



Argentina  
programa  
4.0



Ministerio de  
Desarrollo Productivo  
Argentina

Secretaría de  
Economía del Conocimiento

# PROGRAMADOR JUNIOR EN MACHINE LEARNING I

UNIVERSIDAD NACIONAL DE MISIONES  
FACULTAD DE CIENCIAS EXACTAS QUÍMICAS Y NATURALES

## **CURSO 1: Fundamentos de Machine Learning (ML) Python - Nivel 1 -**

### **FUNDAMENTACIÓN**

Python es un lenguaje de programación ampliamente utilizado en las aplicaciones web, el desarrollo de software, la ciencia de datos y el machine learning (ML). Los desarrolladores utilizan Python porque es eficiente y fácil de aprender, además de que se puede ejecutar en muchas plataformas diferentes.

Python es un lenguaje sencillo de leer y escribir debido a su alta similitud con el lenguaje humano. Además, se trata de un lenguaje multiplataforma de código abierto y, por lo tanto, gratuito, lo que permite desarrollar software sin límites. Python puede utilizarse en proyectos de inteligencia artificial, para crear sitios web escalables, realizar cálculos estructurales complejos con elementos finitos, y diseñar videojuegos, entre otras muchas aplicaciones

### **Objetivos**

- Familiarizar al estudiante con el entorno y las herramientas de programación de Python.
- Desarrollar las capacidades de plasmar una propuesta de resolución a un problema sencillo planteado utilizando las herramientas planteadas.
- Identificar las ventajas del trabajo colaborativo y las herramientas disponibles.
- Promover la indagación sobre los temas planteados.

## Semana 1

- Teoría: En qué consiste el Machine Learning y qué se diferencia de la Inteligencia Artificial y Deep Learning.
- Práctica: Instalación Python + Jupyter.

# Qué es machine learning y data science

Data science o ciencia de datos definición formal

Ciencia para responder preguntas por medio del análisis de datos

Disciplinas relacionadas: ingeniería de software, comunicación de datos y matemática, estadística y algoritmos. Se utiliza para tomar decisiones pero considerando un análisis científico de los datos. Se usan diferentes técnicas entre ellos, la estadística, las consultas dirigidas (SQL/NoSQL) y los métodos de aprendizaje automático (machine learning -ML)

Machine learning estaría incluido (en parte) por el data science, aunque el que se especializa en ML debe desarrollar habilidades de programación y manejo de conceptos inteligencia artificial IA relacionados al aprendizaje automático en mayor profundidad que el experto en ciencia de datos, aunque suelen complementarse no necesariamente son la misma persona en un equipo de trabajo.

El Machine Learning, también conocido como aprendizaje automático, es una rama de la inteligencia artificial que permite a los ordenadores aprender y mejorar su rendimiento en una tarea específica, sin necesidad de ser programados explícitamente para ello.

El aprendizaje automático se basa en algoritmos que utilizan modelos matemáticos para aprender a partir de los datos, identificando patrones y relaciones entre ellos, y luego aplicando ese conocimiento para realizar predicciones o tomar decisiones.

“El ML se enfrenta con el desafío de construir programas computacionales que automáticamente mejoran con la experiencia. Estos son sistemas capaces de adquirir conocimientos de alto nivel y/o estrategias para la resolución de problemas mediante ejemplos, en forma análoga a la mente humana. A partir de ejemplos provistos por un tutor o instructor y de los conocimientos de base o conocimientos previos, el sistema de aprendizaje crea descripciones generales de conceptos.”<sup>1</sup>

Para tener una relación cuando hablamos de algoritmos tenemos los diferentes paradigmas de programación entre ellos el procedimental, el orientado a objetos, el funcional por mencionar algunos. Pero en todos y cada uno de ellos el desarrollador debe tener un profundo dominio del campo de aplicación, o al menos el suficiente como para poder expresar por medio de la

---

<sup>1</sup> Ochoa, M. A. (2004). Herramientas inteligentes para explotación de información. Trabajo Final Especialidad en Ingeniería de Sistemas Expertos url: <https://ri.itba.edu.ar/server/api/core/bitstreams/a848d640-0277-459d-9104-b37017309d31/content>

lógica del código como se resolverán las situaciones. El manejo de datos los que aportaba es pertinencia y aplicación del conocimiento a situaciones específicas, a tener acceso a información histórica, tanto la lógica de los procedimientos y de los informes quedaba expresada de manera estática dentro del código. Más allá de la parametrización que busca dotar de cierta flexibilidad a los sistemas, pero aun así cuando cambian las reglas de negocio el programador deberá modificar el código, habida cuenta de haber entendido previamente cuales son los cambios y los resultados esperados.

