



## Machine Learning Techniques – MLT

**MISIS 4219**

**Semestre 2024-10**

**Lunes 18h30 – 21h15**

**Profesores:** Haydemar Nuñez [h.nunez@uniandes.edu.co](mailto:h.nunez@uniandes.edu.co)  
Rubén Manrique [rf.manrique@uniandes.edu.co](mailto:rf.manrique@uniandes.edu.co)

**Monitor:** Héctor Julio Melo

### 1. Descripción general del curso

El aprendizaje automático (machine learning) es un área de la inteligencia artificial (IA) que se dedica al estudio y desarrollo de algoritmos para dotar a una máquina de la capacidad de aprender de forma autónoma a partir de datos. Actualmente, es parte integral de muchos proyectos de investigación y productos comerciales en una gran diversidad de áreas. Un aspecto clave para el desarrollo de aplicaciones exitosas basadas en este paradigma de la IA, es disponer de un conocimiento sólido sobre cómo funcionan los algoritmos de aprendizaje y cuáles son los aspectos esenciales que hay que tener en cuenta para la construcción de modelos.

En este curso se abordan los fundamentos y algunas técnicas del machine learning, con énfasis en las redes neuronales artificiales y el aprendizaje profundo (*deep learning*). Igualmente, trata sobre cómo utilizar, de manera efectiva, algoritmos de aprendizaje en una diversidad de dominios, tomando en cuenta la naturaleza del problema y los recursos informáticos disponibles.

### 2. Objetivos pedagógicos.

Se espera desarrollar y reforzar las siguientes competencias:

- **Comprender** los fundamentos básicos del aprendizaje a partir de datos.
- **Reconocer** diferentes contextos de aplicación de las técnicas de aprendizaje, con base en la naturaleza del problema y la representación de los datos.



- **Entender** cómo funcionan algunas técnicas de aprendizaje en la construcción de modelos de conocimiento a partir de datos.
- **Comprender** qué es el aprendizaje de profundo y cómo se relaciona con las redes neuronales artificiales.
- **Analizar** el funcionamiento de algunas arquitecturas de aprendizaje profundo y cómo pueden ser aplicadas para resolver problemas complejos.
- **Aplicar**, de manera efectiva, algoritmos de aprendizaje automático para resolver problemas en diferentes dominios siguiendo el proceso de aprendizaje a partir de datos.
- **Reconocer** las implicaciones éticas en el desarrollo de soluciones a partir de datos.

### 3. Plan de temas.

**Tema 1.** Principios básicos del aprendizaje artificial. ¿Qué es un algoritmo de aprendizaje? ¿Qué significa aprender de los datos? Elementos de un sistema de aprendizaje. Tipos de aprendizaje: supervisado, no supervisado, por refuerzo. Conceptos de generalización, sobreajuste y complejidad, dilema (*trade-off*) bias-varianza, regularización e hiperparámetros. Técnicas de selección de modelos, métricas de evaluación y curvas de aprendizaje. Proceso de aprendizaje a partir de datos.

**Tema 2.** Estudio de algunos algoritmos de aprendizaje. Regresión logística, máquinas de vectores de soporte, métodos bayesianos y no supervisados.

**Tema 3.** Redes neuronales artificiales (ANN). Características de una red neuronal, funciones de activación, modelos supervisados lineales y no lineales (y los algoritmos de aprendizaje asociados). Aspectos de diseño y control de la complejidad. Modelos no supervisados.

**Tema 4.** Redes neuronales profundas (DNN). ¿Qué es aprendizaje profundo? Redes neuronales convolucionales, modelos secuenciales, arquitecturas encondedecoder y transformadores. Tendencias y aplicaciones actuales del machine learning.

#### Conocimientos previos recomendados.

Nivel intermedio de programación (se recomienda Python). Conocimientos intermedios de álgebra lineal, probabilidad y estadística, y optimización numérica.



#### 4. Metodología

- Clases teóricas dirigidas a la adquisición de conceptos, con discusión de casos de estudio y aplicaciones.
- Actividades prácticas centradas en la experimentación con conjuntos de datos de repositorios públicos, utilizando el lenguaje de programación Python, junto con librerías de *machine learning* y *deep learning*. A lo largo del curso se desarrolla un proyecto definido por los estudiantes.

#### 5. Porcentaje de notas

- Examen 1: 20% (Individual)
- Examen 2: 20% (Individual)
- Laboratorios: 30% (Grupal, 2 personas por grupo)
- Proyecto Final: 30% (Grupal, máximo 4 personas – mínimo 3)

#### Proyecto Final:

- **Entrega 1** (marzo 11): Documento donde se especifique el objetivo del proyecto, justificación, idea clara de los datos que se van a emplear y técnica de extracción si aplica. También se deberá incluir en conjunto con la propuesta, tres resúmenes de artículos académicos relacionados a su proyecto que les parezcan interesantes y de los cuales puedan extraer ideas.
- **Entrega 2** (Semana examen finales). Entregables:
  - Se debe entregar un documento estilo artículo del proyecto.
  - Código fuente del proyecto (notebooks) con una guía de entrenamiento y posterior ejecución.
  - Demostración/despliegue de su modelo en aplicación.
  - Video de 10 min explicando su proyecto.

Sustentación presencial en horarios a convenir.



