

**Programa para curso de postgrado según Ord. "C.S." Nº 291**

1- Institución:

*Facultad de Ciencias de la Administración*

2- Denominación del curso:

***Minería de Datos y Aprendizaje Automatizado***

3- Tipo de curso:

Curso de actualización, de capacitación, o de profundización (según Ord. C.S. Nº 291)

***Curso de profundización (nivel posgrado)***

4- Responsable académico:

***Dr. Carlos Iván Chesñear (Dep. Cs. e Ing. de la Computación, Universidad Nacional del Sur)***

5- Perfil de destinatarios:

***Graduados y docentes con titulación de grado interesados en la investigación en el Área de Informática***

6- Síntesis del curso:

***Datamining y el descubrimiento de conocimiento a partir de técnicas de aprendizaje automatizado se han transformado en un activa área de investigación, atrayendo a gente de distintas disciplinas (tales como base de datos, Inteligencia Artificial, Visualización de Información, etc.). Este curso introducirá los conceptos fundacionales de datamining y aprendizaje automatizado, estudiando sus principios, algoritmos, implementaciones y aplicaciones.***

7- Objetivos del curso:

*Los objetivos centrales de este curso son:*

- *Lograr que los alumnos manejen los conceptos centrales de datamining y aprendizaje automatizado, conociendo sus principios, algoritmos, implementaciones y aplicaciones principales.*
- *Desarrollar competencias para representar problemas en términos de los conceptos centrales de aprendizaje automatizado, evaluando la performance de distintos acercamientos, así como sus ventajas y desventajas.*
- *Desarrollar competencias para desenvolverse en la solución de problemas de aprendizaje automatizado y datamining utilizando la plataforma gratuita WEKA.*

8- Duración y carga horaria:

**90 Horas**

9- Programa:

*1. Introducción. Visión general del curso. ¿Qué es Datamining? Motivación. Perspectiva histórica. Algunas aplicaciones: análisis de mercados, detección de fraudes. Niveles de Datamining. Datamining como confluencia de disciplinas. Aprendizaje Automatizado y Datamining. Paradigmas de aprendizaje computacional. Técnicas de Datamining. Datamining predictivo vs. prescriptivo. Clasificación. Clustering. Regresión. Sintetización. Asociación. Ejemplos.*

*2. Preprocesamiento de Datos. Motivaciones. Data Cleaning. Integración, reducción y discretización de datos. Generación de Jerarquías de Conceptos. Operaciones de datamining.*

*3. Introducción al Aprendizaje Automatizado. Tipos de aprendizaje. Aprendizaje Inductivo a partir de Ejemplos. Lenguajes para expresar hipótesis. Bias en aprendizaje automatizado.*

*4. Reglas de Asociación. Introducción. Nociones preliminares: confianza y soporte de reglas.*

*Algoritmo A-Priori. Visualización de reglas de asociación. Medidas de interés. Críticas a los conceptos de soporte y confianza. Utilización de la Plataforma WEKA.*

*5. Clasificación y predicción de datos. Clustering. Introducción. Concepto. Componentes de la tarea de clustering. Definiciones. Notación. Técnicas de Clustering. Clasificación. Taxonomía de Acercamientos. Agrupamiento Conceptual. Técnicas de clasificación: árboles de decisión, redes neurales, clasificación bayesiana. Utilización de la Plataforma WEKA.*

*6. Árboles de Decisión. Algoritmos básicos. Criterios de Selección de atributos. Aprendizaje a partir de información con ruido. Poda de árboles de decisión. Cómo evitar el "overfitting" de datos. Bias inductivo en árboles de decisión. Especificidad en preferencia de hipótesis. Utilización de la Plataforma WEKA para obtención de árboles de decisión.*

*7. Aprendizaje basado en Instancias. Introducción. Aprendizaje por los K vecinos más cercanos. Algoritmos basados en pesos por distancias. Funciones de base radial. Razonamiento basado en casos. Ejemplos. Uso de WEKA para utilizar vecinos más cercanos.*

*8. Aprendizaje de Reglas. Reglas de Primer orden. Aprendizaje de conjuntos de reglas via FOIL. Inducción como Deducción Invertida. Resolución Invertida. Aprendizaje por*

*Refuerzo. Q-Learning. Recompensas y acciones no deterministas. Aprendizaje por diferencia temporal. Generalización a partir de ejemplos.*

*9. Integración de las distintas técnicas presentadas en el curso. Evaluación del error. Clases de error. Uso de matrices de costo para ponder el peso del error y definir estrategias de aprendizajes. Sistemas de agregación con metaclasificadores. Lift chart. Curva ROC. Evaluación con Voting, Boosting, Stacking.*

10- Bibliografía:

*H. Witten, E. Frank, M. Hall. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) Paperback – January 20, 2011*

*U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, 1996.*

*J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2011.*

*P. Flach. Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge Univ. Press, 2012. ISBN 978-1107422223*

*T. Imielinski and H. Mannila. A database perspective on knowledge discovery. Communications of ACM, 39:58-64, 1996.*

*G. Piatetsky-Shapiro, U. Fayyad, and P. Smith. From data mining to knowledge discovery: An overview. In U.M. Fayyad, et al. (eds.), Advances in Knowledge Discovery and Data Mining, 1-35. AAAI/MIT Press, 1996.*

*G. Piatetsky-Shapiro and W. J. Frawley. Knowledge Discovery in Databases. AAAI/MIT Press, 1991.*

*Quinlan, J.R. - C4.5: Programs For Machine Learning. San Mateo, CA: Morgan Kaufmann, 1993.*

*Mitchel, Tom – Machine Learning – McGraw Hill, 1997*

11. Metodología de trabajo:

*La materia se compone de un conjunto de clases teóricas a cargo del profesor en 9 encuentros intensivos de carácter interactivo de 5 horas cada uno. En estas clases se abordan los contenidos teóricos y algunos contenidos prácticos. Durante las clases se abordará también la resolución de ejercicios prácticos, vinculados al aprendizaje y utilización de la plataforma WEKA para problemas de datamining.*

12. Estrategias pedagógicas y comunicacionales:

*Para el desarrollo de los contenidos teóricos se utiliza una proyección multimedia y para la resolución práctica se utilizará un pizarrón convencional. Esto constituirá las 45 horas presenciales de los alumnos. Las 45 horas no presenciales estarán dedicadas a la resolución de ejercicios prácticos, interactuando con el profesor a través de la web (email) y al aprendizaje y utilización de la plataforma WEKA para problemas de datamining.*

13. Modalidad de evaluación

*Elaboración de un trabajo final y/o un proyecto.*

14. Criterios para la acreditación:

*Asistencia a instancias presenciales del 75%*

*Cumplimiento en tiempo y forma de actividades e instancias de evaluación.*

15. Inscripción:

Cantidad de alumnos mínimos y máximos

**Cupo máximo de 25 alumnos**

16. Certificación:

*El alumno deberá cumplir con la asistencia estipulada y aprobar la evaluación para recibir el certificado de asistencia y aprobación.*

17. Fecha propuesta:

**A convenir con la UNER (Fac. Cs. Admin) entre agosto y noviembre 2015.**