

Programa del curso

1. Descripción general del curso.

El aprendizaje automático (*machine learning*) se define como la adquisición automática de conocimiento a partir de la experiencia. Es un campo de investigación muy activo, debido a la calidad de los resultados que se han reportado en la resolución de problemas complejos y a la diversidad de áreas donde puede ser utilizado. La premisa de aplicación del machine learning es que no se dispone de un conocimiento explícito que permita derivar una solución analítica para un problema, pero sí de un conjunto de datos a partir del cual es posible inducir o aprender una solución empírica.

Por otra parte, los progresos en las tecnologías de adquisición y almacenamiento de grandes volúmenes de información procedentes de una diversidad de fuentes han determinado el surgimiento y la disponibilidad de datos masivos, heterogéneos y no estructurados. Debido a la magnitud y complejidad de estos repositorios, la información útil contenida en ellos no es fácil de analizar con métodos tradicionales. Surge entonces la necesidad de técnicas y herramientas para el análisis inteligente de la información.

En este curso se ilustra cómo las técnicas de machine learning pueden ser utilizadas en el análisis de información mediante la construcción de modelos que apoyen los procesos de toma de decisiones, tomando en consideración los objetivos planteados, así como las características y complejidad de las representaciones de los conjuntos de datos disponibles a partir de diferentes fuentes.

2. Objetivos de aprendizaje.

El objetivo general de este curso es promover las competencias necesarias para una aplicación efectiva de las técnicas de machine learning en la construcción de modelos de conocimiento que apoyen la toma de decisiones. Así, se espera desarrollar y reforzar las siguientes competencias:

- Diferenciar los tipos de problemas que pueden ser resueltos con las técnicas de machine learning.
- Analizar el proceso de aprendizaje a partir de datos, desde la formulación de los objetivos del negocio hasta la explotación del conocimiento inducido.
- Aplicar técnicas de machine learning, supervisadas y no supervisadas, para construir soluciones alineadas con los objetivos de negocio y las características de los datos disponibles.
- Seleccionar las técnicas y medidas de evaluación que pueden ser utilizadas para determinar la validez de los modelos construidos con métodos de aprendizaje.

- Utilizar librerías y herramientas de software para la resolución de problemas de aprendizaje a partir de datos.
- Explicar los resultados obtenidos con técnicas de machine learning.
- Identificar las limitaciones y riesgos en el uso de datos para desarrollar aplicaciones basadas en machine learning, garantizando el uso ético y transparente de estos.

3. Plan de temas.

Fundamentos de machine learning (ML). Tipos de aprendizaje y contextos de aplicación. Arquitectura de un sistema de aprendizaje. Técnicas de aprendizaje y representación de modelos. Conceptos de generalización, complejidad de modelos y sobreajuste. Trade-off bias-varianza. Técnicas de regularización. Selección de modelos y técnicas de validación.

Diseño de sistemas de aprendizaje. ¿Cómo implementar un proyecto de machine learning? Proceso de aprendizaje a partir de datos: el ciclo de machine learning, desde la definición del problema hasta el despliegue y monitorización de los modelos. Sesgos en ML. Interpretabilidad de modelos. Ética de la inteligencia artificial.

Preparación de los datos. Tipos de conjuntos de datos y su representación para el proceso de aprendizaje. Calidad de datos y técnicas de limpieza de datos. Ingeniería de características. Reducción de la dimensionalidad.

Aprendizaje supervisado. Análisis de tareas predictivas. Evaluación de modelos predictivos. Algunas técnicas de aprendizaje: basadas en árboles de decisión, basadas en instancias, basadas en probabilidades, basadas en funciones, redes neuronales artificiales. Combinación de clasificadores (ensembles). Clasificación en contextos no balanceados.

Aprendizaje no supervisado. Análisis de tareas descriptivas. Agrupación: conceptos básicos. Aspectos de diseño. Algunas técnicas de aprendizaje: por particiones, jerárquicas, por densidad. Evaluación e interpretación de modelos de agrupación. Análisis de textos y modelos de tópicos. Detección no supervisada de anomalías.

Introducción al deep learning. Presentación de algunas arquitecturas básicas. Tendencias y aplicaciones actuales del machine learning.

Conocimientos previos recomendados.

Nivel intermedio de programación (se recomienda Python). Conocimientos básicos 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*. A lo largo del curso se desarrolla un proyecto definido por los estudiantes.

5. Bibliografía

- Geron, A. (2022). "Hands-On Machine Learning with Scikit-Learn and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems “. O'Reilly Media.
- Bruce, P. Bruce, A. Gedeck, P. (2020). Practical Statistics for Data Scientists. 50+ Essential Concepts Using R and Python. Second edition. O'Reilly Media.
- 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).
- Kuhn, M., Johnson, K. (2020). Feature Engineering and Selection. A Practical Approach for Predictive Models. CRC press.
- Ozdemir, Sinan; Susarla, Divya (2018). Feature Engineering Made Easy: Identify Unique Features from your Dataset in order to bBuild powerful machine learning systems. Packt Publishing.
- Pruksachatkun, Y., Mcateer, M., Majumdar, S. (2022). Practicing Trustworthy Machine Learning. O'Reilly
- Artículos de revistas y congresos especializados.

6. Evaluación y aspecto Académicos.

A) Evaluación.

Exámenes	Se realizarán exámenes teóricos sobre los temas vistos en clase.	35%
Laboratorios	Se entregará un problema después de cada laboratorio, con el conjunto de datos asociado, el cual deberá ser resuelto utilizando los algoritmos que se indiquen.	30%
Proyecto	Cada grupo desarrollará un caso de estudio, aplicando el proceso de aprendizaje a partir de datos.	35%

B) Generalidades.

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

donde encuentra información del curso (<https://sistemas.uniandes.edu.co/es/mine-academico/plan-estudios/enfasis-profundizacion>).

C) 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 3.00/5.00.
- No existe aproximación automática en la nota definitiva. En particular, no hay aproximación a 3.00 de puntajes menores a esta nota (e.g., 2.99 no es 3.00).

D) Protocolo MAAD.

El miembro de la comunidad que sea sujeto, presencia 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
2. **Decanatura de Estudiantes:** centrodeapoyo@uniandes.edu.co
3. **Red de Estudiantes:**
 - PACA (Pares de Acompañamiento contra el Acoso) paca@uniandes.edu.co
4. **Consejo Estudiantil Uniandino(CEU)** comiteacosoceu@uniandes.edu.co

E) Política de momentos difíciles.

Estamos viviendo situaciones complejas de salud pública y entorno social, en el mundo y particularmente en nuestro país. Por ello, la Universidad brinda a toda la comunidad apoyo en muchas dimensiones. Entre ellas, debemos recordar que contamos con apoyo emocional, de salud, tecnológico para la virtualidad. Hay líneas directas de atención a través de los canales de

<https://virtualidad.uniandes.edu.co/>, virtualidad@uniandes.edu.co,
<https://uniandes.edu.co/es/soy/servicios/salud> y
<https://decanaturadeestudiantes.uniandes.edu.co/images/stories/Imagenes/2020-10/apoyo/960/lineas-de-apoyo.jpg>

Todas las personas pueden pasar por un momento difícil que de alguna manera pueda afectar nuestra vida en la Universidad. Pueden ser problemas en casa, con la pareja, incluso estrés por esta u otra materia. Si usted siente que está pasando por un momento complicado, sin importar el motivo, siéntase con la tranquilidad de hablar con las profesoras para pedir tiempo o apoyo. Ningún trabajo o entrega puede sobrepasar su salud mental y física. Su bienestar es lo más importante.