

# POLITÉ

UNIVERSIDAD POLITÉCNICA DE MADRID

# CNICA

## GUIA DEL CURSO IA: Aprendizaje Automático

SERVICIO DE FORMACIÓN CONTINUA  
VICERRECTORADO DE ESTRATEGIA ACADÉMICA E  
INTERNACIONALIZACIÓN





## ÍNDICE

1. INTRODUCCIÓN .....	3
2. PROFESORADO .....	5
3. RESUMEN DE CONTENIDOS POR MÓDULOS .....	5
4. COMPETENCIAS ESPECÍFICAS A ALTO NIVEL .....	11
5. METODOLOGÍA DE IMPARTICIÓN .....	13
6. EVALUACIÓN .....	14
7. CRONOGRAMA .....	15

<b>Aprendizaje Automático</b>	
Duración:	400 h
Semanas de aprendizaje	20

## 1. INTRODUCCIÓN

---

El aprendizaje automático es un campo emergente, a caballo entre las matemáticas, la estadística y la informática, que requiere un enfoque multidisciplinar y debe construirse sobre la base de un conjunto de nuevas tecnologías que están generando un efecto transformador en todos los dominios de la industria y de la investigación.

Por su parte, la inteligencia Artificial (IA) es un área tradicional de la informática que comparte algunas técnicas con otras disciplinas, como las matemáticas y la estadística o la ciencia cognitiva. Tal y como señala la Estrategia Española de I+D+I en Inteligencia Artificial<sup>1</sup>, en este momento la IA es una de las disciplinas que puede influir más en la rápida transición hacia una nueva sociedad y economía. La relación entre ciencia de datos e IA es incuestionable. Por una parte, la IA provee de técnicas y tecnologías específicas para gestionar, reutilizar y explotar los datos de una institución (pública o privada). Por otra parte, la ciencia de datos desempeña un papel crucial para alcanzar todo el potencial de la IA. Las principales asociaciones internacionales del ámbito, como la Big Data Value Association (BDVA) a través de su informe de abril de 2019<sup>2</sup> y su documento de visión para la creación de un partenariado público privado en IA<sup>3</sup>, o el propio Ministerio de Ciencia, Innovación y Universidades en su documento de Estrategia Española de I+D+I en IA, reconocen ya el valor esencial de los datos para permitir alcanzar todo el potencial de la IA y plantean la implementación y adopción de un espacio europeo de compartición de datos que permita la interoperación entre espacios de datos verticales, intersectoriales, personales e industriales, nuevos o existentes. En particular, recomiendan dotar a los nuevos perfiles profesionales relacionados con la ciencia de datos y la IA de las competencias y habilidades requeridas para permitir las necesarias sinergias.

El curso: **“Aprendizaje Automático”** tiene una carga docente de 400 horas lectivas con la siguiente estructura modular:

- Python para Data Science
- Visualización
- Preparación de datos
- Fundamentos matemáticos para Data Science
- Optimización exacta y aproximada
- Aprendizaje Supervisado
- Aprendizaje No Supervisado
- Deep Learning

El curso pretende dotar a los/as alumnos/as participantes con los conocimientos avanzados en aprendizaje automático, altamente demandados por el mercado laboral, que permita mejorar las posibilidades de contratación de los/as alumnos/as participantes o, su mejora laboral. El curso proporciona los conocimientos y competencias necesarias para desarrollar soluciones en el área del aprendizaje automático con sólidos conocimientos y habilidades en matemáticas, estadística y computación (y en IA en particular), y con habilidades propias de la ingeniería, que les permitan afrontar problemas complejos de algoritmia, modelización, optimización, análisis y visualización de grandes volúmenes de datos.

La acción formativa se programará entre los meses de febrero y junio de 2020, de lunes a jueves/viernes en horario de tarde de 16:00 a 21:00 y se impartirá en el aula Samsung-UPM ubicada en el Campus Sur de la UPM, Escuela Técnica Superior de Ingeniería de Sistemas Informáticos.

