



Universidad Central de Venezuela
Facultad de Ciencias
Escuela de Computación

Programa de la asignatura
APRENDIZAJE AUTOMÁTICO

Tipo de asignatura:	Electiva/Optativa (Teórico y Práctica)
Código:	XXXX
Unidades de crédito:	5
Requisitos:	Probabilidad y Estadística (6104)
Horas semana:	4 horas semanales teóricas y 2 de práctica o laboratorio
Vigencia:	Pensum 2004-rev2005, materia incorporada en julio de 2023

OBJETIVO DE LA ASIGNATURA

El aprendizaje automático (*Machine Learning*) ha adquirido una influencia significativa en diversas áreas de nuestra vida hoy en día siendo uno de los perfiles profesionales con mayor demanda hoy día en el campo de ciencias de la computación. Desde aplicaciones de reconocimiento facial en nuestros dispositivos móviles hasta recomendaciones de productos personalizadas en plataformas de comercio electrónico, culminando en aplicaciones de procesamiento de lenguaje natural, el aprendizaje automático ha transformado la forma en que interactuamos con la tecnología. El curso se centra en comprender los principios y algoritmos clave utilizados en el aprendizaje automático, así como en su aplicación práctica a través de proyectos y ejercicios.

Objetivos Competenciales Cognitivos (OCC)

1. Comprender los conceptos básicos del aprendizaje automático, incluidos los tipos de problemas que se pueden abordar y las diferentes técnicas utilizadas.
2. Familiarizarse con los algoritmos de aprendizaje automático más comunes y asociarlos con el tipo de aprendizaje correcto.
3. Aprender a aplicar técnicas de preprocesamiento de datos para mejorar la calidad de los conjuntos de datos utilizados en el aprendizaje automático.
4. Adquirir habilidades para evaluar y medir el rendimiento de los modelos de aprendizaje automático utilizando métricas y técnicas de validación cruzada.
5. Desarrollar la capacidad de implementar algoritmos de aprendizaje automático utilizando bibliotecas y herramientas populares.
6. Aplicar los conocimientos adquiridos a través de proyectos prácticos, resolviendo problemas del mundo.

CONTENIDO PROGRAMÁTICO TEÓRICO

Tema 1: Introducción al aprendizaje automático

Definición y conceptos básicos de aprendizaje automático. Breve historia de aprendizaje automático y línea de tiempo del área. Breve repaso sobre conceptos de Álgebra Lineal y Cálculo Multivariable. Tipos de aprendizaje automático: supervisado, no supervisado y por refuerzo. Ejemplos de aplicaciones del aprendizaje automático en diferentes campos. Aspectos éticos en el aprendizaje automático. Sesgos y discriminación en conjuntos de datos. Privacidad y protección de datos. Reproducibilidad de flujos de trabajo.

Considera los Objetivos Competenciales Cognitivos (OCC): 1, 2.

Tema 2: Preprocesamiento de datos

Limpieza de datos: manejo de valores faltantes, datos inconsistentes y ruido. Normalización y estandarización de datos. Selección y extracción de atributos. Aumento de datos (*data augmentation*).

Considera los Objetivos Competenciales Cognitivos (OCC): 3, 6.

Tema 3: Algoritmos de aprendizaje supervisado

Introducción y conceptos básicos sobre regresión y clasificación. Regresión lineal y regresión logística. Máquinas de vectores de soporte (*Support-Vector Machines, SVM*). Árboles de decisión y bosques aleatorios. Redes neuronales artificiales.

Considera los Objetivos Competenciales Cognitivos (OCC): 1, 4, 6.

Tema 4: Algoritmos de aprendizaje no supervisado

Conceptos básicos de agrupamiento (*clustering*). Algoritmos y modelos: k-medias, agrupamiento jerárquico, reglas de asociación, entre otros.

Considera los Objetivos Competenciales Cognitivos (OCC): 1, 4, 5, 6.

Tema 5: Algoritmos de reducción de dimensionalidad

Conjuntos de datos de alta dimensionalidad. Medición de la dimensionalidad. Algoritmos y modelos: Análisis de Componentes Principales (PCA), *Local Linear Embedding* (LLE), *t-Distributed Stochastic Neighbor Embedding* (t-SNE), *Uniform Manifold Approximation and Projection* (UMAP), entre otros.

