

EXP-UNC 26816/2014

RESOLUCIÓN CD N° 163/2014

VISTO

El pedido presentado por el Dr. Daniel Fridlender para que se incorpore la materia "Aprendizaje automático sobre grandes volúmenes de datos" como materia Optativa de la Licenciatura en Ciencias de la Computación; y

CONSIDERANDO

Que se cuenta con el acuerdo de la Comisión Asesora de Computación y el aval del Consejo de Grado;

Que es conveniente agregar a la nómina de materias optativas, aprobada por Res. HCD N°207/02, la asignatura que se propone;

Que mediante Resolución HCS N° 122/02 se ha delegado en este cuerpo la facultad de modificar la nómina de materias optativas del Plan de Estudios de la Licenciatura en Ciencias de la Computación.

EL CONSEJO DIRECTIVO
DE LA FACULTAD DE MATEMÁTICA, ASTRONOMÍA Y FÍSICA

RESUELVE:

ARTÍCULO 1º: Modificar la nómina de materias optativas del Plan de Estudios de la Licenciatura en Ciencias de la Computación, incorporando a la misma la materia "Aprendizaje automático sobre grandes volúmenes de datos".

ARTÍCULO 2º: Fijar como programa, correlativas y carga horaria de la materia, los detallados en el Anexo que forma parte de la presente Resolución.

ARTÍCULO 3º: En cumplimiento con lo establecido en el Artículo 2º de la Res. HCS N° 122/02, remítase a la Secretaria de Asuntos Académicos de la Universidad la presente resolución para su conocimiento y efectos.

ARTÍCULO 4º: Comuníquese y archívese.

DADA EN LA SALA DE SESIONES DEL CONSEJO DIRECTIVO DE LA FACULTAD DE MATEMÁTICA, ASTRONOMÍA Y FÍSICA, A NUEVE DÍAS DEL MES DE JUNIO DE DOS MIL CATORCE.

st


Dr. SERGIO A. CANNAS
Secretario General
FAMAF


Dra. ESTHER GALINA
DECANA
FAMAF

ANEXO RESOLUCIÓN CD N° 163/2014

PROGRAMA DE ASIGNATURA

ASIGNATURA: Aprendizaje Automático Sobre Grandes Volúmenes de Datos		AÑO: 2014
CARÁCTER: Optativa / Especialidad		
CARRERA: Lic. En Ciencias de la Computación		
RÉGIMEN: cuatrimestral	CARGA HORARIA: 120 hs.	
UBICACIÓN en la CARRERA: 5to. año - 2do. Cuatrimestre		

FUNDAMENTACIÓN Y OBJETIVOS

Fundamentación:

El aprendizaje automático permite un nuevo paradigma de programación en el cual algoritmos genéricos son adaptados para problemas dados mediante el uso de datos específicos. Cuando los datos a procesar exceden la capacidad de cómputo de una computadora, se hacen necesarias técnicas especiales adaptadas del cálculo distribuido. En esta materia se estudiarán dichas técnicas primero a nivel teórico y después mediante casos de estudio prácticos. Es de especial interés para aquellos estudiantes de la FaMAF interesados en la aplicaciones tecnológicas y en la investigación interdisciplinaria.

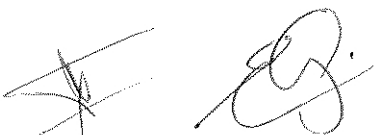
Objetivos:

Al finalizar la materia los estudiantes estarán en condiciones de:

- Comprender los lineamientos básicos de técnicas de aprendizaje automático, en particular las empleadas en aprendizaje de grandes volúmenes de datos.
- Comprender los lineamientos básicos de técnicas de cómputo distribuido, en particular las empleadas en aprendizaje de grandes volúmenes de datos.
- Tener las herramientas fundamentales para generalizar algoritmos de aprendizaje automático a ambientes de cómputo distribuido.
- Analizar un problema que involucre grandes volúmenes de datos y diseñar experimentos de aprendizaje automático sobre dichos datos.

CONTENIDO

Parte I: Conceptos de Aprendizaje Automático



Generalidades de Aprendizaje Automático. Clasificación, Sistemas de Recomendación y Aprendizaje Sin Supervisión. Ciclo del Aprendizaje Automático. Características (features) para el Aprendizaje. Modelos. Sesgo (bias) de Modelos. Modelos Generativos vs. Discriminantes.

Algoritmos específicos de Clasificación. Naive Bayes. Técnicas de Smoothing. Árboles de decisión. Random Forests. Regresión logística. Ingeniería de Features. Máquinas a base de Vectores de Soporte (Support Vector Machines).

Aprendizaje Sin Supervisión. Clustering. kMeans. Clustering basado en modelado estadístico.

Parte II: Conceptos de Cómputo Distribuido

Metodología MapReduce. Función Map. Función Reduce. Claves intermedias. Conceptos de Análisis Numérico. Multiplicación distribuida de Matrices. Inversión distribuida de matrices. Búsqueda distribuida. Óptimo local vs. global. Algoritmos actualizables. Procesamiento por lotes. Modelos alternativos: Colas de Mensajes y Memoria Compartida.

Parte III: Implantación y Casos de Estudio

Plataforma Hadoop como implantación del paradigma MapReduce. Orquestado de clusters de cómputo. OpenStack. Salt.

Paquete Mahout. Recomendación automática. Clustering. Clasificación. Preparación de datos. Monitoreo, distribución de nodos.

Plataforma ActiveMQ como implementación de colas de mensajes. Paradigmas híbridos. Storm. Sparck. Mlib.

Casos de estudio. Recomendación de Películas. Identificación de Páginas co-editadas en Wikipedia. Clasificación de Programas Compilados.

BIBLIOGRAFÍA

Mahout in Action por Sean Owen, Robin Anil, Ted Dunning, y Ellen Friedman (2011) ISBN 9781935182689.

METODOLOGÍA DE TRABAJO

Consistirá en clases teóricas magistrales, clases teórico-prácticas, resolución de guías de ejercicios y clases en laboratorio.



EVALUACIÓN

FORMAS DE EVALUACIÓN

Las primeras dos partes serán evaluadas mediante una batería de ejercicios escritos y parcial. La tercera parte será evaluada con un proyecto que involucre la implementación de un proyecto sobre los datos presentados en clase. La materia será promocionable como se detalla más adelante.

CONDICIONES PARA OBTENER LA REGULARIDAD

El alumno deberá:

- Cumplir un mínimo de 70% de asistencia a clases teóricas, prácticas, o de laboratorio.
- Aprobar dos evaluaciones parciales o sus correspondientes recuperatorios (sólo se puede recuperar uno de los parciales).
- Aprobar el proyecto de clase.

CONDICIONES PARA OBTENER LA PROMOCIÓN

El alumno deberá:

- Cumplir un mínimo de 80% de asistencia a clases teóricas, prácticas, o de laboratorio.
- Aprobar las dos evaluaciones parciales con una nota no menor a 6 (seis), y obteniendo un promedio no menor a 7 (siete).
- Aprobar el proyecto de la clase.

Nota: Esta materia también será ofrecida a alumnos de posgrado, quienes podrán realizar un proyecto implementando nuevos algoritmos de aprendizaje y rendir un final que involucra escribir un ensayo diseñando un experimento relacionado.

CORRELATIVIDADES

MATERIA OPTATIVA	CORRELATIVAS		
	PARA CURSAR		PARA RENDIR
Aprendizaje Automático Sobre Grandes Volúmenes de Datos	REGULARIZADA	APROBADA	APROBADA
		Probabilidad y Estadística, Redes y Sistemas Distribuidos	Álgebra