El curso está orientado para candidatos/as que hayan finalizado o estén en el último año de carrera de estudios prioritariamente en Informática, Telecomunicación y/o carreras de Ciencias e Ingeniería.

Los/as candidatos/as seleccionados/as, no deberán realizar ninguna aportación económica, pero si firmar un compromiso de permanencia/asistencia al curso, que implicará una penalización económica en caso de incumplimiento tal y como se establece en las bases de la convocatoria de selección del curso.

El/la alumno/a deberá asistir como mínimo al 80% de las horas programadas para poder ser evaluado, considerándose únicamente como faltas justificadas a los efectos del cumplimiento del porcentaje anterior, las de enfermedad con la aportación del documento correspondiente que acredite el hecho.

## 2. PROFESORADO

---

El grupo de profesores que participan en el curso a lo largo de su carrera docente e investigadora han desarrollado un perfil adecuado en el ámbito de la ciencia de datos y la inteligencia artificial, incluyendo tanto formación reglada como continua en las áreas de conocimiento relacionadas con el curso que se propone: matemáticas, estadística, ciencia de la computación e inteligencia artificial, lenguajes y sistemas informáticos, arquitectura y tecnología de computadores. El grupo de docentes del curso publica en media más de 15 artículos científicos en revistas indexadas en posiciones relevantes del JCR.

Entre los profesores participantes se encuentran: Jesús Bobadilla Sancho, Abraham Gutiérrez Rodríguez, Fernando Ortega Requena, Raúl Lara Cabrera, Ángel González Prieto y Edgar Talavera Muñoz.

## 3. RESUMEN DE CONTENIDOS POR MÓDULOS

---

La planificación del módulo formativo podrá verse modificada por motivos imprevistos (rendimiento del grupo, disponibilidad de recursos, etc.) y por tanto no deberá considerarse como definitiva y cerrada.

- **Python para Data Science**
  - Objetivos generales
    - Este curso tiene como objetivo enseñar a todos los conceptos básicos de la programación de computadoras con Python. Cubrimos los conceptos básicos de cómo se construye un programa a partir de una serie de instrucciones simples en Python.
  - Resultados de aprendizaje
    - Al finalizar el curso el participante estará en la capacidad de:
    - Crea tus propias aplicaciones para la recuperación y procesamiento de datos.
    - Explicar los conceptos básicos de la programación de computadoras usando Python.
    - Comprender conceptos fundamentales de programación tales como estructuras de datos.
    - Utilizar herramientas de programación básicas como funciones y bucles.
    - Introducir las librería básicas para Data Science utilizadas desde Python.
  - Contenidos (Bloques temáticos)
    - Instalación y uso de Python

- Variables y expresiones
  - Cadenas, listas y ficheros
  - Lógica, control de flujo y filtros
  - Expresiones regulares
  - Funciones y paquetes
  - Librerías básicas para Data Science (NumPy, Pandas, Scikit-Learn,...)
- **Visualización**
    - Objetivos generales
      - En este curso el alumno aprenderá los conceptos básicos de la visualización de la información, desde una perspectiva basada en la abstracción de datos, tareas y codificaciones y su relación con visualizaciones concretas para distintos tipos de datos. Como lenguaje de soporte para la parte práctica del curso se usará Python.
    - Resultados de aprendizaje
      - Formular el problema de visualización en términos de los distintos niveles de abstracción (datos, tareas y codificaciones).
      - Elegir de entre las distintas técnicas de visualización la más adecuada a un problema concreto.
      - Trabajar con múltiples vistas coordinadas sobre los datos y a añadir interacción a las mismas.
    - Contenidos (Bloques temáticos)
      - Introducción a la visualización de la información
      - Tipos de datos
      - Organización en estructuras de datos más complejas
      - Codificación visual
      - Visualizaciones para datos tabulares
      - Visualizaciones para grafos
      - Visualizaciones para datos espaciales
      - Manipulación e interacción
  - **Preparación de datos**
    - Objetivos generales
      - El objetivo general de este curso es que los estudiantes se familiaricen con las técnicas básicas de preprocesado de datos. El preproceso es fundamental antes de aplicar las técnicas de modelado de datos puesto que la calidad de los resultados depende en gran medida de esta fase. De

hecho se estima que el 80% del tiempo de un proyecto de data science se dedica al preprocesado de los datos

