

PLANIFICACIÓN 2020

Minería de Datos en Bioinformática

INFORMACIÓN GENERAL

Carrera	Docente Responsable	
Ingeniería en Informática	Georgina Silvia Stegmayer	
Departamento	Carga Horaria	
Formación Complementaria	Carga Horaria Cuatrimestral	60 hs
Plan de Estudios	<i>TEORÍA</i>	20 hs
Plan 2006	<i>PRÁCTICA</i>	
Carácter	Formación Experimental	6 hs
Cuatrimestral	Resolución de Problemas	8 hs
Equipo Docente	Resolución de Problemas de Ingeniería	8 hs
Diego Humberto Milone	Proyectos y diseños de procesos	6 hs
Georgina Silvia Stegmayer	<i>CONSULTAS Y OTRAS ACTIVIDADES</i>	4 hs
	<i>EVALUACIONES</i>	8 hs

SITIO WEB DE LA ASIGNATURA

CONTENIDOS MÍNIMOS DE LA ASIGNATURA

Dado el actual crecimiento de la cantidad de datos a analizar (la denominada "big-data") en diversas áreas de aplicación, en particular la biología, la informática ha pasado a jugar un rol fundamental para el análisis, interpretación e inferencia de conocimiento a partir de esos datos. Esto se suma al auge actual de la disciplina bioinformática, en la cual se aplican técnicas de minería de datos para el análisis de datos de origen biológico (tales como expresión de genes, perfiles metabólicos, proteínas, entre otros). Existen muchos conceptos, metodologías y técnicas de la minería de datos que pueden potenciar y enriquecer las posibilidades que posee el alumno pronto a recibirse, sobre todo para enfrentar problemas en dominios que difieren de aquellos que tradicionalmente se desarrollan durante el dictado de su carrera de grado. Se propone introducir al alumno en conceptos básicos acerca de los tipos de datos biológicos que se generan hoy en día y pueden ser objeto de estudio. Mostrarle los algoritmos y modelos básicos de análisis, agrupamiento y clasificación de esos datos para búsqueda de conocimiento o comportamientos ocultos en ellos. Se presentarán además las técnicas estadísticas que pueden aplicarse para verificar y validar los resultados obtenidos con los métodos del área. Se organizará el dictado de la asignatura bajo una perspectiva ingenieril: desarrollar los contenidos seleccionados con una fuerte componente de trabajo de los alumnos en implementación computacional de los algoritmos vistos en clase. Se presentarán los temas que se consideran necesarios para que el alumno pueda adquirir los conceptos básicos que le permitan comprender la esencia del área y formar criterios para la aplicación de estas técnicas. Se pretende que las mismas le brinden al futuro ingeniero formación y herramientas que le ayuden a desarrollar actividades específicas de su profesión.

OBJETIVOS DE LA ASIGNATURA

Objetivo general:

El objetivo general de la asignatura es introducir al alumno en el estudio de la minería de datos en bioinformática. Para ello se estudiarán e implementarán los algoritmos básicos de minería de datos, tales como clasificadores (máquinas de vector soporte, redes neuronales) y clustering o agrupamiento (k-medias, agrupamiento jerárquico y mapas auto-organizativos). En cuanto al tipo de datos con el cual se trabaja en el área, se introducirán brevemente los conceptos biológicos necesarios para entender sus características, complejidad y objetivo de estudio desde la perspectiva del campo de aplicación.

Objetivos específicos:

Que el alumno:

- Conozca los diferentes tipos de datos biológicos existentes y sus características.
- Conozca cuáles son los tipos de problemáticas estudiadas en el área de minería de datos en bioinformática.
- Domine y sepa aplicar las técnicas básicas de agrupamiento y clasificación para resolver esos problemas con datos del dominio biológico.
- Sea capaz de resolver problemas con iniciativa autonomía y creatividad.

CONOCIMIENTOS ESPECÍFICOS PREVIOS PARA CURSAR LA ASIGNATURA

Programación y cálculo básico.

La optativa está pensada para alumnos de 4to y 5to año de Ing. en Informática.