Sin embargo del ML se espera poder presentar los datos a los algoritmos de ML y que estos sean capaces de aprender la lógica de la representación del dominio del problema por medio de estos datos. Es decir por medio del modelo implementado aprender la representación de las relaciones entre los datos, de allí la gran importancia que una gran cantidad de datos, al menos representativa del problema bajo análisis, es que debe ser provista de antemano. Luego de creado el modelo que represente el dominio podemos generar nuevas salidas o respuestas a situaciones inesperadas.

## **Diferencias con Inteligencia Artificial y el Deep Learning**

La Inteligencia Artificial se refiere a la capacidad de una máquina para imitar la inteligencia humana, incluyendo la capacidad de razonamiento, aprendizaje, percepción y comprensión de problemas cotidianos, puede incluir o aplicarse a la comprensión del lenguaje natural. Esto puede incluir muchas técnicas de lo más variadas que pueden ser Sistemas Expertos Basados en Reglas, o en Redes Bayesianas, Algoritmos de búsqueda (informada y desinformada), Sistemas difusos, Redes neuronales, entre otros.

El Machine Learning (ML) o Aprendizaje Automático es una rama de la IA que se enfoca en el desarrollo de algoritmos y modelos que pueden aprender a partir de los datos y mejorar su desempeño a medida que se exponen a más datos. El Machine Learning es ampliamente utilizado en una variedad de aplicaciones, desde la clasificación de imágenes hasta la detección de fraudes.

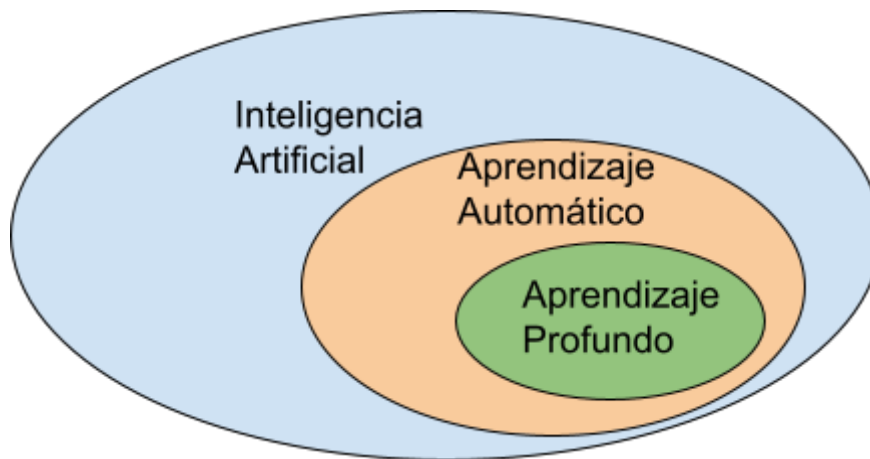
El Deep Learning (DL) o Aprendizaje Profundo es una técnica de Machine Learning que utiliza redes neuronales artificiales para aprender patrones a partir de grandes cantidades de datos. El Deep Learning puede ser utilizado para una amplia variedad de aplicaciones, desde el procesamiento de imágenes hasta el procesamiento del lenguaje natural.

En resumen, el Machine Learning es una técnica de IA que se enfoca en el aprendizaje automático a partir de los datos. El Deep Learning es una técnica de Machine Learning que utiliza redes neuronales profundas para aprender patrones a partir de los datos. Y la IA es un concepto más amplio que se refiere a la capacidad de las máquinas para imitar la inteligencia humana en general.

