



Ministerio de Cultura y Educación  
Universidad Nacional de San Luis  
Facultad de Ciencias Físico Matemáticas y Naturales  
Departamento: Informatica  
Area: Area V: Automatas y Lenguajes

(Programa del año 2024)  
(Programa en trámite de aprobación)  
(Presentado el 06/04/2024 07:51:12)

### I - Oferta Académica

| Materia                                      | Carrera      | Plan  | Año  | Período         |
|----------------------------------------------|--------------|-------|------|-----------------|
| () APRENDIZAJE AUTOMATICO Y MINERIA DE DATOS | LIC.CS.COMP. | 32/12 | 2024 | 1° cuatrimestre |

### II - Equipo Docente

| Docente                  | Función           | Cargo      | Dedicación |
|--------------------------|-------------------|------------|------------|
| ERRECALDE, MARCELO LUIS  | Prof. Responsable | P.Asoc Exc | 40 Hs      |
| CAGNINA, LETICIA CECILIA | Prof. Colaborador | P.Adj Exc  | 40 Hs      |

### III - Características del Curso

| Credito Horario Semanal |          |                   |                                       |       |
|-------------------------|----------|-------------------|---------------------------------------|-------|
| Teórico/Práctico        | Teóricas | Prácticas de Aula | Práct. de lab/ camp/ Resid/ PIP, etc. | Total |
| Hs                      | 2 Hs     | 2 Hs              | 1 Hs                                  | 5 Hs  |

| Tipificación                                   | Periodo         |
|------------------------------------------------|-----------------|
| B - Teoria con prácticas de aula y laboratorio | 1° Cuatrimestre |

| Duración   |            |                     |                   |
|------------|------------|---------------------|-------------------|
| Desde      | Hasta      | Cantidad de Semanas | Cantidad de Horas |
| 11/03/2024 | 21/06/2024 | 15                  | 75                |

### IV - Fundamentación

A partir de la popularización de Internet y el incremento del número de bases de datos accesibles, se ha producido en los últimos años un gran crecimiento en los volúmenes y disponibilidad de datos alrededor del mundo. Se ha estimado que la cantidad de datos almacenados en las bases de datos del mundo se duplica cada veinte meses. Este crecimiento cuantitativo de los datos no se ha reflejado en un crecimiento cualitativo de la información disponible. Existe un "gap" creciente entre la generación de los datos y nuestro entendimiento de los mismos.

Una alternativa para solucionar este problema que ha recibido un interés creciente es el área de minería de datos. La minería de datos (en inglés Data Mining y de ahora en más MD) es el proceso de descubrir patrones en grandes volúmenes de datos. Este proceso debe ser automático (o más usualmente) semi-automático. Los patrones descubiertos deben ser significativos en el sentido que deben conducir a alguna ventaja (generalmente económica).

### V - Objetivos / Resultados de Aprendizaje

Introducir al alumno en los principales conceptos vinculados al aprendizaje automático y la minería de datos. Al finalizar el curso los alumnos deberán ser capaces de identificar las principales etapas y procesos de la minería de datos, los mecanismos estadísticos y de aprendizaje automático frecuentemente utilizados en este área. Se espera además que los mismos sean

## VI - Contenidos

### Unidad 1. Aprendizaje Automático (AA).

Definiciones. Tipos de Aprendizaje. Importancia y Aplicaciones del AA. Minería de Datos. Definición. Tareas Predictivas. Tareas Descriptivas. Aplicaciones y Tipos de MD. El Proceso de KDD. Preparación de los datos. Minería de datos. Evaluación, visualización e interpretación. Entrada y Salida del Proceso de MD. Big Data (Analytics)

### Unidad 2. Aprendizaje Supervisado

Tipos de aprendizaje. Aprendizaje automático (supervisado). Aprendizaje de una hipótesis. Esquema General. Aprendizaje como Búsqueda. La necesidad del sesgo inductivo. Evaluación de una hipótesis. Medidas de evaluación de clasificador.

### Unidad 3: Regresión Lineal, descenso del gradiente y redes neuronales

Regresión Lineal (RL) Simple. RL múltiple. Mínimos cuadrados ordinarios. Descenso del gradiente. Redes Neuronales (RN) como una herramienta para hacer regresión no lineal. Estructura de las RN. Perceptrones. RN feed-forward con múltiples capas. Algoritmo de Back-Propagation.

### Unidad 4: Aprendizaje de árboles de decisión.

Algoritmo ID3. Medidas de entropía y ganancia de información. Sesgo inductivo en ID3. Métodos para evitar el sobre-entrenamiento.

### Unidad 5: Máquinas de vectores soporte (SVM).

Introducción. SVM para clasificación binaria. Justificación teórica. Aplicaciones de SVM. Extensiones y temas avanzados. Software para SVM.

### Unidad 6. Redes Neuronales Profundas

Redes neuronales convolucionales. Operación y capas de convolución. Operación de pooling. Capa de dropout. Desarrollos recientes. Redes neuronales recurrentes. Ideas Iniciales, Tipos de secuencia. Unidades de memoria, Long Short Term Memory. Gated Recurring Units. Ejemplos de aplicación.