METODOLOGÍA DE ENSEÑANZA

Las 4 horas semanales se distribuyen de la siguiente manera en:

- 2 horas de clases teóricas: presentación, exposición dialogada y ejemplificación de los temas, con apoyo de material bibliográfico y recursos didácticos. Se presentan ejemplos y problemas mediante el empleo de transparencias y/o desarrollos en la pizarra.

- 2 horas de clases prácticas: para afianzar el dominio de los tópicos discutidos, donde se presentan actividades de fijación, revisión e implementación de los conceptos vistos en las clases teóricas. Las prácticas serán orientativas para permitir el aprendizaje activo del alumno, se promoverá la discusión de soluciones y análisis de casos propuestos por los estudiantes.

Se resuelven problemas presentados de las guías de ejercicios prácticos para cada unidad del programa analítico. Las guías que se presentan son las siguientes:

Guía de datos biológicos (Unidad 2): enunciados con distintas situaciones problemáticas en las cuales se deben pre-procesar distintos tipos de datos biológicos, extraer características de secuencias, calcular distancias semánticas entre anotaciones de genes, entre otros.

Guía de clustering y clasificación (Unidad 3): implementación de algoritmos de clasificación y clustering con bases de datos libres, artificiales y biológicas reales.

Guía de análisis de resultados (Unidad 4): aplicación de medidas de calidad en clustering (cohesión, separación, etc). Análisis estadístico de los resultados (tests de hipótesis, análisis de la varianza)

Guía de inferencia de redes (Unidad 5): búsqueda de relaciones entre componentes de un cluster mediante algoritmos de búsqueda.

Los alumnos deberán desarrollar un Proyecto Integrador (PI) grupal a lo largo del cuatrimestre, aplicando e implementando las técnicas vistas en clase. En el PI se presenta un problema real de bioinformática a resolver incluyendo el diseño de soluciones y la resolución de problemas abiertos. Por ejemplo, a partir de una base de datos biológica hay que procesarla, encontrar relaciones entre sus datos, validarlos y reconstruir la vía metabólica que los vincula. Cada trabajo práctico se presenta con una

exposición oral en donde el alumno debe demostrar cómo resolvió el problema planteado por los profesores.

PROGRAMA ANALÍTICO

Título: UNIDAD 1 – INTRODUCCION.
Descripción/ Qué es la minería de datos. Qué es la Bioinformática. Antecedentes y
Contenidos: contexto histórico. Desafíos actuales en la minería de datos biológicos y “Big-data”. Proceso de minería de datos biológicos.

Título: UNIDAD 2 - DATOS BIOLOGICOS.
Descripción/ Tipos y características particulares.. Bases de datos biológicas.
Contenidos: Secuencias de ADN: tipos, plegado, alineación de secuencias y extracción de características. Genes, Gene Ontology (GO) y similaridad. Cómo medir distancias entre datos biológicos (distancia euclídea, correlación, similaridad semántica). Redes metabólicas y de regulación de genes.

Título: UNIDAD 3 - MINERIA DE DATOS.
Descripción/ Clustering. Técnicas y modelos de clustering. Agrupamiento Jerárquico.
Contenidos: K-medias y Partición en torno a Medoides (PAM). Mapas auto-organizativos (SOM). Ensamble de clusters. Clasificación. Redes neuronales. Máquinas de vector soporte (SVM).

Título: UNIDAD 4 – VALIDACION DE RESULTADOS.
Descripción/ Análisis estadístico de los resultados. Análisis de la varianza. Medidas de
Contenidos: calidad en clustering. Medidas de validación externas e internas. Medidas de estabilidad en clustering.

Título: UNIDAD 5 – GENERACION DE NUEVO CONOCIMIENTO.
Descripción/ Algoritmos de búsqueda de redes de relaciones. Búsqueda en profundidad y en
Contenidos: amplitud, algoritmos evolutivos. Inferencia de redes metabólicas. Reconstrucción de redes de regulación entre genes a partir de mediciones temporales.

