



UCU

Universidad
Católica del
Uruguay

“Introducción a los Métodos de Aprendizaje Automático – Machine Learning”

1 - 5 de agosto de 2022

Dr. Ernesto Ocampo

Dr. Gustavo Vázquez

CONTENIDO

Desarrollo del curso.....	3
Objetivos.....	3
Resultados Esperados del Aprendizaje.....	3
Estructura general del curso y metodología.....	4
Instrucciones Generales de la Materia.....	6
Horarios y asistencia.....	6
Webasignatura.....	6
Bibliografía básica.....	7
Bibliografía adicional.....	7

DESARROLLO DEL CURSO.

OBJETIVOS.

El objetivo del "Aprendizaje Automático" (una de las traducciones más utilizadas de Machine Learning) es analizar y descubrir patrones que ocurren en conjuntos de datos. El verdadero potencial se manifiesta cuando se analizan grandes volúmenes de información, o cuando las relaciones entre los datos son tan complejas o desconocidas que humanamente no es posible descubrirlas. El Aprendizaje Automático se enmarca dentro de una disciplina mayor que se denomina Data Science (o Ciencia de Datos), y es un componente crítico de la Inteligencia Artificial. Ciencia de Datos es la intersección del conocimiento de diversas áreas, tales como el aprendizaje automático, matemática (fundamentalmente análisis estadístico), desarrollo de software, visualización y comunicación de resultados.

La intención del curso es familiarizarte con el ciclo completo de resolución de un problema basado en Aprendizaje Automático ("Machine Learning" - ML), presentando diversas técnicas y metodologías actualmente muy utilizadas en la industria. Es también un objetivo el procurar que mejores tu capacidad de analizar y contrastar diversas soluciones basadas en sistemas inteligentes para resolución de variados problemas computacionales en tu vida profesional.

RESULTADOS ESPERADOS DEL APRENDIZAJE.

Al aprobar el curso el curso serás capaz de:

- **Aplicar** un proceso completo de definición y resolución de problemas basados en aprendizaje automático
- Dados diferentes problemas computacionales, **determinar** si es apropiado o no aplicar técnicas de Aprendizaje Automático para resolverlos eficazmente.
- **Formalizar y representar** un problema en el marco de diferentes algoritmos de aprendizaje automático y conjuntos de datos a utilizar.
- **Identificar los problemas existentes en conjuntos de datos** y **Aplicar** técnicas apropiadas, globalmente reconocidas, para la preparación previa de los mismos.
- **Evaluar los costos** de recursos, implementación y tiempo que insumen diversos enfoques para la resolución de un problema con técnicas de Aprendizaje Automático.
- **Investigar y explorar** diversos algoritmos de Aprendizaje de Máquina para la resolución de eventuales problemas en contextos heterogéneos.

ESTRUCTURA GENERAL DEL CURSO Y METODOLOGÍA

En este curso tendremos un enfoque aplicado, de “afuera hacia adentro”: primero veremos las características de problemas, conjuntos de datos, algoritmos y modelos, procurando usarlos y discernir entre alternativas disponibles. Eventualmente luego se podrá ir profundizando y analizando en detalle los algoritmos, a efectos de mejorar el rendimiento, precisión y eficiencia.

El curso se ha organizado en “Unidades Temáticas” que concentran tópicos relacionados. En la webasignatura podrás observar que cada Unidad Temática tiene su sección, especificando objetivos, resultados esperados del aprendizaje, recursos de lectura / estudio y dinámicas de trabajo aplicado en clase.

UT01 – INTRODUCCIÓN a ML

- El proceso de Machine Learning
- Tipos de algoritmos y sus aplicaciones
- Revisión de Conceptos básicos de estadística descriptiva
- Terminología de Machine Learning
- Algoritmos Paramétricos y no Paramétricos
- Aprendizaje Supervisado, No Supervisado y Semi-Supervisado
- Sesgo, varianza y el compromiso entre ambos
- Sobre y sub-ajuste

UT02 – Tratamiento previo de los datos y fundamentos de los algoritmos de ML

- Visualización de las interacciones de atributos
- Normalización y estandarización de los datos
- Tratamiento de valores faltantes en los conjuntos de datos
- “Feature engineering”
- Estimación del rendimiento de algoritmos de ML
 - Métricas
 - Cross-validation, Estratificación, k-fold, otras técnicas
 - Estimación de una línea base de rendimiento para los modelos