- Resultados de aprendizaje
  - Ser capaces de analizar los datos que serán sometidos a tareas de aprendizaje
  - Ser capaz de realizar el preproceso necesario dependiendo de los datos que se tienen y de los que se necesitan para poder extraer el conocimiento
- Contenidos (Bloques temáticos)
  - Introducción
  - Análisis exploratorio de datos
  - Detección y tratamiento de valores atípicos
  - Detección y tratamiento de valores nulos: Imputación
  - Normalización / estandarización
  - Discretización
  - Reducción de dimensionalidad
  - Desbalanceo de clases
  - Derivación de nuevas variables
- **Fundamentos matemáticos para Data Science**
  - Objetivos generales
    - En este curso el alumno aprenderá los fundamentos matemáticos y estadísticos necesarios para comprender el funcionamiento de las diferentes técnicas de Data Science. El curso se centra en aquellos conceptos con los que el alumno debe estar familiarizado para comprender cómo funcionan dichas técnicas y para interpretar la salida de las mismas. Como soporte para ilustrar los conceptos teóricos cubiertos en el curso se usará como herramienta el lenguaje de programación Python. Por último, el enfoque del curso será eminentemente de aplicación, por lo que se relacionarán los conceptos teóricos con su aplicación en Data Science.
  - Resultados de aprendizaje
    - Comprender los fundamentos teóricos de muchas técnicas de Data Science (algoritmos de clasificación lineal, algoritmos de clustering, algoritmos probabilísticos, etc.).
    - Interpretar la salida devuelta por muchos de estos algoritmos (estimadores estadísticos, intervalos de confianza, contraste de hipótesis, etc.).



- Utilizar las distintas técnicas de estadística descriptiva y representaciones gráficas básicas para realizar un análisis exploratorio interactivo previo al proceso de preparación de datos para su posterior análisis.
- Contenidos (Bloques temáticos)
  - Álgebra lineal
  - Probabilidad
  - Estadística descriptiva
  - Estadística inferencial
- **Optimización exacta y aproximada**
  - Objetivos generales
    - En este curso el alumno aprenderá acerca de las distintas técnicas de optimización y algoritmos bio-inspirados y cómo pueden ser usados para resolver problemas de caja negra. Se relacionarán dichas técnicas de optimización con otras técnicas similares presentes en muchos algoritmos de Data Science (como, por ejemplo, todas las basadas en el descenso del gradiente). Asimismo, se presentarán distintos casos prácticos de complejidad variable para ser resueltos durante la clase. Por último, se presentarán algunos conceptos avanzados relacionados con la paralelización de los algoritmos bio-inspirados, la evaluación y comparación de los resultados de los mismos, la complejidad de un problema de optimización, etc.
  - Resultados de aprendizaje
    - Formular un problema en términos de problema de optimización.
    - Usar las técnicas evolutivas presentadas para resolver problemas de diversa índole.
    - Analizar la naturaleza del problema y los resultados obtenidos por el algoritmo seleccionado.
  - Contenidos (Bloques temáticos)
    - Introducción a las técnicas de optimización. Optimización heurística vs. Metaheurística
    - Definición de un problema de optimización
    - Algoritmos bio-inspirados
    - Algoritmos híbridos (meméticos, portfolio optimization, etc.)
    - Paralelización de algoritmos bio-inspirados
    - Optimización multi-objetivo
    - Ajuste de parámetros en algoritmos bio-inspirados



