	<b>Nombre del curso: APRENDIZAJE AUTOMÁTICO</b>		<b>Semestre A– 2021</b>			
	<b>Profesores</b>	PABLO GUILLEN RONDON		<a href="mailto:guillenpablo1@gmail.com">guillenpablo1@gmail.com</a>		
	<b>Coordinador:</b>	VICTOR ROMERO CANO		<a href="mailto:varomero@uao.edu.co">varomero@uao.edu.co</a>		
<b># Créditos</b>		<b>4</b>	<b>Hora presenciales</b>	<b>48</b>	<b>Horas de trabajo independiente</b>	<b>144</b>

### Descripción del curso


Machine learning is the science of developing statistical methods that quantify relationships within data. This branch of mathematics/computer science has seen an explosive growth over the past decade as our ability to store and process digital data has dramatically increased. Prediction, classification, regression, and identification are the aims of learning from data. All of these problems are routinely performed in data analytics.

### Objetivos del curso

- To obtain an overview of the literature in learning-based methods and applications.
- To obtain an understanding of a variety of machine learning techniques for classification, regression, and prediction.
- To obtain the ability to implement and experiment with a wide range of machine learning algorithms in Python with examples.
- To apply: Unsupervised and supervised learning and clustering concepts, dimensionality reduction, Kernels and kernel-based classifiers such as SVM, and Deep Learning algorithms.
- To understand and implement learning-based methods for detection, diagnosis, and classification of data coming from biomedical and oil/gas systems.

### Contenidos

- Introduction to data mining
- Introduction to exploratory data analysis
- Featurization: feature extraction, transformation, dimensionality reduction, and selection
- Pipelines: tools for constructing, evaluating, and tuning ML Pipelines
- Introduction to unsupervised and supervised learning
  - Concepts and Definitions
- Unsupervised learning
  - Clustering
  - Principal Component Analysis
- Supervised machine learning techniques
  - K-Nearest Neighbors
  - Decision Tree
  - Random Forest
  - XGBoost

	<b>Nombre del curso: APRENDIZAJE AUTOMÁTICO</b>		<b>Semestre A– 2021</b>			
	<b>Profesores</b>	PABLO GUILLEN RONDON		<a href="mailto:guillenpablo1@gmail.com">guillenpablo1@gmail.com</a>		
	<b>Coordinador:</b>	VICTOR ROMERO CANO		<a href="mailto:varomero@uao.edu.co">varomero@uao.edu.co</a>		
<b># Créditos</b>		<b>4</b>	<b>Hora presenciales</b>	<b>48</b>	<b>Horas de trabajo independiente</b>	<b>144</b>

- Support Vector Machine
- Neural Networks
- Support vector machine and optimization of parameters
  - Linear, Polynomial and Gaussian Radial Kernels
  - Grid search
- Deep learning algorithms
  - Concepts and Definitions
  - Autoencoders, CNN, RNN
  - Deep Learning based on a multi-layer feedforward artificial neural network
  - Deep Learning based on Convolutional Neural Networks
- Auto machine learning systems
  - Concepts and Definitions
  - auto-sklearn
  - auto Keras
- Examples with machine learning frameworks applied to oil/gas:
  - Well logs prediction
  - Time series prediction
  - Lithofacies classification
  - Seismic image classification
- Examples with machine learning frameworks applied to biomedical systems:
  - Single-Cell RNA Sequencing in Cancer - Classification
  - Alzheimer image classification
  - MRI image classification
  - MRI, CT-Scan image segmentation

### **Metodología**


---

The emphasis in this course will be focus on theory and a series of practical case studies, **Hands on**. During the course the participants will be able to use Python notebooks to (1) load a data set from a file, (2) prepare the data for analysis, (3) build and tune an appropriate machine learning algorithm, (4) assess the future performance of the algorithm on new data, and (5) apply the algorithm on a new dataset to predict the outcome. Python notebooks used in the course will make use of software tools (modules and functions) available in the Scikit-Learning and Keras libraries.

### **Lengua extranjera**

---

- Professor Pablo Guillen will teach 20% of his classes in the English language.


	<b>Nombre del curso: APRENDIZAJE AUTOMÁTICO</b>		<b>Semestre A– 2021</b>			
	<b>Profesores</b>	PABLO GUILLEN RONDON	<a href="mailto:guillenpablo1@gmail.com">guillenpablo1@gmail.com</a>			
	<b>Coordinador:</b>	VICTOR ROMERO CANO	<a href="mailto:varomero@uao.edu.co">varomero@uao.edu.co</a>			
<b># Créditos</b>		<b>4</b>	<b>Hora presenciales</b>	<b>48</b>	<b>Horas de trabajo independiente</b>	<b>144</b>

- 100% of the class material: Slides, Notes, papers, documentation of the software platforms will be in the English language.
- At the end of the course, in the final project, each student will present their results of the research carried out in a 2-page report following the IEEE format, the writing of this report will be in the English language.