### Unidad 7: Introducción al aprendizaje no supervisado

Introducción al Análisis de Clusters. Tipos de clusters. Medidas de Evaluación externas e internas. Reducción no supervisada de dimensionalidad. Introducción al Análisis de Asociaciones

### Unidad 8: Introducción a la Minería de Textos.

Representación de Documentos. Bolsa de Palabras. Ponderado de términos. Modelización de Tópicos. Enfoques Distribucionales. Enfoques neuronales al Procesamiento del Lenguaje Natural. Embeddings. Representaciones de textos como secuencias

### Unidad 9: Aplicaciones

Aspectos prácticos involucrados en el desarrollo de aplicaciones de MD para un problema particular. Caso de estudio: Minería de Textos y la Web. Herramientas de desarrollo de aplicaciones para MD. Scikit-learn. Keras. TensorFlow. Bases de datos públicas.

## VII - Plan de Trabajos Prácticos

Práctico 1: Aspectos generales.

Práctico 2: Modelos Lineales

Práctico 3: Redes Neuronales

Práctico 4: Redes Neuronales Profundas

Práctico 5: Análisis de Textos

Práctico 6: Desarrollo de un proyecto de programación a ser entregado a final de cuatrimestre.

Práctico de Máquina Nro. 1. Introducción al aprendizaje automático con Scikit-Learn.

Práctico de Máquina Nro.2. Aprendizaje de árboles de decisión

Práctico de Máquina Nro. 3. Redes Neuronales Artificiales

Práctico de Máquina Nro. 4. Redes Neuronales Convolucionales

Práctico de Máquina Nro. 5. Representación y análisis de textos

## VIII - Regimen de Aprobación

Los contenidos de la materia serán impartidos en modalidad presencial. No se admite rendir la materia en condición de libre.

A. Para regularizar la materia, los alumnos deberán cumplir:

A1) Un porcentaje de asistencia mínima a clases del 80%

A2) La entrega de la totalidad de los trabajos prácticos solicitados en las distintas unidades, los cuales serán provistos en un repositorio digital (tipo Google Classroom) de la materia. Cada uno de estos trabajos, tendrá dos instancias de recuperación en caso de no alcanzarse una nota mínima de 7 puntos.

A3) Entrega de un proyecto de programación propuesto vinculado a las temáticas del curso, con un informe adicional con las principales conclusiones.

B. Para promocionar la materia, los alumnos deberán:

B1). Cumplir con los requerimientos A1, A2 y A3 enunciados anteriormente.

B2) Aprobar con un mínimo de 7 (siete) puntos un examen integrador oral y/o escrito al final del cuatrimestre.

La nota final se computará promediando las notas obtenidas en los puntos A2, A3 y B2.

## IX - Bibliografía Básica

- [1] "Introduction to Data Mining". Second Edition. Pang-Ning Tan, Michael Steinbach y Vipin Kumar. Pearson. 2018.
- [2] "Introduction to Machine Learning with Python". Andreas C. Müller y Sarah Guido. O'Reilly Media, Inc., 2017.
- [3] "Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow". Aurélien Géron. 2da. Edición. O'Reilly Media, Inc., 2019.
- [5] "Data Mining: Practical Machine Learning Tools and Techniques". Third Edition. Witten y Frank. Morgan Kaufmann. 2011.
- [6] "Introducción a la Minería de datos". Orallo, J. H., Ramírez Quintana, Ma. J. y Ramírez, C.F.. Pearson Prentice Hall. 2004.
- [7] "Machine Learning". Tom Mitchell. McGraw-Hill Series in Computer Science, 1997.

## X - Bibliografía Complementaria

- [1] "Deep Learning". Ian Goodfellow and Yoshua Bengio and Aaron Courville. Mit Press, 2016.
- [2] "Python Machine Learning". Sebastian Raschka. 2015.

## **XI - Resumen de Objetivos**

Introducir al alumno en los principales conceptos vinculados al aprendizaje automático y la minería de datos. Al finalizar el curso los alumnos deberán ser capaces de identificar las principales etapas y procesos de la minería de datos y los mecanismos de aprendizaje automático frecuentemente utilizados en este área. Se espera además que los mismos sean capaces de detectar y aplicar las técnicas de minería de datos en aplicaciones concretas del mundo real.

## **XII - Resumen del Programa**

Unidad 1. Aprendizaje Automático (AA).  
Unidad 2. Aprendizaje Supervisado  
Unidad 3: Regresión Lineal, descenso del gradiente y redes neuronales  
Unidad 4: Aprendizaje de árboles de decisión.  
Unidad 5: Máquinas de vectores soporte (SVM).  
Unidad 6. Redes Neuronales Profundas  
Unidad 7. Aprendizaje No Supervisado  
Unidad 8: Introducción a la Minería de Textos.  
Unidad 9: Aplicaciones

## **XIII - Imprevistos**

|  |
|--|
|  |
|--|

## **XIV - Otros**

|  |
|--|
|  |
|--|

| ELEVACIÓN y APROBACIÓN DE ESTE PROGRAMA |                             |
|-----------------------------------------|-----------------------------|
|                                         | <b>Profesor Responsable</b> |
| Firma:                                  |                             |
| Aclaración:                             |                             |
| Fecha:                                  |                             |