- Métricas de dificultad de problemas
- Aplicaciones de las técnicas de optimización
  
- **Aprendizaje Supervisado**
  - Objetivos generales
    - El objetivo general de este curso es que los estudiantes se familiaricen con las técnicas básicas de aprendizaje supervisado. En particular se analizan los tipos de problemas distinguiendo entre la clasificación y la predicción de valores y se analizan las técnicas básicas para solventar estos tipos de problemas. Se analizará también como evaluar los resultados obtenidos y como comparar los resultados obtenidos usando distintas técnicas.
  - Resultados de aprendizaje
    - Ser capaces de distinguir un problema de aprendizaje supervisado
    - Ser capaz de elegir la mejor técnica del problema de aprendizaje dependiendo del problema y de la naturaleza de los datos y aplicarla en un entorno de programación real
    - Ser capaz de evaluar y comparar los resultados obtenidos
  - Contenidos (Bloques temáticos)
    - Introducción
    - Clasificación vs Predicción de Valores
    - Técnicas de Clasificación
    - Combinación de clasificadores
    - Predicción de Valores
    - Máquinas de vector soporte
    - Evaluación de resultados de clasificación
    - Comparación de clasificadores
  
- **Aprendizaje No Supervisado (Clustering)**
  - Objetivos generales
    - El objetivo general de este curso es que los estudiantes se familiaricen con las técnicas básicas de aprendizaje No supervisado. En particular se analizan los tipos de problemas distinguiendo entre los distintos tipos de clustering y el problema de calcular la similitud y se analizan las técnicas básicas para obtener clusterings. Se analizará también como evaluar los resultados obtenidos.
  - Resultados de aprendizaje
    - Ser capaces de distinguir un problema de aprendizaje No supervisado

- Ser capaz de elegir la mejor técnica del problema de aprendizaje dependiendo del problema y de la naturaleza de los datos y aplicarla en un entorno de programación real
- Ser capaz de evaluar y comparar los resultados obtenidos
- Contenidos (Bloques temáticos)
  - Introducción a los problemas de aprendizaje no supervisado: clustering
  - Tipos de clusters
  - Midiendo la similitud entre clusters
  - Validación de clusters
  - Técnicas de clustering particional
  - Técnicas de clustering basadas en densidad
  - Técnicas de clustering basadas en gráficos
- **Aprendizaje No Supervisado (Reducción de dimensionalidad)**
  - Objetivos generales
    - El objetivo general de este curso es que los estudiantes se familiaricen con las técnicas de reducción de dimensionalidad. Este es un problema que nos enfrentamos cuando vamos a realizar modelos de clasificación.
  - Resultados de aprendizaje
    - Ser capaces de analizar cuando se requiere reducción de dimensionalidad
    - Ser capaces de aplicar las técnicas
    - Ser capaces de entender los resultados obtenidos y evaluarlos
  - Contenidos (Bloques temáticos)
    - Necesidad de la reducción de dimensionalidad.
    - Selección de características vs. reducción de dimensionalidad.
    - Técnicas para la selección de características
    - Técnicas para la reducción de la dimensionalidad
    - Detección de anomalías.
- **Deep Learning**
  - Objetivos generales
    - Este curso tiene como objetivo enseñar los conceptos de Deep Learning, entendido como redes neuronales multicapa. Un segundo objetivo es dominar las técnicas más modernas de procesamiento de imágenes, basadas en aprendizaje automático.
  - Resultados de aprendizaje

