



---

## MINERIA DE DATOS

Carrera: *Maestría en Inteligencia de Datos orientada a Big Data*

Carga Horaria: 64 Hs.

**Año 2019**

Profesor a Cargo: *Dra. Laura Lanzarini*

---

### OBJETIVO

Brindar a los alumnos los conceptos necesarios para analizar un problema y presentar los resultados obtenidos de una manera útil para la toma de decisiones. Para ello

- Se introducirán los conceptos básicos de la Minería de Datos.
- Se analizarán e implementarán distintas técnicas que permitan construir modelos predictivos y descriptivos sobre datos de repositorios.
- Se resolverán problemas concretos haciendo hincapié en el análisis de los modelos obtenidos.

### MODALIDAD DE EVALUACION

Para obtener la cursada es necesario asistir al 80% de las clases, resolver, entregar y aprobar las actividades correspondientes al marco teórico respetando el cronograma indicado al inicio del curso.

Para aprobar el curso además deberá realizar un trabajo final integrador referido al análisis de un problema concreto (con datos reales o de repositorio) utilizando las distintas técnicas vistas en clase y elaborar un informe explicando los modelos y resultados obtenidos.



## PROGRAMA

- Introducción. Obtención de conocimiento a partir de los datos. El concepto de patrón. El proceso KDD. Fases del proceso de extracción del conocimiento. La Minería de Datos como fase del proceso KDD. Relación con otras disciplinas.
- Recuperación de información vs recuperación de datos. Proceso de recuperación de información.
- Preparación de Datos. Metadatos. Análisis de la información de entrada. Construcción y análisis de representaciones gráficas. Limpieza y transformación. Transformación y creación de atributos. Discretización y Numerización, Normalización de rango, escalado y centrado. Exploración mediante visualización y selección de datos.
- Técnicas de Minería de Datos. Extracción de Patrones. Introducción. Tareas y Métodos. Tareas predictivas y descriptivas. Aprendizaje supervisado y aprendizaje no supervisado. La Minería de Datos y el aprendizaje inductivo. Comparación de las técnicas de Minería de Datos.
- Árboles de decisión. Métricas de selección de atributos. Entropía. Ganancia de Información. Tasa de Ganancia. Índice Gini. Poda y Sobreajuste. Algoritmos Id3, C4.5 y Random Forest. Construcción de árboles para grandes volúmenes de datos.
- Reglas de clasificación. Partición vs cobertura. Métodos ZeroR, OneR, PRISM y PART. Métricas de una regla: soporte, cobertura, confianza, interés y convicción.
- Reglas de asociación. Calidad de las reglas. Algoritmo A priori. Concepto de ítem frecuente. Mejoras del algoritmo a priori: FP-Growth y FP-Tree.
- Técnicas de Agrupamiento. Métricas de calidad del agrupamiento. Tipos de agrupamiento: Jerárquico, partitivo y probabilista. Medidas de distancia y de conectividad. Proceso de agrupamiento. Clustering partitivo. Algoritmo k-medias. Algoritmos de clustering jerárquicos aglomerativos y divisivos. Dendrogramas. Algoritmo probabilista EM (Expectation - Maximization)
- Análisis y difusión del modelo obtenido. Evaluación de modelos. Comparación de técnicas de aprendizaje. Evaluación y mejora del modelo obtenido. Performance del modelo. Matriz de confusión. Sensibilidad, especificidad, precisión y recall. F-measure. Visualización utilizando curvas ROC.

## BIBLIOGRAFIA

- Hernández Orallo, Ramírez Quintana, Ferri Ramírez. Introducción a la Minería de Datos. Prentice Hall. 2004. ISBN 84-205-4091-9.



- Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, (Fourth Edition). Morgan Kaufmann. 2017. ISBN 978-0-12-804291-5
- Robert Layton. Learning Data Mining with Python. 2da.Edición. Packt Publishing Ltd, 2017. ISBN 9781787129566.
- Charu C. Aggarwal . Data Mining: The Textbook. Springer, 2015. ISBN 978-3319141428.
- Nong Ye. Data Mining: Theories, Algorithms, and Examples. CRC Press. 2013. ISBN 9781439808382
- Jiawei Han, Micheline Kamber, Jian Pei. Data Mining: Concepts and Techniques, (Third Edition). Morgan Kaufmann. 2013. ISBN-13: 978-0123814791.
- J. Ross Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann series in machine learning. Elsevier, 2014. ISBN 978- 0080500584.
- Valentina Porcu . Learning Python for data mining. StreetLib, 2017. ISBN 9788822803955