

Programa del curso

Semestre 2021-10

Nombre del curso:	Análisis de información con <i>Machine Learning</i>
Course Name:	Information Analysis with Machine Learning
Créditos:	4
Profesor:	Haydemar Núñez
Versión PDF	Click Aquí

Descripción

Se ilustran técnicas de análisis de información mediante la construcción de modelos apoyados en **aprendizaje automático**, tomando en consideración las características de representación, generación y uso de los datos en la toma de decisiones. [tab title="Objetivos pedagógicos"]

Objetivos pedagógicos

El **Aprendizaje Automático** o **Machine Learning** (también conocido como aprendizaje de máquinas o aprendizaje a partir de datos), 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 aprendizaje automático es que no se dispone de un conocimiento explícito que permita derivar una solución analítica a 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, en muchos casos, 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, que permitan construir soluciones de apoyo a la toma de decisiones en diferentes contextos. 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 del negocio, así como las características y complejidad de las representaciones de los conjuntos de datos disponibles a partir de diferentes fuentes. El objetivo general de este curso es promover las competencias necesarias para la aplicación efectiva de las técnicas de *machine*

learning, en la construcción de modelos de apoyo a la toma de decisiones en diferentes contextos. De esta forma, se espera desarrollar y reforzar las siguientes competencias:

- **Comprender** los fundamentos asociados al *machine learning* y cómo puede plantearse una solución basada en este paradigma de inteligencia artificial.
- **Analizar** el proceso de aprendizaje a partir de datos, desde la formulación de los objetivos del negocio hasta la explotación y monitorización del conocimiento inducido.
- **Identificar** los tipos de problemas que pueden ser resueltos con las técnicas de *machine learning*.
- **Comprender** los aspectos relacionados con la recolección y preparación de datos de diferentes fuentes, tomando en cuenta las características y complejidad de su representación.
- **Conocer** las técnicas de *machine learning*, supervisadas y no supervisadas, y cómo utilizarlas para la construcción de soluciones alineadas con los objetivos del negocio y las características de los datos disponibles.
- **Identificar** las técnicas y medidas de evaluación que pueden ser utilizadas para determinar la validez de los modelos construidos con métodos de aprendizaje.
- **Utilizar** herramientas de software libre para la resolución de problemas de aprendizaje sobre conjuntos de datos disponibles en repositorios públicos.

Plan de temas

El curso se desarrolla en 6 módulos, así:

1. *Machine learning*: conceptos básicos y fundamentos
2. Diseño de sistemas de aprendizaje
3. Recolección y preparación de datos
4. Aprendizaje supervisado: clasificación, análisis de tareas predictivas. SVM. evaluación.
5. Aprendizaje no supervisado: agrupación. evaluación, interpretación de modelos
6. Tópicos y aplicaciones. Detección de anomalías

Metodología

- Clases teóricas presenciales, 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 herramientas de software libre (RapidMiner) y el lenguaje de programación Python con librerías de machine learning (scikit-learn) en el ambiente Jupyter Notebook.

Se realizan actividades prácticas individuales y trabajo en grupos. La evaluación es individual.

Conocimientos previos

Saber programar. Es deseable tener conocimiento de Python. Se espera un nivel intermedio de programación, como mínimo. Matemáticas básicas: teoría de conjuntos, álgebra de vectores Conceptos básicos de probabilidad y estadística