El proceso de aplicación de la técnica de Machine Learning

Contar con los datos validados y depurados, este proceso de contar con un set de datos válidos es sumamente importante y se considera tanto la calidad como la cantidad de datos presentes. Cuando hablamos de calidad es identificar si cada tupla registrada (filas) cuenta con datos en todos sus campos (columnas), si estos datos registrados en cada campo (columna) son del mismo tipo, si el tipo de datos es el adecuado para el procesamiento y creación del modelo

(cuantificar datos cualitativos, crear escalas de representación, entre otras opciones que podemos detallar más adelante), entre otras opciones de revisión que como outliers (observación de datos anómalos o que se alejan sensiblemente de la media determinando valores extremos o no representativos de la muestra que pueden deberse a ruidos durante la captura de los mismos, estos en una muestra estadística o serie temporal de datos que puede afectar potencialmente a la estimaciones o proyecciones esperadas por parte del modelo). Cuando hablamos de cantidad se refiere a que los datos deben ser suficientes como para representar al dominio del problema en estudio (al menos una cantidad considerable en cada uno de los casos en estudio).



## Tipos de Aprendizaje Automático

Los tipos básicamente son supervisado y no supervisado y definen la existencia o no de un agente externo (supervisor) que controle el aprendizaje de la red<sup>2</sup>

### Aprendizaje supervisado:

En el aprendizaje supervisado, el modelo se entrena utilizando un conjunto de datos de entrenamiento etiquetados. Es decir, cada muestra del conjunto de datos de entrenamiento viene con una etiqueta que indica la respuesta correcta que el modelo debe aprender a predecir. Se podría decir que le damos ejemplos basados en un conjunto de datos de entrada y salidas esperadas relacionadas a las entradas correspondientes.

El objetivo es que el modelo aprenda a generalizar a partir de los datos de entrenamiento para hacer predicciones precisas sobre datos nuevos no vistos previamente. Algunas aplicaciones típicas de aprendizaje supervisado incluyen la clasificación de imágenes, la predicción de precios de bienes raíces o la detección de spam.

Aprendizaje supervisado:  
En el aprendizaje supervisado, el modelo se entrena utilizando un conjunto de datos de entrenamiento etiquetados. Es decir, cada muestra del conjunto de datos de entrenamiento viene con una etiqueta que indica la respuesta correcta que el modelo debe aprender a predecir. El objetivo es que el modelo aprenda a generalizar a partir de los datos de

---

<sup>2</sup> Hilera, J. R. y Martinez, V. J. (2000) Redes Neuronales Artificiales. Fundamentos. modelos y aplicaciones. Alfaomega Ed

entrenamiento para hacer predicciones precisas sobre datos nuevos no vistos previamente. Algunas aplicaciones típicas de aprendizaje supervisado incluyen la clasificación de imágenes, la predicción de precios de bienes raíces o la detección de spam. Se pueden identificar dos grandes grupos.

**Clasificación:** por medio de las características presentes en los datos identifican su pertenencia a un grupo u otro, los grupos deben estar predefinidos en las etiquetas. Por ejemplo, saber si se corresponde un conjunto de datos o características de algún activo financiero como ser una acción a una operación de compra o venta del activo. Saber la posibilidad dadas las características del paciente si se corresponde a la pertenencia a un grupo de enfermedades o a otro (es importante resaltar que un posible subsistema de explicación debería programarse por otros medios, no es un sistema experto basado en reglas, solamente podrá determinar si es una enfermedad u otra).

**Regresión:** saber cual sería la predicción posible de un valor específico. en este sentido se podría predecir el posible valor futuro de una acción.

Tipos de Aprendizaje supervisado (Hilera, Martinez, 2000)

Es importante destacar aunque no será ampliado en este curso que los modelos de redes neuronales artificiales son modelos creados inspirados en los modelos biológicos existentes y por este motivo la lógica de su funcionamiento se vio de alguna manera relacionado al entendimientos que a lo largo de la historia se ha llegado a abarcar sobre el funcionamiento del cerebro humano. Por este motivo el auge de esta tecnología viene de la mano de los avances en los campos de la neurociencia, aunque cabe destacar que los modelos artificiales son modestas representaciones de solamente algunas de las funcionalidades del modelo humano, aunque incluso así se han logrado avances muy importantes.

