



Técnicas inteligentes para el análisis de datos Año 2024	Carrera: Especialización en Bioinformática
	Docente responsable: Dr. Facundo Manuel Quiroga
	Docentes: Dr. Franco Ronchetti, Lic. Santiago Andrés Ponte Ahón
	Carga horaria: 70hs.

OBJETIVOS

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.

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. Modelos Supervisados y No supervisados. Aplicaciones.
- 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.



Normalización de datos. Selección de atributos. Reducción de Características

- Modelos no Supervisados con énfasis en Técnicas de Agrupamiento. Tipos de agrupamiento: Jerárquico, partitivo, aglomerativo y probabilístico. Métricas de calidad del agrupamiento. Medidas de distancia y de conectividad. Proceso de agrupamiento. Ejemplos: k-medias, DB-SCAN.
- Modelos de clasificación. Ejemplos: Regresión Logística, Árboles de Decisión, Random Forest, SVM, Redes Neuronales, Reglas de Clasificación.
- Modelos de Regresión. Ejemplos: Regresión Logística, Árboles de Regresión, SVR, Redes Neuronales.

METODOLOGÍA Y MODALIDADES DE EVALUACIÓN

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

Las clases teóricas tienen autoevaluaciones asociadas para que los alumnos puedan realizar un seguimiento de su progreso y tengan retroalimentación constante. Las autoevaluaciones no tienen una nota asociada pero su realización es obligatoria.

La evaluación se realizará mediante un examen escrito al final de las clases del curso para evaluar el grado de conocimientos del alumno. Se requiere la realización del 100% de las autoevaluaciones para poder rendir el examen final.

ACTIVIDADES EXPERIMENTALES Y DE INVESTIGACION

En las actividades prácticas se utilizarán ejemplos y casos de estudio con distintos problemas de Aprendizaje Automático así como distintos modelos y conjuntos de datos biológicos. De esta forma, se hará énfasis en que los alumnos desarrollen habilidades prácticas para la carga y procesamiento de conjuntos de datos, el diseño de características, la elección de modelos y métricas, y su entrenamiento y evaluación correspondiente.

BIBLIOGRAFÍA



- 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.
- Sarah Guido, Andreas Müller. Introduction to Machine Learning with Python. O'Reilly Media. 2016.
- Jason Bell. Machine Learning: Hands-On for Developers and Technical Professionals. Wiley. 2015.