Component (hours)	# total hours	Weight of component	# weighted hours
Oral component: speak in english and to interact with students.	9.6	70%	6.72
Written component: slides, notes, papers, etc.	48	30%	14.4
			21.12

### Innovación y el emprendimiento de base tecnológica

- Both data mining and machine learning are rooted in data science and generally fall under that umbrella.
  - One key difference between machine learning and data mining is how they are used and applied in our everyday lives. For example, data mining is often used by machine learning to see the connections between relationships.
  - Machine learning embodies the principles of data mining but can also make automatic correlations and learn from them to apply to new algorithms.
- 
- Machine learning, on the other hand, can actually learn from the existing data and provide the foundation necessary for a machine to teach itself.
  - Machine learning can look at patterns and learn from them to adapt behavior for future incidents, while data mining is typically used as an information source for machine learning to pull from. Although data scientists can set up data mining to automatically look for specific types of data and parameters, it doesn't learn and apply knowledge on its own without human interaction. Data mining also can't automatically see the relationship between existing pieces of data with the same depth that machine learning can.
  - It's true that data mining can reveal some patterns through classifications and sequence analysis. However, machine learning takes this concept a step further by

	<b>Nombre del curso: APRENDIZAJE AUTOMÁTICO</b>		<b>Semestre A– 2021</b>			
	<b>Profesores</b>	PABLO GUILLEN RONDON		<a href="mailto:guillenpablo1@gmail.com">guillenpablo1@gmail.com</a>		
	<b>Coordinador:</b>	VICTOR ROMERO CANO		<a href="mailto:varomero@uao.edu.co">varomero@uao.edu.co</a>		
<b># Créditos</b>		<b>4</b>	<b>Hora presenciales</b>	<b>48</b>	<b>Horas de trabajo independiente</b>	<b>144</b>

using the same algorithms data mining uses to automatically learn from and adapt to the collected data.

- The future is bright for data science as the amount of data will only increase. As we amass more data, the demand for advanced data mining and machine learning techniques will force the industry to evolve in order to keep up. We'll likely see more overlap between data mining and machine learning as the two intersect to enhance the collection and usability of large amounts of data for analytics purposes.

### **Libro texto y lecturas recomendadas**

Machine Learning in Python®: Essential Techniques for Predictive Analysis. Michael Bowles. Published by John Wiley & Sons, Inc. 2015

Building Machine Learning Systems with Python. Willi Richert and Luis Pedro Coelho. Published by Packt Publishing Ltd. 2013.

Deep Learning: <http://www.deeplearningbook.org/contents/TOC.html>

Links sobre automated machine learning (automl):

<https://automl.github.io/auto-sklearn/master/>

<https://autokeras.com/>

Notes and Papers published by the Instructors:

[http://www.researchgate.net/profile/Pablo\\_Guillen3](http://www.researchgate.net/profile/Pablo_Guillen3)

<https://scholar.google.com/citations?user=x3M1JIAAAAAJ&hl=es>


### **Requerimientos previos**

Basic knowledge of Python.

### **Aclaraciones adicionales del curso**

Each Professor will teach 8 weeks of classes.

### **Computational Tools:**

	<b>Nombre del curso: APRENDIZAJE AUTOMÁTICO</b>		<b>Semestre A– 2021</b>
	<b>Profesores</b>	PABLO GUILLEN RONDON	<a href="mailto:guillenpablo1@gmail.com">guillenpablo1@gmail.com</a>
	<b>Coordinador:</b>	VICTOR ROMERO CANO	<a href="mailto:varomero@uao.edu.co">varomero@uao.edu.co</a>
<b># Créditos</b>		<b>4</b>	<b>Hora presenciales</b> <b>48</b> <b>Horas de trabajo independiente</b> <b>144</b>

- **Anaconda** (free and open distribution) platform will be used in this course.
- **Python** 3.7 language will be used.
- Python notebooks (over 2000 lines of code) developed by the Instructors will be delivered to the participants

### Criterio de evaluación

Evaluación	Porcentaje
3 homework assignments (10% each)	30 %
<hr/>	
Evaluación para registrar en el primer corte de calificaciones	30 %
4 homework assignments (10% each)	40 %
Final project	20 %
Final Exam	10 %
Evaluación para registrar en el segundo corte de calificaciones	70 %