UT03 – Algoritmos Lineales

- Regresión Lineal
- Regresión Logística
- Resumen de ventajas, desventajas y requerimientos de estos métodos lineales

UT04 – Algoritmos No Lineales

- Árboles de clasificación y de regresión
- Naive Bayes
- K-Nearest Neighbors
- Support Vector Machines
- Resumen de ventajas, desventajas y requerimientos de estos algoritmos

UT05 – Clustering y Modelos Jerárquicos

- Concepto de distancia – medidas de similitud
- K-means
- Métodos jerárquicos – Dendogramas, método aglomerativo
- Métodos basados en densidad - DBSCAN

UT06 – Ensamblados

- Bagging y Random Forest
- Boosting y AdaBoost

UT07 – Comparación del rendimiento de algoritmos de ML y sintonía de parámetros

- Sintonía o Ajuste de los parámetros de algoritmos de Machine Learning.
- Comparación del rendimiento de algoritmos de Machine Learning
 - Selección del mejor algoritmo para un problema dado
 - Diseño y ejecución de experimentos
 - Revisión de los resultados de experimentos y conclusiones

INSTRUCCIONES GENERALES DE LA MATERIA.

HORARIOS Y ASISTENCIA.

El curso se dictará desde el lunes 1 al viernes 5 de agosto, de 14:30 a 18:30 horas, en dos módulos de 110 minutos cada día.

Para aprobar este curso necesitas asistir al menos al **75%** del total de módulos dictados y realizar todos los ejercicios indicados en cada unidad temática.

WEBASIGNATURA

La Web asignatura de nuestro curso contiene todos los materiales puestos a tu disposición por los docentes.

El foro de anuncios generales de la web asignatura es el medio de comunicación básico que tenemos para relacionarnos contigo continuamente.

Toda la bibliografía y recursos necesarios (lecturas, videos, ejemplos, ejercicios, recursos web, etc.) para la preparación de cada Unidad Temática será indicada en la sección correspondiente de cada tema en la Webasignatura.

BIBLIOGRAFÍA BÁSICA

En términos generales, se utilizarán los siguientes materiales:

[BROWNLEE2017a] Jason Brownlee: “Master Machine Learning Algorithms - Discover How They Work and Implement Them From Scratch” Edition, v1.10 <http://MachineLearningMastery.com>

[NORTH2012] Matthew North, “Data Mining for the masses”, disponible en <https://docs.rapidminer.com/downloads/DataMiningForTheMasses.pdf>

[BROWNLEE2017b] Jason Brownlee: “Machine Learning Mastery with Weka - Analyze Data, Develop Models and Work Through Projects” Edition, v1.2 <http://MachineLearningMastery.com>

[JAMES2014] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani: “An Introduction to Statistical Learning with Applications in R” Springer (2014), (descargable por Timbó)

[WITTEN2011] Ian H. Witten, Eibe Frank, Mark A. Hall: “Data Mining: Practical Machine Learning Tools and Techniques” - Morgan Kaufmann (2011)

[KUHN2013] Max Kuhn, Kjell Johnson: “Applied predictive Modelling” – Springer (2013) (descargable por Timbó)

[CHIS2013] Andrew Chisholm, “Exploring Data with RapidMiner” - [PACT] Publishing

Como bibliografía complementaria a la anterior se puede usar:

[RUSSELL2004] S. Russell & P. Norvig: "Inteligencia Artificial: Un enfoque moderno". Pearson Prentice Hall, 2da edición ISBN 84-205-4003-X. (2da edición en inglés "Artificial Intelligence: A modern approach". ISBN 978-0137903955).

[MITCHELL1997] Mitchell, Tom M.: "Machine Learning". WCB/McGraw-Hill, ISBN 0-07-042807-7.

[NILSSON2005] Nils J. Nilsson: "Introduction to Machine Learning".

BIBLIOGRAFÍA ADICIONAL

En la webasignatura del curso se publicarán una variedad de materiales y referencias adicionales, actualizadas, para cada unidad temática.