- Comprender el funcionamiento de las redes neuronales backpropagation.
- Diseñar arquitecturas Deep Learning para distintos tipos de información.
- Realizar aplicaciones de clasificación de patrones.
- Crear soluciones de reconocimiento de imágenes.
- Utilizar con solturas diversas técnicas para mejorar la calidad de los resultados en clasificación de imágenes.
- Visualizar y analizar la información de aprendizaje de una arquitectura Deep Learning.
- Comprimir y descomprimir información mediante diversas arquitecturas de red neuronal.
- Generar información basada en resultados intermedios de aprendizaje profundo.
- Contenidos (Bloques temáticos)
  - Redes neuronales
  - Clasificación usando redes neuronales
  - Redes Convolucionales. Conceptos básicos
  - Clasificación usando Redes Convolucionales en datasets sencillos
  - Uso de data generators y redes convolucionales
  - Enriquecimiento de los datos (data augmentation)
  - Visualización de las capas ocultas
  - Aprendizaje por transferencia (transfer learning)
  - Autoencoders
  - Aprendizaje generativo.

#### 4. COMPETENCIAS ESPECÍFICAS A ALTO NIVEL

---

- Utilizar con destreza los conceptos y métodos matemáticos que subyacen los problemas de la ciencia de datos y la inteligencia artificial para su modelización y resolución.
- Diseñar, implementar y evaluar soluciones algorítmicas eficientes para problemas computacionales de ciencia de datos e inteligencia artificial de acuerdo con los requisitos establecidos.
- Analizar fenómenos complejos mediante la probabilidad y estadística, y plantear modelos matemáticos de los mismos en situaciones concretas. Formular, modelizar y resolver problemas de optimización matemática relacionados con la ciencia de datos y la inteligencia artificial.
- Utilizar las técnicas adecuadas de análisis de datos y estadística sobre datos disponibles para descubrir nuevas relaciones y proporcionar conocimiento y una

comprensión intuitiva precisa y profunda sobre problemas de investigación o procesos organizacionales y respaldar la toma de decisiones.

- Desarrollar e implementar estrategias de gestión de datos para su posterior procesamiento.
- Utilizar los principios de ingeniería y las tecnologías informáticas actuales para investigar, diseñar e implementar nuevas aplicaciones de análisis de datos; desarrollar experimentos, procesos, instrumentos, sistemas e infraestructuras que den soporte a la gestión de datos durante todo el ciclo de vida de los datos.

## 5. METODOLOGÍA DE IMPARTICIÓN

Todas las actividades tienen como objetivo principal el de potenciar el aprendizaje de los alumnos, facilitando la adquisición de cuantos conocimientos y competencias precise. Por ello, dependiendo del tipo de contenidos de los diferentes módulos formativos los métodos a utilizar variarán: expositivo, por descubrimiento, interactivo, enseñanza colaborativa, enseñanza programada, trabajo autónomo, metodología participativa (trabajos grupales, aprendizaje cooperativo, ...) etc.

Al comenzar cada módulo del curso, el o la estudiante dispondrá de todo el material de la asignatura. Se incluirán las diapositivas presentadas en clase, como guión (nunca como único contenido a evaluar) de los conocimientos a adquirir, para facilitar su seguimiento. Se proporcionarán las hojas de ejercicios, así como toda la documentación que se considere necesaria relativas a los contenidos formativos.

Los enunciados de las prácticas también estarán disponibles en cada módulo, así como hojas de ejercicios para trabajo personal de los estudiantes.

A continuación se detallan los diferentes métodos docentes que se emplearán a lo largo del curso en este

- Clases magistrales: Se expondrán la teoría básica del programa en cada uno de los temas que se vayan a tocar.
- Prácticas: Se realizará un trabajo práctico propuestos por el profesor y relacionados con la materia estudiada. El trabajo realizado por cada alumno/a será revisado y valorado por el profesor, tanto en contenidos como en presentación, pudiendo ser requeridas de los alumnos cuantas explicaciones se consideren oportunas. Cada alumno tendrá acceso a su informe, debidamente revisado y valorado

Los contenidos se desarrollarán en el aula (clase magistral + prácticas en aula), resolviendo supuestos teórico prácticos). Cada sesión constará de explicaciones teóricas seguidas de actividades prácticas. La explicación teórica se estructurará, cada día, en dos partes de manera que se pueda practicar inmediatamente lo explicado. Durante la parte teórica se introducirán los conceptos clave y se presentarán los elementos tecnológicos más relevantes de cada tema, y durante la parte práctica se aplicarán estos conceptos a través de una serie de ejercicios propuestos. Además de los ejercicios propuestos, se sugiere que cada estudiante se plantee qué sitio web le gustaría crear para poder realizarlo por sí mismo en las sesiones finales.

