

## **Programa del curso de doctorado: Introducción a la Ciencia de Datos**

Docente: Elmer A. Fernández (PhD)

### **Fundamentación:**

La ciencia de datos ha cobrado una importancia significativa en los últimos años, transformando radicalmente la manera en que las organizaciones abordan la toma de decisiones y la resolución de problemas. Por ejemplo, se destaca su prominencia en el mercado laboral actual, donde las empresas buscan activamente profesionales con la capacidad de analizar datos para extraer información valiosa y tomar decisiones respaldadas por evidencia. Las organizaciones que implementan prácticas de ciencia de datos tienen el potencial de optimizar sus operaciones, logrando una mayor eficiencia, reducción de costos y una ventaja competitiva notable.

Además, promoveremos la interdisciplinariedad, una característica cada vez más esencial en el entorno laboral actual, alentando a los estudiantes a combinar conocimientos de estadística, programación, matemáticas y dominio del tema. Un curso de introducción en ciencia de datos puede empoderar a los estudiantes con un conjunto versátil de habilidades que son aplicables en diversos campos y sectores.

En resumen, consideramos que la ciencia de datos representa una habilidad fundamental en la era de la información. La capacidad de recopilar, analizar y tomar decisiones basadas en datos es esencial en una amplia variedad de ámbitos y sectores. Un curso de "Introducción a la Ciencia de Datos" proporciona a los estudiantes los cimientos necesarios para prosperar en este entorno en constante evolución.

### **Objetivos de aprendizaje:**

**Comprender los fundamentos de la ciencia de datos:** Al finalizar el curso, los estudiantes deben tener una comprensión sólida de qué es la ciencia de datos, su relevancia en diversas industrias y cómo se relaciona con otras disciplinas, como la estadística y la informática.

**Manejo de herramientas y tecnologías:** Los estudiantes deben adquirir competencia en el uso de herramientas y tecnologías comunes en ciencia de datos, como Python, R y Jupyter Notebooks, así como bibliotecas de análisis de datos, bases de datos y sistemas de almacenamiento de datos.

**Manipulación de datos:** Los estudiantes deben ser capaces de adquirir, limpiar, transformar y unificar datos de diversas fuentes para su análisis posterior.

**Análisis exploratorio de datos (EDA):** Los estudiantes deben aprender a realizar un EDA efectivo, incluyendo la generación de estadísticas descriptivas, la creación de visualizaciones y la identificación de patrones y valores atípicos en los datos.

Modelado predictivo básico: Los estudiantes deben comprender los conceptos fundamentales del aprendizaje automático y ser capaces de construir modelos de regresión y clasificación sencillos.

Comunicación de resultados: Los estudiantes deben poder comunicar de manera efectiva los resultados de su análisis de datos a través de informes escritos, presentaciones y visualizaciones.

Aplicación en estudios de caso reales: Los estudiantes deben aplicar sus habilidades a través de estudios de caso prácticos que reflejen situaciones del mundo real y problemas reales de la industria.

Trabajo en equipo y colaboración: Los estudiantes deben tener la capacidad de trabajar en equipos interdisciplinarios y colaborar en proyectos de ciencia de datos.

Desarrollo de habilidades de resolución de problemas: Los estudiantes deben ser capaces de abordar problemas complejos y aplicar técnicas de ciencia de datos para encontrar soluciones efectivas.

### **Evaluación:**

La evaluación será continua durante el cursado, donde se tendrán en cuenta los siguientes criterios: los tiempos de resolución en función de las fechas estipuladas, calidad de la presentación y capacidad de mostrar los resultados al resto de la clase.

La evaluación final consistirá en la resolución de un problema real o de un problema asignado por la cátedra. El mismo consistirá en una base de datos y una situación problemática que deberá ser resuelta en grupos de no más de 3 personas.

Los mismos deberán al finalizar realizar una presentación al frente de todo el curso, presentando su trabajo y defendiendo de preguntas de sus pares y de los profesores.

### **Temas:**

- 1) Conceptos básicos de ciencia de datos  
Descubrimiento de conocimiento en Bases de datos, aplicaciones
- 2) Introducción al entendimiento del negocio/problema  
Conceptos de problemas, relevamiento, recursos y proyectos
- 3) Familiarización de datos  
Tipos de datos, fuentes de datos, tipos de bases de datos
- 4) Preparación de datos  
Generación de la base de datos de trabajo, conceptos de integración de datos, concepto de tensores, conceptos de normalización de datos, conceptos de ingeniería de datos, preparación de datos para algoritmos de machine learning, preparación de datos para algoritmos de Deep Learning
- 5) Modelos  
Problemas de data mining/Big Data, clasificación, regresión, clustering. Técnicas de abordaje de cada uno.  
Algoritmos de machine learning  
Algoritmos de deep learning
- 6) Evaluación de modelos

Técnicas de validación, K-fold, jack knife, permutación, etc.

7) Aplicaciones

Clasificación, Regresión, clustering. Clasificación de imágenes, de señales temporales, datos tabulares. Aplicaciones en el ámbito de la salud y otras

Práctica:

Se realizará capacitación en lenguaje R y Python, paralelamente con el curso se darán clases iniciales en ambos lenguajes y la práctica se realizará en ambos lenguajes.

Horas: 60 Hrs (30 teóricas, 30 prácticas), examen final.

Docente:

Dr. Elmer Fernández , Investigador Independiente - CONICET

Bibliografía:

MacKay DJ. Information Theory, Inference, and Learning Algorithms. Copyright Cambridge University Press 2003.

Banks D, Statistical data mining. Volume 2, January/February 2010 © 2009 John Wiley & Sons, Inc.

Wim P. Krijnen, Applied Statistics for Bioinformatics using R, 2009

Zhao Y., R and Data Mining: Examples and Case Studies, 2011

Pyle D. Data preparation for data Mining 1999

Recursos web

[Introducción a Deep Learning with keras in R](#)

[A gentle introduction to deep learning with keras in r](#)

[Deep neural Networks with keras](#)