

### Programa de curso

Unidad Académica	:Centro de Informática Médica Y Telemedicina Programa de Anatomía y Biología del Desarrollo Centro de Informática Médica Y Telemedicina Programa de Anatomía y Biología del Desarrollo
Nombre del curso	:Métodos de Análisis Estadístico II
Nombre en inglés del curso	:Methods for Statistical Analysis II
Idioma en que se dicta	:Español
Código ucampus	:CAMAEL
Versión	:v. 1
Modalidad	:A distancia
Semestre	:1
Año	:2023
Días/Horario	:Lun 18:00-21:20, Jue 18:00-21:20, Vier 18:00-21:20, Mar 18:00-21:20, Mie 18:00-21:20,
Fecha inicio	:05/06/2023
Fecha de término	:11/07/2023
Lugar	:
Cupos mínimos	:3
Cupos máximo	:15
Créditos	:3

#### Tipo de curso

AVANZADO

#### Datos de contacto

Nombre	: Mauricio Cerda
Teléfono	: +56985927543
Email	: mauricio.cerda@uchile.cl
Anexo	: 89641

#### Horas cronológicas

Presenciales:	: 0
A distancia:	: 92

#### Tipos de actividades(Horas directas estudiante)

Clases(horas)	: 30
Seminarios (horas):	: 0
Evaluaciones (horas)	: 0
taller/trabajo práctico	: 0
Trabajo/proyecto	: 2
investigación:	: 2
Créditos	: 3

**PROFESOR ENCARGADO/A DEL CURSO (PEC)**

Cerde Villablanca Mauricio

Docente Participantes	Unidad Academica	Función	Horas directas.	Horas indirectas.	Horas totales
Fabian Villena	Centro de Informática Médica Y Telemedicina	Profesor Participante	19	57	76
Pablo Baez	Centro de Informática Médica Y Telemedicina	Profesor Participante	3	9	12

**Fundamentos, Antecedentes que justifican la necesidad de dictar el curso**

Una importante proporción de la ficha clínica se encuentra en formato de texto libre, lo cual dificulta la extracción de información clave y el uso secundario de datos. El texto clínico tiene desafíos particulares debido al uso extensivo de abreviaciones, la presencia de negación o la variabilidad del lenguaje clínico entre especialidades médicas. Por otro lado, el procesamiento del lenguaje natural (PLN) es una rama de la inteligencia artificial que permite la comunicación entre humanos y máquinas a través del lenguaje. Este curso es una introducción al PLN clínico.

**Destinatarios**

Estudiantes del programa de magister en informática médica. También esta abierto a recibir estudiantes de otros programas de magister o doctorado que deseen una formación básica en computación.

**Requisitos**

Cursos del Primer y Segundo Semestre del Magister en Informática Medica o Conocimientos a nivel de pregrado en biología, microscopía de fluorescencia, matemática aplicada, o computación.

**Resultado de aprendizaje**

- Poder preprocesar, visualizar, y hacer clasificación básica de textos clínicos
  - Entender la importancia de crear recursos lingüísticos para analizar textos clínicos producidos en Chile y
  - Comprender de donde emerge el uso de redes neuronales para el análisis de texto
  - Comprender y calcular word embeddings
- cuáles son las técnicas para ello problemas clásicos así como los desafíos actuales

**Metodologías de enseñanza y aprendizaje**

Clase teórica

Cantidad

30

**Metodologías de evaluación**

Cantidad

Duración horas

Ponderación

Informe, trabajo o proyecto de investigación

1

2

100.0 %

**Suma (Para nota presentación examen)**

100.0 %

**Total %**

%

**Requisitos de aprobación y asistencia.**Nota Informe, trabajo o proyecto de investigación  $\geq 4.0$

## Unidades

### Unidad: Preprocesamiento de texto

Encargado: Fabian Villena

#### Logros parciales de aprendizajes:

- Conocer y manejar las herramientas básicas del preprocesamiento de texto
- Poder realizar visualizaciones básicas de texto
- Aprender a usar Python para el preprocesamiento de texto