## 6. Cronograma

Semana	Fecha	Tema	Actividad (a realizar/ entregar)
1	Enero 22	Fundamentos del aprendizaje artificial. ¿Qué es un algoritmo de aprendizaje?  ¿Qué significa aprender de los datos? Elementos de un sistema de aprendizaje. Tipos de aprendizaje: supervisado, no supervisado, por refuerzo.  Regresión Lineal.	
2	Enero 29	Discriminantes lineales. Regresión logística.	Lab 1 Introducción a pandas, numpy, matplotlib y scikitlearn. Regresión Lineal y Logística.
3	Febrero 05	Conceptos de generalización, sobreajuste y complejidad, trade-off bias-varianza, regularización e hiperparámetros.  Métricas de evaluación. Técnicas de selección de modelos y curvas de aprendizaje.	Lab 2 Sobreajuste, y regularización. Métricas en aprendizaje supervisado.
4	Febrero 12	Naive Bayes Métodos generativos vs métodos discriminatorios. La ingeniería de características. Estrategias para extracción de características: PCA	Lab 3 Naive Bayes para tareas de clasificación y PCA.
5	Febrero 19	Máquinas de vectores de soporte.	Lab 4 SVM
6	Febrero 26	Árboles y Métodos Ensamblados (Random forest), XGBoost  Estrategias para selección de características: SHAP, <i>Permutation feature importance</i>	Lab 5. Árboles y métodos ensamblados.
7	Marzo 04	Métodos de aprendizaje no supervisado: K-means, Jerárquico, K-medoids.	Lab 6. Aprendizaje no supervisado.
8	Marzo 11	<b>Examen 1</b>	
		<b>Receso+Semana Santa</b>	<b>receso</b>
9	Abril 01	Introducción a redes neuronales y aprendizaje profundo.  Como se entrena una red neuronal, problema del gradiente desvaneciente.	Lab 7: Intro a Tensorflow, Keras



		Introducción a Tensor-Flow y Keras. Como construir el grafo de cómputo y arquitecturas densas.	
10	Abril 08	Redes convolucionales. Kernels, Pooling. Aumentación de datos.	Lab 8: Redes Convolucionales. Clasificación de imágenes médicas.
11	Abril 15	Modelos pre-entrenados. Transferencia de aprendizaje.	Lab 9: Transferencia del aprendizaje.
12	Abril 22	Procesamiento de Lenguaje Natural. Modelos secuenciales. Redes recurrentes, LSTM y GRU.	Lab 10: Generación de secuencias
13	Abril 29	Incrustaciones de texto: Vectores dispersos y Densos. Word2Vec. Incrustaciones contextuales: BERT, ROBERTA, XLNet	
14	Mayo 06	Transformers, Modelos de lenguaje, GPT-3	Lab 11: Transformers (posible Workshop Nvidia)
15	Mayo 13	Festivo	
16	Mayo 20	<b>Examen 2</b>	
17	Mayo 27-31	<b>Sustentaciones de proyectos finales</b>	

## 7. Aspecto Académicos.

### a) Generalidades.

- Clases: 3 horas semanales, de las cuales se espera la participación activa de los estudiantes.
- El curso tiene como canales oficiales de comunicación el correo electrónico Uniandes, la lista de correo del curso, y el sistema de apoyo a la docencia BloqueNeón (<http://bloqueneon.uniandes.edu.co>).

### b) Política de aproximación de notas finales.

- Las notas del curso varían entre 0.00 y 5.00 con dos decimales y no hay aproximaciones.
- Para pasar el curso es indispensable lograr en el puntaje ponderado 3.00 o superior.
- No existe aproximación automática en la nota definitiva. En particular, no hay aproximación a 3.00 de puntajes menores a esta nota (e.g., 2.9 no es 3.00).

### c) Protocolo MAAD.

El miembro de la comunidad que sea sujeto, presencie o tenga conocimiento de una conducta de maltrato, acoso, amenaza, discriminación, violencia sexual o de género (MAAD) deberá poner el caso en conocimiento de la Universidad. Ello, con el propósito de que se puedan tomar acciones institucionales para darle manejo al caso, a la luz de lo previsto en el protocolo, velando por el bienestar de las personas afectadas.

Para poner en conocimiento el caso y recibir apoyo, usted puede contactar a:

1. Ombudsperson: [ombudsperson@uniandes.edu.co](mailto:ombudsperson@uniandes.edu.co)
2. Decanatura de Estudiantes: [centrodeapoyo@uniandes.edu.co](mailto:centrodeapoyo@uniandes.edu.co)
3. Red de Estudiantes: - PACA: [paca@uniandes.edu.co](mailto:paca@uniandes.edu.co)
4. Consejo Estudiantil Uniandino (CEU): [comiteacosoceu@uniandes.edu.co](mailto:comiteacosoceu@uniandes.edu.co)

## 8. Bibliografía

- Geron, A. (2019). "Hands-On Machine Learning with Scikit-Learn and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems".
- Sebastian Raschka, Vahid Mirjalili (2019). Python Machine Learning. Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. Packt Publishing, tercera edición (la segunda edición se encuentra en la biblioteca).
- Prosise, J. (2022). "Applied Machine Learning and AI for Engineer". O'Reilly.
- Müller, A., Guido, S. (2016). Introduction to Machine Learning with Python\_ A Guide for Data Scientists. O'Reilly Media.
- Ahmed Menshawy (2018). Deep Learning by Example: A Hands-on Guide to Implementing Advanced Machine Learning Algorithms and Neural Networks. Packt Publishing.
- Aggarwal, C. (2018). Neural Networks and Deep learning. A textbook. Springer.
- Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press.