Considera los Objetivos Competenciales Cognitivos (OCC): 3, 4, 5, 6.

Tema 6: Evaluación de modelos de aprendizaje automático

Métricas de evaluación: precisión, *recall*, F1-score, matriz de confusión, curva ROC (*Receiver Operating Characteristic*), entre otras. Validación cruzada y selección de modelos. Sobreajuste (*overfitting*) y subajuste (*underfitting*). Técnicas de visualización de estadística descriptiva útiles en aprendizaje automático: gráficos de dispersión, histogramas de frecuencia, líneas de tiempo, entre otros.

Considera los Objetivos Competenciales Cognitivos (OCC): 2, 4, 6.

Tema 7: Aplicaciones y casos de estudio

Reconocimiento de imágenes y visión por computadora. Sistemas de recomendación. Aprendizaje automático en medicina y finanzas. Análisis deportiva (sabermetría y extensión del modelo xG en fútbol), procesamiento de lenguaje natural (síntesis automática de texto, reconocimiento automático de entidades), entre muchos otros.

Considera los Objetivos Competenciales Cognitivos (OCC): 5, 6.

ESTRATEGIAS DE ENSEÑANZA Y DE APRENDIZAJE

El curso podría combinar clases teóricas con sesiones prácticas en laboratorio.

Durante las clases teóricas, se hace énfasis en cómo leer artículos de investigación asociados a cada tema y comprender las limitaciones de qué necesitamos para el curso. Se presentarán los conceptos y algoritmos fundamentales, y se discutirán las aplicaciones de aprendizaje automático en diversos campos.

Las sesiones prácticas permitirán a los estudiantes implementar y experimentar con los algoritmos utilizando librerías y herramientas populares en el campo.

Además de las clases y el material proporcionado por el profesor, se recomendarán recursos adicionales, como libros, tutoriales en línea y artículos científicos, para aquellos estudiantes interesados en profundizar en aspectos específicos.

ESTRATEGIAS DE EVALUACIÓN

La evaluación se llevará a cabo a través de una combinación de exámenes, proyectos individuales o en grupo, y ejercicios prácticos. Los proyectos, siendo el foco principal del curso, permitirán a los estudiantes aplicar los conceptos y técnicas aprendidas para resolver problemas reales utilizando conjuntos de datos reales.

Distribución propuesta: (por definir)

Tareas y asignaciones	Exámenes	Proyectos	Total
20%	40%	40%	100%

BIBLIOGRAFÍA RECOMENDADA

Aggarwal, C. C. (2014). *Data Classification: Algorithms and Applications*. Boca Raton, FL: CRC Press.

Aggarwal, C. C. (2015). *Data mining*. (2015th ed.) [PDF]. doi:10.1007/978-3-319-14142-8.

Aggarwal, C. C. (2018). *Neural networks and deep learning* (1st ed.). Cham, Switzerland: Springer International Publishing.

Aggarwal, C. C. (2020). *Linear algebra and optimization for machine learning* (1st ed.). doi:10.1007/978-3-030-40344-7.

Aggarwal, C. C., & Reddy, C. K. (2018). *Data Clustering: Algorithms and Applications*. Boca Raton, FL: CRC Press.

Geron, A. (2022). *Hands-on machine learning with scikit-learn, keras, and TensorFlow*. (3rd ed.). Sebastopol, CA: O'Reilly Media.

Goodfellow, I., Bengio, Y, Courville, A. (2016). *Deep Learning*. MIT Press. (<http://www.deeplearningbook.org>)

James, G., Witten D., Hastie T., Tibshirani R., Taylor J (2023). *An Introduction to Statistical Learning: with Applications in Python* (1st de.). Springer. (<https://www.statlearning.com/>)

Müller A., Guido S. (2017). *Introduction to Machine Learning with Python: a guide to data scientists* (1st ed., second release). O'Reilly.

Strang, G. (2019). *Linear algebra and learning from data*. Wellesley, MA: Wellesley-Cambridge Press.