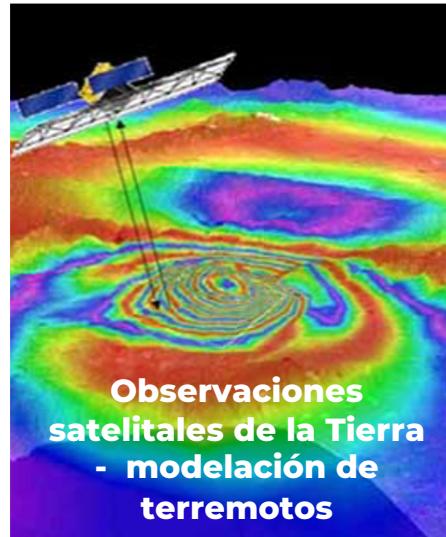
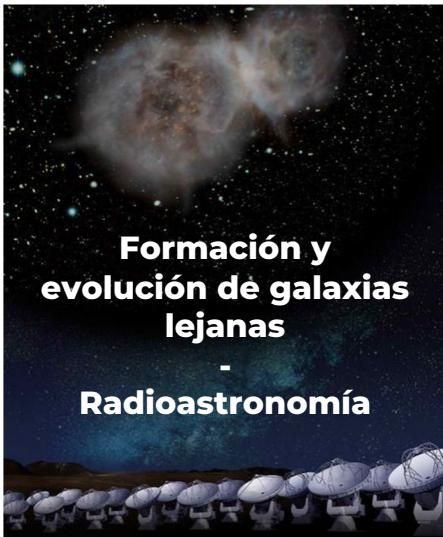


Paula Aguirre Aparicio - paaguirr@ing.puc.cl

Ing. Civil Mecánico / PhD en Astrofísica

Profesora Instituto Ingeniería Matemática y Computacional UC
Jefe de Programa Licenciatura en Ingeniería en Ciencia de Datos

Investigadora del Centro de Investigación para la Gestión Integrada del Riesgo de Desastres (CIGIDEN, www.cigiden.cl)



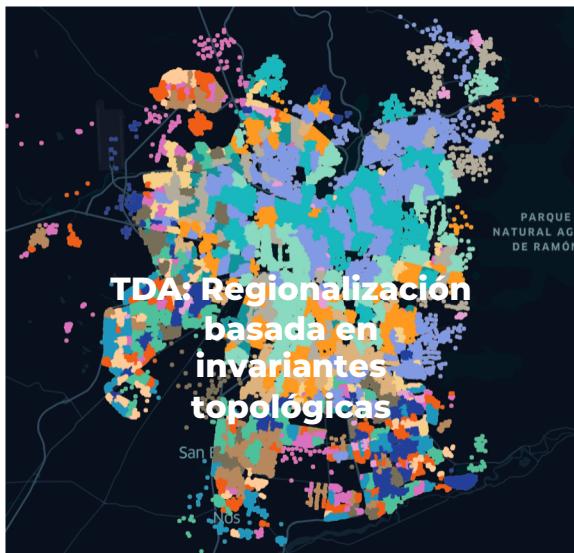
Equipo Docente

Ayudante



Felipe Gutiérrez- figutier@uc.cl

- Planificador Urbano
- Magister en Desarrollo Urbano y en Procesamiento y Gestión de información
- Alumno de Doctorado en Ciencias de la Ingeniería

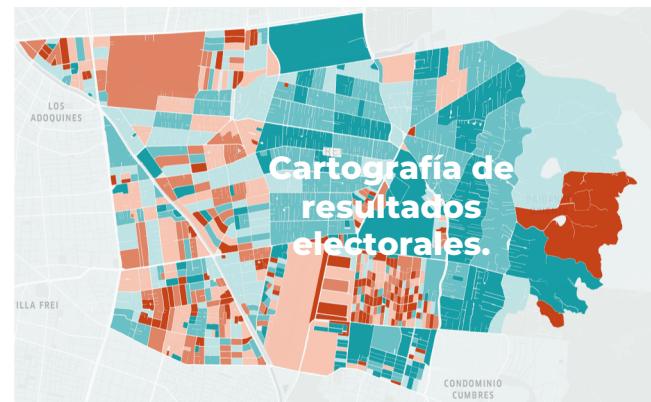


Inédito plan piloto en Ñuble: Monitorean aguas servidas para detección temprana del Covid-19

Publicado: Viernes, 26 de Junio de 2020 a las 14:52hrs. Autor: Cooperativa.cl

- La herramienta permite obtener información anticipada en grupos sociales.
- Según estudios europeos, test de heces fecales permite detectar el virus al tercer día del contagio.