#### Acciones Asociadas:

- Clases expositivas
- Sesiones prácticas de programación

#### Contenidos:

- Tokenización, lematización y stemming - Expresiones regulares - Preprocesamiento de texto en Python - Bag-of-words, TF-IDF - Nubes de palabras - Naive Bayes

### Unidad: Anotación de textos clínicos

Encargado: Pablo Baez

#### Logros parciales de aprendizajes:

- Entender la necesidad de crear recursos lingüísticos para el texto clínico
- Aprender cómo se comparan distintos etiquetados
- Conocer a BRAT Rapid Annotation Tool

#### Acciones Asociadas:

- Clases expositivas
- Sesiones prácticas de programación

#### Contenidos:

- Anotado de texto clínico - Introducción a BRAT Rapid Annotation Tool - Representación de anotaciones: Standoff y CoNLL

### Unidad: Redes neuronales para analizar texto

Encargado: Cerda Villablanca Mauricio

#### Logros parciales de aprendizajes:

- Entender el concepto de word embedding y saber como construirlos y testarlos
- Entender por qué las redes recurrentes son usadas para analizar texto
- Aprender a evaluar modelos de deep learning
- Conocer cuales son los problemas actuales en el campo de procesamiento del lenguaje natural en medicina

#### Acciones Asociadas:

- Clases expositivas
- Sesiones prácticas de programación

#### Contenidos:

- Word embedding - Redes neuronales recurrentes - Transformers - Métricas para el análisis de la calidad de los modelos predictivos. - Utilización de motores de procesamientos de lenguaje natural en línea.

Bibliografía							
Caracter	Título	Autor	Edición	Idioma	Formato	Vínculo(Url)	Fecha de consulta
Obligatorio	Clinical Text Mining	H. Dalianis	1ra	Inglés	Libro digital	<a href="https://www.spr...">https://www.spr...</a>	16/01/2021
Complementario	Speech and Language Processing	D. Juraksky y J. H. Martin	3ra	Inglés	Libro digital	<a href="https://doi.org...">https://doi.org...</a>	16/01/2021

Plan de clases					
Fecha	Horario	Actividad	Condición	Tema	Profesor(es)
2023-06-05,Lun	18:00 - 21:20	Clase 1	Libre	Introducción al texto clínico en medicina	Cerda Villablanca Mauricio;Fabian Villena
2023-06-08,Jue	18:00 - 21:20	Clase 2	Libre	Preprocesamiento de texto	Cerda Villablanca Mauricio;Fabian Villena
2023-06-12,Lun	18:00 - 21:20	Clase 3	Libre	Visualización y clasificación de texto	Cerda Villablanca Mauricio;Fabian Villena
2023-06-16,Vier	18:00 - 21:20	Clase 4	Libre	Word embedding	Cerda Villablanca Mauricio;Fabian Villena
2023-06-20,Mar	18:00 - 21:20	Clase 5	Libre	Anotación de textos	Pablo Baez
2023-06-23,Vier	18:00 - 21:20	Clase 6	Libre	Introducción a redes recurrentes	Cerda Villablanca Mauricio;Fabian Villena
2023-06-27,Mar	18:00 - 21:20	Clase 7	Libre	Transformers y métricas de evaluación	Cerda Villablanca Mauricio;Fabian Villena
2023-06-28,Mie	18:00 - 21:20	Clase 8	Libre	Problemas en Chile de texto clínico y consultas de proyectos personales	Cerda Villablanca Mauricio;Fabian Villena
2023-07-03,Lun	18:00 - 21:20	Clase 9	Libre	Desafíos actuales en área de procesamiento del lenguaje natural clínico	Cerda Villablanca Mauricio;Fabian Villena
2023-07-11,Mar	18:00 - 21:20	Examen	Libre	Examen	Cerda Villablanca Mauricio;Fabian Villena