



## Especialización en Bioinformática

---

### Técnicas inteligentes para el análisis de datos

Año 2022

Duración total: 70hs.

Responsable: Facultad de Informáticas

Docente Responsable: Dra. Laura Lanzarini

Docentes: Dr. Hasperue Waldo  
Dr. Quiroga Facundo

---

### **OBJETIVO GENERAL**

La Inteligencia de datos es la rama de la Inteligencia Artificial capaz de transformar datos en información e información en conocimiento. En este curso se analizarán distintas técnicas y se resolverán problemas concretos. Al finalizar el curso el alumno conocerá los principales modelos y algoritmos inteligentes aplicables al análisis de datos y podrá seleccionar el más adecuado para los casos prácticos que se presenten.

### **COMPETENCIAS A DESARROLLAR EN RELACION CON EL OBJETIVO DE LA CARRERA**

C. 2 - Utilizar distintas técnicas de procesamiento de datos biológicos para su representación/visualización y análisis eficiente, mediante algoritmos de software ejecutados sobre plataformas adecuadas para el tipo y volumen de datos en cuestión.

C. 3 - Tener capacidad para evaluar la eficiencia (en tiempo, recursos utilizados y consumo energético) así como la calidad de las soluciones obtenidas, considerando el ámbito crítico de la aplicación de las mismas.

### **PROGRAMA**

- Inteligencia de Datos. Introducción. Aprendiendo patrones a partir de los datos. Diferencias entre Inteligencia Artificial, Inteligencia de Datos, Ciencia de Datos, Aprendizaje Automático, Minería de Datos.

- Almacenamiento y estructura de datos. Abstracción. Generalización. Evaluación. Aprendizaje supervisado y no supervisado. Tipos de datos de entrada y tipos de algoritmos de aprendizaje.

- Atributos nominales y numéricos Preprocesamiento y generación de características. Visualización y preprocesamiento. Visualización en espacios n-dimensionales. Diagramas de dispersión. Importancia de la normalización de los datos. Selección de atributos.

- Árboles de decisión. Métricas de selección de atributos. Entropía. Ganancia de Información. Tasa de Ganancia. Índice Gini. Poda y Sobreajuste. Algoritmos Id3, C4.5 y Random Forest. Construcción de árboles para grandes volúmenes de datos.



- Reglas de clasificación. Partición vs cobertura. Métodos de obtención de reglas de clasificación. Reglas de asociación. Calidad de las reglas. Métricas de una regla: soporte, cobertura, confianza, interés y convicción. Algoritmo A priori. Concepto de ítem frecuente. Mejoras del algoritmo a priori: FP-Growth y FP-Tree.

- Técnicas de Agrupamiento. Métricas de calidad del agrupamiento. Tipos de agrupamiento: Jerárquico, partitivo y probabilista. Medidas de distancia y de conectividad. Proceso de agrupamiento. Clustering partitivo. Algoritmo k-medias. Algoritmos de clustering jerárquicos aglomerativos y divisivos. Dendrogramas. Algoritmo probabilista EM (Expectation - Maximization).

- Redes Neuronales Feedforward. Descripción de la arquitectura. Regla delta generalizada. Algoritmo de entrenamiento backpropagation. Incorporación del término de momento. Capacidad de generalización de la red. Resolución de problemas de clasificación y predicción. Aprendizaje profundo.

## **METODOLOGIA Y MODALIDAD DE EVALUACION**

La metodología se basa en clases presenciales combinadas con actividades experimentales en el Aula y/o Laboratorio para aplicar los conceptos teóricos y que así el alumno adquiriera las competencias y habilidades sobre cada uno de los temas que forman parte del contenido de la asignatura.

Se requiere un 80% de asistencia a las clases presenciales (teóricas y prácticas).

El trabajo se complementa con un proyecto que debe desarrollar el alumno para cumplimentar las horas asignadas con soporte tutorizado por el profesor / docentes de la asignatura.

La evaluación se realizará mediante un examen escrito al final de las clases del curso para evaluar el grado de conocimientos del alumno (20%), el proyecto/desarrollo experimental que deberá entregar el alumno al final de las horas programadas (70%) y la participación y aportaciones de calidad/excelencia a las soluciones propuestas (10%).

## **ACTIVIDADES EXPERIMENTALES y DE INVESTIGACION**

El alumno debe desarrollar un proyecto, cuyo despliegue práctico se realiza sobre equipamiento físico (o virtualizado en el caso de aplicaciones informáticas).

En las actividades prácticas se utilizarán ejemplos y casos de estudio con distintas técnicas de procesamiento de datos biológicos para su representación/visualización y análisis eficiente, mediante algoritmos de software ejecutados sobre plataformas adecuadas para el tipo y volumen de datos en cuestión. Interesa evaluar la eficiencia (en tiempo, recursos utilizados y consumo energético) así como la calidad de las soluciones obtenidas, considerando el ámbito crítico de la aplicación de las mismas.

Los trabajos experimentales pueden desarrollarse en cada clase o continuarse en más de una clase. Parten de una especificación/consigna del docente (explicada en la clase) y un trabajo individual o en grupos que interactúan en el que los alumnos resuelven un problema experimental concreto relacionado con la temática.

Los trabajos podrán ser individuales o grupales (máximo 3 alumnos).

Estos trabajos pretenden desarrollar y/o fortalecer las aptitudes de opinión crítica en los temas relativos del curso. Los alumnos deberán sintetizar su comprensión de los temas, al realizar



correctamente la tarea experimental propuesta.

También se pretende desarrollar la capacidad de poder comunicar y transmitir los resultados, en presentaciones pautadas a lo largo del curso.

En general, finalizada una actividad, hay una sesión de discusión conjunta donde los participantes comunicarán sus opiniones e intercambiarán los distintos puntos de vista.

Como materiales de estudio, se dispone de:

- Presentaciones multimedia desarrolladas ad-hoc para introducir cada uno de los diferentes ejes temáticos.
- Ejemplos donde se aplican los conceptos teóricos
- Ejercicios prácticos que son desarrollados en clase
- Píldoras formativas con la explicación de algunos temas
- Material de lectura para estudiar y profundizar conceptos abordados en las clases
- Enlaces a artículos de actualidad de repositorios reconocidos en el área
- Libros digitales
- Acceso a equipamiento de laboratorio físico disponible en la Facultad de Informática y la Facultad de Ciencias Exactas, relacionado directamente con el contenido de la asignatura.
- Software específico para determinadas actividades de laboratorio que se detallan en este programa.

### **BIBLIOGRAFÍA BASICA**

- Introducción a la Minería de Datos (2004). Hernández Orallo, Ramírez Quintana, Ferri Ramírez. Editorial Pearson – Prentice Hall.
- Advanced Data Mining Technologies in Bioinformatics (2006). Hui-Huang Hsu. Idea Group Pub.
- Machine Learning: An Algorithmic Perspective – 2nd Ed. (2015). Stephen Marsland. CRC Press.
- Data Mining for Bioinformatics Applications (2015). Zengyou He. Woodhead Publishing.
- Python Data Science Handbook: Essential Tools for Working with Data (2016). 1st Edition. Jake VanderPlas. O'Reilly Media.
- Data Mining: Practical Machine Learning Tools and Techniques (Fourth Edition) (2017). Ian H. Witten, Eibe Frank, Mark A. Hall. Morgan Kaufmann. ISBN 978-0-12-804291-5.
- Bioinformatics Algorithms: Design and Implementation in Python (2018). Rocha Miguel, Ferreira Pedro G. Academic Press.