Recuerda nuestra campaña ["Todos podemos hacer más"](#), si tienes una iniciativa solidaria que quieras comunicar, llámanos al 223548000 o visita nuestro [sitio](#) si quieres hacer tu aporte





02

Programa del Curso

Descripción del Curso y Objetivos

Este curso **introduce** a los alumnos al campo de la Ciencia de Datos.

Estudiaremos en **amplitud**, más que en profundidad, sus **principios básicos** y las **principales técnicas y herramientas** utilizadas:

- ❖ Recolección, integración, transformación y almacenamiento de datos.
- ❖ Inferencia estadística.
- ❖ Análisis exploratorio de datos.
- ❖ Análisis descriptivo y predictivo.
- ❖ Creación de productos de información.
- ❖ Visualización de datos.
- ❖ Herramientas prácticas (Python, Pandas, entre otros).
- ❖ Problemas éticos y de privacidad emergentes en Ciencia de Datos.

Descripción del Curso y Objetivos

Este curso **introduce** a los alumnos al campo de la Ciencia de Datos.

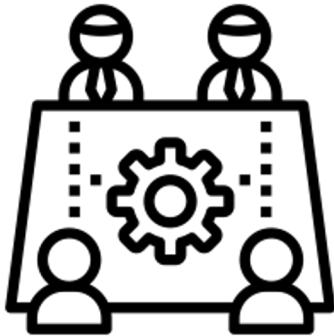
Estudiaremos en **amplitud**, más que en profundidad, sus **principios básicos** y las **principales técnicas y herramientas** utilizadas:

Bimestre 2	Bimestre 3	Bimestre 4	Bimestre 5	Bimestre 6	Bimestre 7	Bimestre 8
<ul style="list-style-type: none">• EPG4001 Aprendizaje Supervisado• EPG4002 Aprendizaje no Supervisado• FIL2000 Ética aplicada a Inteligencia Artificial	<ul style="list-style-type: none">• EPG4003 Métodos Probabilísticos• IMT3870 Computación de Alto Rendimiento• INF3812 Deep Learning	<ul style="list-style-type: none">• INF3813 Deep Learning Avanzado <p>Optativos</p> <ul style="list-style-type: none">• INF3822 Proyecto Aplicado I o INF3250 Seguridad Computacional o EPG4008 Análisis Estadístico con Datos Faltantes• INF3803 Minería de Procesos o INF3863 Sistemas Recomendadores o EPG4005 Métodos Bayesianos	<ul style="list-style-type: none">• INF3862 Plataformas de Big Data <p>Optativos</p> <ul style="list-style-type: none">• INF3823 Proyecto Aplicado II o INF3260 Procesos de Negocio o EPG4007 Análisis de Tiempos de Falla• INF3801 Minería de Medios Sociales o EPG4004 Series de Tiempo o EPG4006 Modelos Lineales Generalizados	<ul style="list-style-type: none">• INF3841 Recuperación de Información• INF3842 Visualización de Información <p>Optativos</p> <ul style="list-style-type: none">• INF3821 Aplicaciones o INF3570 Aspectos Legales de las TI o EPG4009 Análisis Multivariado	<ul style="list-style-type: none">• INF3851 Gobernanza de Datos• INF4980 Actividad de Graduación <p>Optativos</p> <ul style="list-style-type: none">• INF3852 Gestión de Proyectos o INF3280 Gestión Estratégica o EPG4010 Métodos Bayesianos Avanzados	<ul style="list-style-type: none">• INF4980 Actividad de Graduación
Bimestre 1						

Estrategias Metodológicas

- ❖ Clases Expositivas.
- ❖ Talleres Prácticos.
- ❖ Lecturas y actividades de discusión.