BIBLIOGRAFÍA BÁSICA

Título: Artificial Intelligence: A Modern Approach
Autores: Russell, S., Norvig, P.
ISBN: **Editorial:** Prentice Hall
Formato:
Selección de No se ha especificado la selección de páginas.
Páginas:

Título: Clustering
Autores: R. Xu and D. C. Wunsch.
ISBN: **Editorial:** IEEE
Formato:

Selección de Páginas: No se ha especificado la selección de páginas.

Título: Computational Intelligence in Bioinformatics
Autores: G. Fogel and D. Corne,
ISBN: **Editorial:** Morgan Kaufmann,
Formato:

Selección de Páginas: No se ha especificado la selección de páginas.

Título: Data Mining
Autores: I. Witten & E. Franck
ISBN: **Editorial:** Morgan-Kauffmann
Formato:

Selección de Páginas: No se ha especificado la selección de páginas.

Título: Data mining over biological datasets: an integrated approach based on computational intelligence. IEEE Computational Intelligence Magazine, Special Issue on Computational Intelligence in Bioinformatics
Autores: Stegmayer, G., Gerard, M., Milone, D.
ISBN: **Editorial:** IEEE
Formato:

Selección de Páginas: No se ha especificado la selección de páginas.

Título: Intelligent Bioinformatics: The Application of Artificial Intelligence Techniques to Bioinformatics Problems.
Autores: E. Keedwell and A. Narayanan,
ISBN: **Editorial:** Wiley
Formato:

Selección de Páginas: No se ha especificado la selección de páginas.

BIBLIOGRAFÍA COMPLEMENTARIA

No se ha carga bibliografía complementaria para esta asignatura.

CRONOGRAMA DE ACTIVIDADES

Actividad: Unidad 1. Introducción. Modalidad de cursado. Condiciones de regularidad.
Semana: 1
Horas: 4
Tipo: T
Docentes a Cargo: Georgina Silvia Stegmayer

Actividad: Unidad 2. Explicación de los distintos tipos de datos biológicos. KEGG. Practica con genes y metabolitos su pre-procesamiento (normalización).
Semana: 2
Horas: 2
Tipo: T
Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 2. Explicación de los distintos tipos de datos biológicos. KEGG. Practica con genes y metabolitos su pre-procesamiento (normalización).
Semana: 2
Horas: 2
Tipo: EP
Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 2. Genes y GO. Anotaciones semánticas. Distancias. Práctica de cálculo de distinto tipo de matrices de distancias entre datos biológicos.
Semana: 3
Horas: 2
Tipo: EP
Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 2. Genes y GO. Anotaciones semánticas. Distancias. Práctica de cálculo de distinto tipo de matrices de distancias entre datos biológicos.
Semana: 3
Horas: 2
Tipo: T
Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 3. Clustering. Ensamble de clustering. Implementación y comparación de resultados en diferentes tipos de datos artificiales.
Semana: 4
Horas: 2
Tipo: T
Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 3. Clustering. Ensamble de clustering. Implementación y comparación de resultados en diferentes tipos de datos artificiales.
Semana: 4
Horas: 2
Tipo: EP
Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 3. Clasificación. Perceptrón multicapa y máquinas de vector soporte. A partir de secuencias: plegado, extracción de características y clasificación de miRNAs.

Semana: 5

Horas: 2

Tipo: T

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 3. Clasificación. Perceptrón multicapa y máquinas de vector soporte. A partir de secuencias: plegado, extracción de características y clasificación de miRNAs.

Semana: 5

Horas: 2

Tipo: EP

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Defensa Proyecto Integrador 1

Semana: 6

Horas: 2

Tipo: P/D

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Defensa PI 1

Semana: 6

Horas: 2

Tipo: PI

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Parcial 1: Unidades 1, 2 y 3

Semana: 7

Horas: 4

Tipo: E

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Consultas

Semana: 7

Horas: 2

Tipo: C

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 4. Validación de resultados. Estabilidad. Medidas. Implementación y práctica de las medidas sobre bases de datos artificias y reales disponibles.

Semana: 8

Horas: 2

Tipo: T

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 4. Validación de resultados. Estabilidad. Medidas. Implementación y práctica de las medidas sobre bases de datos artificias y reales disponibles.

Semana: 8

Horas: 2

Tipo: PL

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 4. Análisis estadístico de resultados. Implementación y práctica sobre bases de datos artificias y reales disponibles.

