

Prontuario

- I. **Título del curso:** Minería de Datos
- II. **Codificación:** ESTA 5504
- III. **Horas / Crédito:** 3 horas semanales / 3 Créditos
- IV. **Prerequisitos:** ESTA 3042 Estadística para la Administración de Empresas II, MECU 3032 Métodos Cuantitativos para Administración de Empresas II o equivalentes. Estudiantes Graduados deberán contar con el permiso del Director.
- V. **Descripción del curso:**

Seminario de los conceptos fundamentales de minería de datos. Mediante el trabajo investigativo individual y colaborativo, se estudian las técnicas de aprendizaje automático que apoyan a la toma de decisiones al facilitar el análisis de grandes cantidades de datos. El contenido incluye técnicas de pre-procesamiento y visualización de datos, estudio y aplicación de técnicas de clasificación supervisada, clasificación no supervisada y métodos de detección de reglas de asociación. Se utilizan programas de aplicación que permiten al estudiante aplicar las técnicas estudiadas y el análisis de los resultados obtenidos. El estudiante preparará y presentará en forma escrita y oral, un proyecto de investigación donde aplique e integre conceptos del curso a un problema concreto.

Seminar on fundamental concepts of data mining. Through individual and collaborative research, students study machine learning techniques that support decision making by the analysis of large amounts of data. Course contents includes pre-processing and data visualization techniques, study and application of supervised and unsupervised classification techniques, and methods of discovery of association rules. Students apply data mining techniques and analyze the results using application software. The student will prepare and

¹ **Misión de la Facultad de Administración de Empresas**

Desarrollar líderes gerenciales, empresariales y académicos, profesionales y éticos, mediante una educación de excelencia e iniciativas de investigación y servicio en el contexto de Puerto Rico y el mundo.

² **Misión del programa de BAE en Estadística Aplicada**

El programa de Estadística Aplicada provee formación interdisciplinaria en los principios, metodologías y enfoques de la Estadística Aplicada, su fase computacional y la aplicación en escenarios diversos, en particular en la administración de empresas. De esta forma el programa contribuye a capacitar la sociedad en los enfoques analíticos para la investigación y la toma de decisiones con el fin de mejorar la calidad de vida de sus miembros.

present both in written and oral form, a research project in which studied concepts will be integrated and applied to a concrete problem.

VI. **Objetivos instruccionales:**

Al concluir el curso, el estudiante podrá:

- Describir conceptos y técnicas fundamentales en minería de datos.
- Aplicar técnicas de pre-procesamiento de datos, tales como tratamiento de datos faltantes, normalización, discretización, selección y extracción de variables.
- Aplicar técnicas para crear modelos predictivos por medio de técnicas de clasificación supervisada.
- Evaluar el rendimiento de los clasificadores.
- Aplicar algoritmos de clasificación no supervisada y técnicas de agrupamiento (“clustering”) a un conjunto de datos.
- Evaluar los resultados al aplicar algoritmos de clasificación no supervisada y técnicas de agrupamiento (“clustering”) a un conjunto de datos
- Aplicar distintos métodos estadísticos para la detección de datos anómalos.
- Usar técnicas para la detección de reglas de asociación para descubrir hechos que ocurren en común dentro de un conjunto de datos.
- Seleccionar el método de minería de datos adecuado para problemas específicos e información disponible.
- Usar programas de aplicación para minería de datos y evaluar los resultados obtenidos.
- Comunicar los resultados obtenidos correctamente y de una manera clara y organizada, escrita y verbal.
- Identificar implicaciones éticas y sociales del uso de minería de datos.
- Demostrar una actitud crítica hacia la aplicación de métodos de minería de datos en la solución de una diversidad de problemas.

VII. **Bosquejo de contenido:**

<i>Temas</i>	<i>Distribución de tiempo (horas)</i>
1 Introducción a. ¿Qué es minería de datos? b. Aplicaciones de minería de datos. c. Pasos en el proceso de minería de datos. d. Consideraciones éticas en minería de datos.	2
II. Pre-procesamiento de datos a. Tratamiento de datos faltantes b. Reducción de la dimensionalidad: i. Selección de variables ii. Extracción de variables	9

c. Discretización d. Normalización	
III. Visualización a. Plot de coordenadas paralelas b. "Surveyplots" c. Visualización radial d. Visualización "star-coordinates" e. "Heatmaps": Visualizar correlación y valores perdidos	6
IV. Clasificación Supervisada a. Clasificación Bayesiana b. Evaluación de un clasificador c. Clasificación usando vecinos cercanos (k-nn) d. Árboles de decisión e. Otros métodos de clasificación	9
V. Métodos de Agrupamiento ("Clustering") a. Métodos de Particionamiento b. Métodos Jerárquicos c. Métodos basados en densidad d. Mapas auto-organizantes (SOM) de Kohonen e. Métodos de agrupamiento basados en modelos	9
VI. Reglas de asociación a. Análisis de la canasta de Mercado b. El algoritmo a priori c. De "itemsets" frecuentes a reglas de asociación d. Minería de reglas de asociación multidimensional e. De reglas de asociación a análisis de correlación	7
Exámenes y presentaciones	3
Total	45

VIII. Estrategias de instrucción:

Seminarios de trabajo colaborativo en grupo, análisis de casos y discusiones en clase. Se asignarán trabajos de aplicación para hacer utilizando un programado estadístico. Se asignará un proyecto de clase como medio para aprender, hacer avalúo y evaluar lo aprendido. Se utilizarán otras estrategias instrucción tales como conferencias, simulación y uso de programado tal como SAS, R o XLMiner, trabajo en equipo, exposiciones orales, asignaciones.