Estrategias Evaluativas



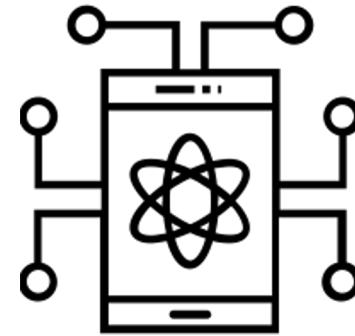
Talleres prácticos

- 30%
- Individuales
- 4 Talleres prácticos durante el curso
- Trabajo en clase y entrega dentro de la semana siguiente.
- + Taller 5: participación en discusiones y actividades
0% → 1.0, 50% → 4.0, 70% → 7.0



Controles

- 30%
- Individuales
- 2 durante el curso
- Formato online, en horario de clase.



Proyecto de Aplicación

- 40%
- Grupal (3 personas)
- Avances + presentación + reporte.

Política de atrasos: cada estudiante tiene derecho a 24 h de atraso en total durante el curso, sin expresión de causa ni penalización. Luego, se descuenta 0.5 pto/hora. En el caso de entregas grupales, el atraso se contabiliza para todos los integrantes. Para atrasos con causa justificada, comunicarse con la profesora.

Calendario de Evaluaciones

	L	M	W	J	V	S	D
Marzo	22	23	24	25 Clase 1	26	27	28
	29	30	31 Clase 2 Taller 1	1	2	3	4
Abril	5	6	7	8 Clase 3 Control 1	9	10	11
	12	13	14	15 Clase 4 Taller 2	16	17	18
Mayo	19	20	21	22 Clase 5 Control 2	23	24	25
	26	27	28	29 Clase 6 Taller 3	30	1	2
Junio	3	4	5	6 Clase 7	7	8	9
	10	11	12	13	14	15	16
Receso universitario							
	17	18	19	20 Clase 8 Taller 4	21	22	23
	24	25	26	27 Presentaciones proyecto	28	29	30
Junio	31	1	2	4	5	6	7
					Entrega reporte Proyecto		

Proyecto de Aplicación

- ❖ **Objetivo:** desarrollar todo el proceso de ciencia de datos para responder preguntas de interés y accionables en un ámbito a elección.

- ❖ **Componentes del proyecto:**
 - Planteamiento del problema.
 - Recolección de datos.
 - Limpieza y de datos.
 - Análisis exploratorio.
 - Preparación de datos.
 - Modelamiento y análisis descriptivo/predictivo.
 - Comunicación y reporte de resultados.

- ❖ **Formato:**
 - Avances parciales sin evaluación.
 - Presentación (27 de mayo 2021)
 - Reporte escrito + códigos (4 de junio 2021).

- ❖ El enunciado detallado del proyecto se presentará y discutirá en la próxima clase.

Bibliografía

Mínima:

- “*Doing Data Science, Straight Talk from the Frontline*”.
Cathy O’Neil and Rachel Schutt, O'Reilly Media, 2013.
- “*Data Science for Business: What you need to know about data mining and data-analytic thinking*”.
Foster Provost and Tom Fawcett. O'Reilly Media, 2013.

Complementaria:

- “*Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*”.
Wes McKinney, O'Reilly Media, 2017
- Blogs y artículos de revistas científicas, prensa, etc.



03

Aspectos Prácticos

Comunicación y Herramientas Computacionales

- ❖ Todos los contenidos, anuncios y material del curso se canalizará a través de Canvas y correo electrónico (al menos hasta que todos tengan acceso a Canvas)
- ❖ Utilizaremos software de código abierto y lenguaje de programación en Python.
- ❖ Todo tipo de dudas, preguntas y comentarios bienvenidos!!! Utilizar correo electrónico, foros de Canvas, preguntar en clases, etc.

¿Preguntas?