



**UNIVERSIDAD NACIONAL DE SANTIAGO DEL ESTERO**  
**FACULTAD DE CIENCIAS FORESTALES**

**CURSO DE POSGRADO:**

**MINERIA DE DATOS EN CIENCIAS AMBIENTALES**

**PROFESOR RESPONSABLE:**

**MS. PRISCILLA G MINOTTI**

**3IA- UNSAM**

**2 AL 6 DE SEPTIEMBRE**

**SANTIAGO DEL ESTERO**



**Nombre del Curso:** Minería de datos en Ciencias Ambientales

**Profesor Responsable:** MS. Priscilla G. Minotti 3iA-UNSAM

**Duración:** 40 horas presenciales + 20 desarrollo de proyecto

**Créditos:** 6 créditos

**Fecha:** 2 al 6 de septiembre 2019

**Horario:** 8:30 – 12:30, y 14:00 a 18:00 hs

### 1. Marco conceptual e importancia actual/Fundamento:

Aunque estamos en la era de la información con una generación continua de datos ambientales, sin embargo, se utiliza menor del 0,5% de lo que se produce, y muchas veces es solo para corroborar valores una única vez. La minería de datos, o data mining, constituye un componente clave de la generación de conocimiento a partir de datos. Es el proceso de identificación automatizada de información relevante extraída de grandes volúmenes de datos, con el objetivo de descubrir patrones y tendencias, estructurando la información obtenida de un modo comprensible para su posterior utilización. Las técnicas de minería de datos se apoyan en la estadística, la gestión de bases de datos y el desarrollo y aplicación de algoritmos de aprendizaje automático, temas que no suelen ser abordados de manera conjunta en las disciplinas ambientales. El curso busca dar herramientas para un mejor aprovechamiento de la información ambiental con foco en de Ingeniería ambiental y Ecología.

### 2. Objetivo General/ 2.1. Objetivos específicos

El objetivo general del curso es brindar una visión integral de las áreas de aplicación de la minería de datos en Ciencias Ambientales.

Como objetivos específicos se espera que los participantes puedan seleccionar métodos que resulten apropiados para sus áreas de estudio o aplicación y se familiaricen con algunas técnicas de minería de datos en entorno del lenguaje R.

### 3. Contenidos y cronograma

#### Día 1

1. Terminología y conceptos claves. Ciencia de datos. Minería de datos. Big Data. Machine learning o Aprendizaje estadístico. Inteligencia Artificial.



2. El Proceso de Minería de datos. Aprendizaje automatizado y estadística. La generalización como búsqueda. Patrones vs. Modelos. Minería de datos y ética.

3. Entrada de Datos: conceptos, instancias, atributos. Preparación de datos. Análisis exploratorio. Transformación de datos.

4. Salidas: Representación del conocimiento. Tablas. Modelos lineares. Árboles. Reglas, Representación por instancias. Grupos (Clusters). Proyección de un modelo a los datos.

Ejercitación 1. Introducción a R y Rstudio. Entrada y estructuración de datos tabulares. Filtrar filas, seleccionar columnas, reordenar filas, crear columnas con valores nuevos. Aplicar funciones por fila o columna. Aplicar funciones por grupos de datos.

Ejercitación 2. Análisis exploratorio visual. Visualizaciones y problemas con grandes volúmenes de datos. Ejemplos de tipos de visualizaciones: geográficas, estadísticas, mapas de calor, ejes paralelos, en redes. Visualizaciones con los paquetes i-graph, ggplot2 y tmap.

## **Día 2**

5. Algoritmos: Métodos básicos. Inferencia rudimentaria de reglas, modelos probabilísticos simples, arboles de decisión, cubrimiento de reglas. Reglas de asociación. Modelos lineares. Modelos basados en instancias. Agrupamiento.

6. Credibilidad y evaluación del aprendizaje. Entrenamiento y testeo. Medidas de Performance. Validación cruzada. Ajuste de parámetros. Comparación de modelos. Predicción probabilística. Costos. Evaluación de predicciones numéricas. Principio de predicción de longitud mínima. Selección de modelos.

Ejercitación 3. Ejemplos de trabajos publicados y en blogs. Asignación de trabajos a presentar en clase.

## **Día 3**

7. Arboles de decisión. En qué consisten. Problemas típicos. Terminología. Arboles de regresión vs arboles de clasificación. Aprendizaje por ensambles, bagging y boosting.

Ejercitación 4. Ejemplos con CART

Ejercitación 5. Ejemplos con Random forest.

## **Día 4**



8. Redes neuronales. En qué consisten. Terminología. Tipos de redes. Perceptrón multicapa. Redes de Kohonen o SOM. Aprendizaje profundo.

Ejercitación 5. Ejemplos con perceptrón multicapa

Ejercitación 6. Ejemplos con redes de Kohonen

#### **Día 5**

7. Panorama actual de aplicaciones de minería de datos en ciencias ambientales con discusión de trabajos a presentar por participantes del curso.

Explicación del trabajo final.

#### **4. Evaluación.**

El curso se aprueba con el desarrollo de un trabajo final, que según su complejidad puede ser realizado de manera grupal o individual.

#### **5. Bibliografía básica**

P. Tan, M. Steinbach, V. Kumar. 2006. Introduction to Data Mining. Addison-Wesley.

Ian H. Witten, Eibe Frank, Mark Hall, Christopher Pal. 2016. Data Mining: Practical Machine Learning Tools and Techniques, Fourth Edition (Morgan Kaufmann Series in Data Management Systems)