Semana: 9

Horas: 2

Tipo: T

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 4. Análisis estadístico de resultados. Implementación y práctica sobre bases de datos artificias y reales disponibles.

Semana: 9

Horas: 2

Tipo: PL

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 5. Reconstrucción de redes de regulación entre genes. Búsqueda de redes metabólicas con algoritmos evolutivos.

Semana: 10

Horas: 2

Tipo: T

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 5. Reconstrucción de redes de regulación entre genes. Búsqueda de redes metabólicas con algoritmos evolutivos.

Semana: 10

Horas: 2

Tipo: PL

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 5. Algoritmos clásicos de búsqueda. Búsqueda en profundidad, en amplitud. Búsqueda óptima y completa. Práctica de búsqueda de relaciones entre un conjunto de metabolitos y genes.

Semana: 11

Horas: 2

Tipo: T

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Unidad 5. Algoritmos clásicos de búsqueda. Búsqueda en profundidad, en amplitud. Búsqueda óptima y completa. Práctica de búsqueda de relaciones entre un conjunto de metabolitos y genes.

Semana: 11

Horas: 2

Tipo: P/D

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Defensa Proyecto Integrador 2

Semana: 12

Horas: 2

Tipo: PI

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Consulta

Semana: 12

Horas: 2

Tipo: C

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Defensa Proyecto Integrador 2

Semana: 12

Horas: 2

Tipo: P/D

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Parcial 1: unidades 4 y 5.

Semana: 13

Horas: 4

Tipo: E

Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

Actividad: Coloquio Final Integrador
Semana: 15
Horas: 4
Tipo: PI
Docentes a Cargo: Georgina Silvia Stegmayer, Georgina Silvia Stegmayer

REQUERIMIENTOS DE LA ASIGNATURA

Detallar cuanto sea necesario para que los alumnos no tengan dudas sobre cada uno de estos requerimientos:

Para Regularizar: Alumno regular: El alumno queda regular cuando cumple con:
- aprobación de 2 (DOS) exámenes parciales, debiendo aprobarse cada uno de ellos con un mínimo del 40%. Se podrá acceder a un recuperatorio si alguno de los parciales resultara desaprobado.
- aprobación de un proyecto integrador con presentación y defensa oral.
Los alumnos que no cumplan con los requisitos quedarán en condición de libres.

Para Promocionar: Promoción directa sin examen final: Para alcanzar la promoción de la asignatura los alumnos deberán cumplir las siguientes condiciones y actividades:
- aprobación de 2 (DOS) exámenes parciales de teoría y práctica, debiendo obtener una calificación no inferior al 60% en cada uno de ellos. Para cada examen se podrá acceder a un recuperatorio.
- aprobación de un proyecto integrador y un coloquio final integrador (CFI) que consistirá en una evaluación global de los temas abordados en la asignatura, mediante un diálogo con los docentes. El alumno que no apruebe esta instancia quedará en condición de regular.
A efectos de la nota final, se hará un promedio ponderado de las notas obtenidas en los parciales, PI y CFI.

EXAMEN FINAL

Para Alumnos Regulares: Escrito individual

Para Alumnos Libres: Dos TP practicos, una actividad practica integradora y un examen integrador escrito individual.

EVALUACIONES

PARCIALES

Fecha: 06-05-2020 **Título:** Parcial 1: unidades 1, 2 y 3

Temas /

Descripción:

Fecha: 10-06-2020 Título: Parcial 2: unidades 4 y 5

Temas /
Descripción:

TRABAJOS PRÁCTICOS

Fecha: 29-04-2020 Título: Defensa Proyecto Integrador parte I

Temas /
Descripción:

Fecha: 03-06-2020 Título: Defensa Proyecto Integrador parte 2

Temas /
Descripción:

RECUPERATORIOS

Fecha: 17-06-2020 Título: Recuperatorios

Temas /
Descripción:

COLOQUIOS

Fecha: 17-06-2020 Título: Coloquio final integrador

Temas /
Descripción:

INFORMACIÓN COMPLEMENTARIA

No se ha ingresado información complementaria para esta asignatura