IX. Recursos de aprendizaje:

Uso del programa R o algún otro programado estadístico tal como SPSS, SAS o XLMiner. El estudiante deberá tener acceso a una computadora personal, sea en un laboratorio de computadoras o en su casa. Salón de clases equipado con computadoras, pizarras, acceso al Internet y proyector que se pueda conectar a una computadora personal para desplegar visuales en una pantalla electrónica.

X. Estrategia de evaluación:

Exámenes 25%
 Proyecto, casos y asignaciones 50%

Examen Final 25%

De ser necesario, se realizará evaluación diferenciada a estudiantes con necesidades especiales.

XI. Estrategia de avalúo:

Se utilizarán estrategias de avalúo tal como pruebas cortas, supervisión de trabajo en grupo o en el computador, exámenes, tareas para realizar fuera del salón de clases que requieran el uso del computador, la participación en clase. Se asignará un proyecto de clase, con una rúbrica, como medio para aprender, hacer avalúo y evaluar lo aprendido.

XII. Sistema de calificación:

100-90%=A, 89-80%=B, 79-70%=C, 69-60%=D and 59-0%=F

XIII. Servicios educativos para personas con impedimentos, ley 51

Según la Ley de Servicios Educativos para Personas con Impedimentos (Ley 51 del 7 de junio de 1996), todo estudiante que requiera acomodo razonable deberá notificarlo al profesor el primer día de clases.

Los estudiantes que reciban servicios de **Rehabilitación Vocacional** deben comunicarse con el profesor al inicio del semestre para planificar el acomodo razonable y el equipo asistido necesario conforme a las recomendaciones de la Oficina de Asuntos para las Personas con Impedimento (OAPI) del Decanato de Estudiantes.

XIV. Conducta estudiantil sujeta a sanciones disciplinarias:

Los actos de deshonestidad académica están sujetos a sanciones disciplinarias, según establece el Reglamento General de Estudiantes de la Universidad de Puerto Rico, Certificación 13, 2009-2010, Parte VI, Artículo 6.2.³

No se permite en momento alguno el uso de teléfonos celulares o cualquier otro artefacto electrónico no autorizado previamente. El profesor podrá tomar las medidas disciplinarias que considere pertinentes para evitar su uso.

XV. Bibliografía :

Texto:

Shmueli, Galit, Patel, Nitin and Bruce, Peter. 2010. *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*. s.l. : Wiley, 2010.

³ El reglamento incluye ejemplos de actos sujetos a sanción tal como: acciones fraudulentas, la obtención de notas o grados académicos valiéndose de falsas o fraudulentas simulaciones, copiar total o parcialmente la labor académica de otra persona, plagiar total o parcialmente el trabajo de otra persona, copiar total o parcialmente las respuestas de otra persona a las preguntas de un examen, haciendo o consiguiendo que otro tome en su nombre cualquier prueba o examen oral o escrito, así como la ayuda o facilitación para que otra persona incurra en la referida conducta.

Referencias:

Bramer, Max. 2007. *Principles of Data Mining (Undergraduate Topics in Computer Science)*. s.l. : Springer, 2007.

Hancock, Monte. 2011. *Practical Data Mining*. s.l. : Auerbach Publications, 2011.

Nisbet, Robert, Elder, Jhon and Miner, Gary. 2009. *Handbook of Statistical Analysis and Data Mining Applications*. s.l. : Academic Press, 2009.

Williams, Graham. 2011. *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*. s.l. : Springer, 2011.

Fernandez, George, 2010. *Statistical Data Mining Using SAS Applications, Second Edition* CRC Press; 2 edition (June 18, 2010)

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

Torgo, Luis. 2010. *Data Mining with R: Learning with Case Studies*. s.l. : Chapman and Hall/CRC, 2010.

Hastie, Trevor, Tibshirani, Robert and Friedman, Jerome. 2010. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. s.l. : Chapman and Hall/CRC, 2010.

Referencias Electrónicas:

Curso de data mining desarrollado por Gregory Piatetsky-Shapiro y Gary Parker
http://www.kdnuggets.com/data_mining_course/index.html

Curso en línea, de introducción a la minería de datos, desarrollado por Ruth Dilly, de la Universidad de Queens. http://www.pcc.qub.ac.uk/tec/courses/datamining/ohp/dm-OHP-final_1.html

“The Knowledge Discovery Mine”, enlaces útiles a otros recursos, publicaciones relacionadas a minería de datos, enlaces a repositorios de datos y un catálogo de software para minería de datos. <http://www.kdnuggets.com/index.html>

Repositorio de conjuntos de datos

<http://archive.ics.uci.edu/ml/>