



**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 2020)**

**I - Oferta Académica**

| <b>Materia</b>                                       | <b>Carrera</b> | <b>Plan</b> | <b>Año</b> | <b>Período</b>  |
|------------------------------------------------------|----------------|-------------|------------|-----------------|
| (OPTATIVA) APRENDIZAJE AUTOMATICO Y MINERIA DE DATOS | LIC.CS.COMP.   | 32/12       | 2020       | 1° cuatrimestre |

**II - Equipo Docente**

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

**III - Características del Curso**

| <b>Credito Horario Semanal</b> |                 |                          |                                              |              |
|--------------------------------|-----------------|--------------------------|----------------------------------------------|--------------|
| <b>Teórico/Práctico</b>        | <b>Teóricas</b> | <b>Prácticas de Aula</b> | <b>Práct. de lab/ camp/ Resid/ PIP, etc.</b> | <b>Total</b> |
| Hs                             | 2 Hs            | 2 Hs                     | Hs                                           | 4 Hs         |

| <b>Tipificación</b>                            | <b>Periodo</b>  |
|------------------------------------------------|-----------------|
| B - Teoria con prácticas de aula y laboratorio | 1° Cuatrimestre |

| <b>Duración</b> |              |                            |                          |
|-----------------|--------------|----------------------------|--------------------------|
| <b>Desde</b>    | <b>Hasta</b> | <b>Cantidad de Semanas</b> | <b>Cantidad de Horas</b> |
| 09/03/2020      | 10/07/2020   | 19                         | 76                       |

**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 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 8: 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. El sistema Weka. RapidMiner. Scikit-learn. Keras. TensorFlow. Bases de datos públicas.

## VII - Plan de Trabajos Prácticos

Práctico 1: Aspectos generales.

Práctico 2: Aprendizaje de Árboles de decisión.

Práctico 3: Redes Neuronales

Práctico 4: Redes Neuronales Convolucionales

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 impartidos durante el periodo no presencial (clases virtuales), así como las evaluaciones que se realicen, tendrán validez académica según estipula el Art. 7 de la Resol. CD 018-20. Las clases virtuales se impartirán mediante el uso de tecnologías como Google Meet, Zoom, videos preparados por la cátedra y material escrito que se proveerá a través de Google Classroom. Por tal motivo, no se considera la asistencia en el régimen de aprobación de la materia.

La materia sólo admite el régimen PROMOCIONAL. En este contexto, la aprobación del curso requiere:

- a. La entrega de la totalidad de los trabajos prácticos solicitados en las distintas unidades, los cuales serán provistos en la plataforma 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.
- b. Entrega de un proyecto de programación propuesto vinculado a las temáticas del curso, con un informe adicional con las principales conclusiones.
- c) Aprobar con un mínimo de 7 (siete) puntos un examen integrador oral y/o escrito al final del cuatrimestre bajo la modalidad que se comunique oportunamente (Arts. 10 y 11 de la Resol. CD 018-20).

La nota final se computará promediando las notas obtenidas en cada uno de los puntos mencionados previamente.

## IX - Bibliografía Básica

- [1] "Introduction to Data Mining". Tan, Steinbach y Kumar. Pearson. 2006.
- [2] "Data Mining: Practical Machine Learning Tools and Techniques". Third Edition. Witten y Frank. Morgan Kaufmann. 2011.
- [3] "Introducción a la Minería de datos". Orallo, J. H., Ramírez Quintana, Ma. J. y Ramírez, C.F.. Pearson Prentice Hall. 2004.
- [4] "Machine Learning". Tom Mitchell. McGraw-Hill Series in Computer Science, 1997.
- [5] "Readings in Machine Learning". Shavlik, J. y Dietterich, Thomas. Morgan Kaufmann, 1990.
- [6] "The Elements of Statistical Learning". T. Hastie, R. Tibshirani, J. H. Friedman. Springer, 2002.
- [7] "Principles of Data Mining (Adaptive Computation and Machine Learning)". David J. Hand, Heikki Mannila, Padhraic Smyth. MIT Press, 2001.

## X - Bibliografía Complementaria

- [1] "The Text Mining HandBook: Advanced Approaches in Analyzing Unstructured Data". Feldman and Sanger. Cambridge University Press. 2007.
- [2] "Deep Learning". Ian Goodfellow and Yoshua Bengio and Aaron Courville. Mit Press, 2016.
- [3] "Python Data Science Handbook: Essential tools for working with data". Jake VanderPlas (1ra. Edición). 2017.
- [4] "Introduction to Machine Learning with Python". Andreas C. Müller and Sarah Guido. (1ra. Edición). 2016
- [5] "Python Machine Learning". Sebastian Raschka. 2015.
- [6] "Python for Data Analysis". Wes McKinney. (2da. Edición). 2016
- [7] Publicaciones y apuntes complementarios sobre los distintos tópicos del curso que estarán disponibles en <http://www.dirinfo.unsl.edu.ar/~aamd>

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

## **XIII - Imprevistos**

Debido al contexto de aislamiento social preventivo y obligatorio, que derivó en un dictado en MODALIDAD NO PRESENCIAL, se decidió que el dictado de los contenidos en lugar de ser en 15 semanas se extendiera a 18 semanas, y con menos carga presencial semanal para facilitar la asimilación a la nueva modalidad.

Si por problemas de conectividad u otra eventualidad referida a la no presencialidad, los alumnos no pudieran entregar el material solicitado en las distintas tareas o participar de alguna forma de evaluación, los docentes propondrán alternativas para subsanar este inconveniente comunicándolo formalmente a la Secretaría Académica de la Facultad y luego a los estudiantes involucrados (Art. 12 de la Resol. CD 018-20).

## **XIV - Otros**