



Universidad de Puerto Rico  
Recinto Universitario de Mayagüez  
Colegio de Artes y Ciencias  
Departamento de Ciencias Matematicas



## ESMA 4016 MINERIA DE DATOS Y APENDIZAJE AUTOMATICO

<b>Horas crédito:</b> 3	<b>Horas contacto:</b> 3 horas de conferencia semanales	
<b>Requisitos previos:</b> (ESMA 3016 y MATE 4031) o Permiso Director		<b>Requisitos concurrentes:</b>
<b>Descripción del curso (español):</b> Introducción a las principales técnicas de minería de datos y aprendizaje automático aplicadas al aprendizaje supervisado y no supervisado. Entre los tópicos que se incluyen están regresión y clasificación, métodos de vecinos más cercanos, arboles de decisión, repotenciar, redes neurales y maquinas de vectores de soporte.		
<b>Descripción del curso (inglés):</b> Introduction to the major techniques for data mining and machine learning applied to both supervised and unsupervised learning. Topics include regression and classification, nearest neighbors methods, decision trees, boosting, neural networks and support vector machines		
<b>Objetivos del curso:</b> Entender los conceptos fundamentales para llevar a cabo minería de datos y descubrimiento de conocimiento en base de datos usando métodos de Machine Learning. Experimentar algunos algoritmos más usados en minería de datos y Machine Learning en conjuntos de datos reales. Usar Python en Jupyter notebook junto con keras y tensorflow para ejecutar los algoritmos aprendidos.		

Profesor: Edgar Acuña  
Oficina:  
Extension:  
Horas de Oficina: Lunes y Miercoles de 1pm a 4pm  
Github del curso: <https://github.com/eacunafer/Data-Mining-Machine-Learning-subgraduado->  
También se pondrá material en piazza.com

### Contenido del curso:

<i>Temas a cubrir</i>	<i>Horas contacto</i>
Introduction to data mining and Machine Learning	2
Data Cleaning	3

Feature Engineering: Binning, Feature Selection and Principal Component Analysis	4										
Data Visualization	2										
Supervised Learning: Regression	3										
Supervised Learning: Logistic Regression	2										
Review of Probability	1										
Supervised learning: Linear Discriminant Analysis	2										
Estimation of Prediction Error	1										
Supervised Learning: Naïve Bayes	2										
K Nearest Neighbors	3										
Decision Trees	3										
Ensembles	1										
Random Forest and XGBoost	2										
Neural Networks and Deep Learning	4										
Unsupervised Clustering: Partitional Algorithms	3										
Unsupervised Clustering: Hierarchical Algorithms	2										
Cluster Validation	1										
Recommendation Systems	2										
<i>Total de horas: (deben ser equivalentes a las horas crédito del curso)</i>	45										
<b>Estrategias instruccionales:</b> <input checked="" type="checkbox"/> conferencia <input type="checkbox"/> discusión <input type="checkbox"/> cómputos <input type="checkbox"/> laboratorio  <input type="checkbox"/> investigación <input type="checkbox"/> otros, especifique:											
<b>Recursos mínimos disponibles:</b> materiales y equipo necesarios para cumplir los objetivos del curso											
<b>Estrategias de evaluación y su peso relativo:</b> <table border="1" data-bbox="500 1161 1149 1350"> <thead> <tr> <th></th><th>Porcentaje</th></tr> </thead> <tbody> <tr> <td><input type="checkbox"/> examen- Take home</td><td>30%</td></tr> <tr> <td><input type="checkbox"/> tareas(2-3)</td><td>40%</td></tr> <tr> <td><input type="checkbox"/> proyecto</td><td>30%</td></tr> <tr> <td><b>TOTAL: 100%</b></td><td></td></tr> </tbody> </table>			Porcentaje	<input type="checkbox"/> examen- Take home	30%	<input type="checkbox"/> tareas(2-3)	40%	<input type="checkbox"/> proyecto	30%	<b>TOTAL: 100%</b>	
	Porcentaje										
<input type="checkbox"/> examen- Take home	30%										
<input type="checkbox"/> tareas(2-3)	40%										
<input type="checkbox"/> proyecto	30%										
<b>TOTAL: 100%</b>											
<b>Texto:</b> 1-James, Witten, Tibshirani & Hastie. 2017. Introduction to Statistical Learning Springer, 4 <a href="https://github.com/eacunafer/Data-Mining-Machine-Learning-subgraduado/blob/master/books/ISLR7.pdf">Programas en R. Disponible en el github del curso</a> <a href="https://github.com/eacunafer/Data-Mining-Machine-Learning-subgraduado/blob/master/books/ISLR7.pdf">https://github.com/eacunafer/Data-Mining-Machine-Learning-subgraduado/blob/master/books/ISLR7.pdf</a>  2- Sebastian Raschka Vahid Mirjalil. 2017. Python Machine Learning Second Edition. Disponible en el github del curso. <a href="https://github.com/eacunafer/Data-Mining-Machine-Learning-subgraduado/blob/master/books/python-machine-learning-2nd.pdf">https://github.com/eacunafer/Data-Mining-Machine-Learning-subgraduado/blob/master/books/python-machine-learning-2nd.pdf</a>  3-Laython, R. Learning Data Mining with Python (2015). Disponible en el github del curso											

<https://github.com/eacunafer/Data-Mining-Machine-Learning-subgraduado/blob/master/books/RLDMPy.pdf>

4. Geron Aurelien. (2019). Hands on Machine Learning with scikit-learn, Keras and tensorflow

<https://github.com/changwookjun/StudyBook/blob/master/DeepLearningBooks/Hands-on-Machine-Learning-with-Scikit-2E.pdf>

### Referencias:

1- Hastie, T., Tibshirani, R. and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd Edition. Springer, 2009.

[https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII\\_print12.pdf](https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf)

2- Alpaydin, E. Introduction to Machine Learning. Third Edition. MIT Press, 2014

[http://cs.du.edu/~mitchell/mario\\_books/Introduction\\_to\\_Machine\\_Learning\\_-\\_2e\\_-\\_Ethem\\_Alpaydin.pdf](http://cs.du.edu/~mitchell/mario_books/Introduction_to_Machine_Learning_-_2e_-_Ethem_Alpaydin.pdf)

3-A Complete Tutorial to Learn Data Science with Python from Scratch

<https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/>

### Videos:

1-Clases del Profesor Nando Freitas (British Columbia University)

<http://www.cs.ubc.ca/~nando/340-2012/lectures.php>

[https://www.youtube.com/playlist?list=PLE6Wd9FR--Ecf\\_5nCbnSQMHqORpiChfJf&feature=view\\_all](https://www.youtube.com/playlist?list=PLE6Wd9FR--Ecf_5nCbnSQMHqORpiChfJf&feature=view_all)

2-Clases de Machine Learning del Profesor Andrew Ng en Stanford: CS 209

-<http://robotics.stanford.edu/~ang/courses.html>

### Sistema de calificación:

☒cuantificable (de letra) ☐no cuantificable

Curva estándar

100-90 A; 89-80 B; 79-70 C; 69-60 D; 59-0 F

Después de identificarse con el profesor y la institución, los estudiantes con impedimento recibirán acomodo razonable en sus cursos y evaluaciones. Para más información comuníquese con Servicios a Estudiantes con Impedimentos en la Oficina del Decano de Estudiantes (Oficina #1) o a los teléfonos 787-265-3864 ó 787-832-4040 x 3772, 2040 y 3864.

EAF/Enero 2021