## 6. EVALUACIÓN

---

### INSTRUMENTOS DE EVALUACIÓN

Para la evaluación del alumno se sigue un tipo criterial, i.e. una evaluación que pretende determinar las competencias que posee el estudiante tras el desarrollo del proceso de aprendizaje. En este tipo de evaluación se recoge información mediante un instrumento, prueba, procedimiento o actividad para poder describir e identificar las competencias adquiridas por los estudiantes acerca de un dominio de referencia, descrito en la planificación docente inicial. Así, los criterios de evaluación deben correlacionarse con los objetivos/competencias y especificar principalmente un dominio conceptual claro, la adquisición de procedimientos, técnicas, instrumentos, habilidades y destrezas de ejecución profesional y académica, que concretan el nivel de capacidades personales y profesionales necesarias para el ejercicio profesional.

En este caso los instrumentos de evaluación serán diferentes pruebas de desarrollo y la elaboración de trabajos en equipo.

### MOMENTOS EN LOS QUE SE REALIZARÁ LA EVALUACIÓN

Los instrumentos de evaluación serán utilizados durante todo el proceso formativo a modo de evaluación continua. Periódicamente, y como mínimo, tras concluir cada tema del bloque temático, se pedirá a los alumnos que resuelvan (individualmente) algún problema o práctica concreta, que deberán entregar para su evaluación.

Un elemento fundamental de este sistema de enseñanza-aprendizaje y de la evaluación continua es el relativo a la retroalimentación permanente que reciben los estudiantes sobre el trabajo realizado. Esto se traduce en la necesidad de devolverles información precisa, objetiva e inmediata acerca de sus resultados de rendimiento, de manera que puedan monitorizar sus niveles de esfuerzo y regular su trabajo, al mismo tiempo que dicha información actúa como un elemento motivador. Con este objetivo, el profesorado deberá establecer un procedimiento de revisión de las pruebas y actividades, indicando en cada caso según sus características la forma en que se llevará a cabo, bien en la propia clase o bien en el horario de tutorías, facilitando a los estudiantes la realización de las mismas.

### RESPONSABLES DE LA EVALUACIÓN

La responsabilidad de la evaluación recaerá principalmente en los profesores que serán los encargados de la revisión de los trabajos prácticos, si bien, los alumnos deberán responsabilizarse de los ejercicios generales que se resuelvan en clases prácticas a modo de autoevaluación.

## 7. CRONOGRAMA

L	M	X	J	V										
3	4	5	6	7	febrero									
10	11	12	13	14										
17	18	19	20	21										
24	25	26	27	28										
2	3	4	5	6	marzo		Python para Data Science (40 horas)							
9	10	11	12	13			Visualización (20 horas)							
16	17	18	19	20			Preparación de datos (40 horas)							
23	24	25	26	27			Fundamentos matemáticos para Data Science (45 horas)							
30	31	1	2	3	abril		Optimización Exacta y Aproximada (45 horas)							
6	7	8	9	10			Aprendizaje Supervisado (Clasificación, Regresión y Personalización) (60 horas)							
13	14	15	16	17			Aprendizaje No Supervisado (Reducción de Dimensiones y Clustering) (60 horas)							
20	21	22	23	24			Deep Learning (90 horas)							
27	28	29	30	1										
4	5	6	7	8	mayo									
11	12	13	14	15										
18	19	20	21	22										
25	26	27	28	29										
1	2	3	4	5	junio									
8	9	10	11	12										
15	16	17	18	19										