En el cerebro, el aprendizaje de un hecho parecido a otro del que ya tenemos conocimiento, se hace por refuerzo de las conexiones entre neuronas y esta actividad conjunta es necesaria para memorizarlo. Se ha demostrado que si un sujeto no está expuesto a ciertos estímulos al comienzo de su vida, no podrá detectar más estos estímulos debido a la degeneración de las conexiones necesarias

En este tipo de aprendizaje se suelen considerar a su vez tres formas de llevarlo a cabo que dan lugar a los siguientes de aprendizajes supervisados:

- Aprendizaje por corrección de error: Consiste en ajustar los pesos de las conexiones de la red en función de la diferencia entre los valores deseados y los obtenidos en la salida de la red, es decir, en función del error cometido en la salida
- Aprendizaje por refuerzo: se basa en la idea de no disponer de un ejemplo completo del comportamiento deseado, es decir de no indicar durante el entrenamiento exactamente la salida que se desea proporcione la red ante una determinada entrada. La función del supervisor se reduce a indicar mediante una "señal de refuerzo" si la salida obtenida en la red se ajusta a la deseada ("éxito"=+1 o "fracaso"=-1). El aprendizaje por refuerzo implica el uso de un modelo que aprende a tomar decisiones a través de la interacción con un entorno y la retroalimentación recibida por el modelo. El modelo debe aprender a maximizar una recompensa dada por el entorno mientras se toman decisiones secuenciales. Algunas aplicaciones típicas de aprendizaje por refuerzo incluyen la planificación de rutas y la optimización de la gestión de inventarios.

- Aprendizaje estocástico: Este tipo de aprendizaje consiste básicamente en realizar cambios aleatorios en los valores de los pesos de las conexiones de la red y evaluar su efecto en términos del objetivo deseado y de distribuciones de probabilidad.

#### Aprendizaje no supervisado:

El aprendizaje no supervisado implica el uso de un conjunto de datos no etiquetados para descubrir patrones y relaciones intrínsecas entre los datos. En este tipo de aprendizaje, el modelo no recibe información sobre las respuestas correctas, sino que debe aprender a estructurar y organizar los datos por sí mismo. Algunas aplicaciones típicas de aprendizaje no supervisado incluyen la segmentación de clientes y la agrupación de imágenes similares.

Estas redes deben encontrar las características, regularidades, correlaciones o categorías que se puedan establecer entre los datos que se presenten en su entrada. Puesto que no hay un supervisor que indique a la red la respuesta que debe generar ante una entrada concreta, cabría preguntarse precisamente por lo que la red genera en estos casos. Existen varias posibilidades en cuanto a la interpretación de la salida de estas redes, que dependen de su estructura y del algoritmo de aprendizaje empleado

Se usan cuando no tenemos datos etiquetas o previamente procesados.

Se usa para segmentación e identificación de posibles agrupamientos entre los datos conocido también como clustering (ejemplo de redes SOM para identificar diferentes tipos de animales según sus características comunes) , también están las reglas de asociación que se encarga de descubrir relaciones entre variables en grandes conjuntos de datos y se definen como implicaciones de antecedentes relacionados con consecuentes en base a determinadas probabilidades de aparición de los campos (columnas) en cada una de las tuplas (filas) Encontrar items frecuentes (con una frecuencia mayor o igual a un determinado valor promedio calculado en base a la cantidad de veces que aparece ese item) pudiendo determinar por ejemplo la asociación que existe entre determinados productos presentes en algunas compras realizadas por clientes frecuentes.

Existen dos tipos que dan lugar a los siguientes aprendizajes:

- Aprendizaje Hebbiano: consiste básicamente en el ajuste de los pesos de las conexiones de acuerdo con la correlación (multiplicación en el caso de valores binarios +1 y -1) de los valores de activación (salidas) de las dos neuronas conectadas:

- Aprendizaje Competitivo y Cooperativo: las neuronas compiten (y cooperan) unas con otras con el fin de llevar a cabo una tarea dada. Con este tipo de aprendizaje se pretende que cuando se presente a la red una cierta información de entrada, sólo una de las neuronas de salida de la red, o una por un cierto grupo de neuronas, se active (alcance su valor de respuesta máximo).

Una vez generado el conjunto de datos o dataset se procede al entrenamiento del modelo (cuando hablamos de modelo se tomó de antemano la decisión de cual seria el mejor modelo aplicable al caso de estudio y con qué parámetros de acuerdo al modelo se procederá al entrenamiento) en este punto se define por ejemplo lenguaje, tipo de red neuronal, cantidad de neuronas, cantidad de capas, y otros parámetros dependiendo del modelo y de la herramienta utilizada. Muchos lenguajes traen librerías que tienen preprogramados los

modelos y simplemente debemos definir los parámetros del mismo. Qué son los que se utilizaran en el presente curso.

En este punto cabe resaltar que es probable más allá de experiencia trabajando con estos modelos, que se deban realizar varias pruebas ajustando los parámetros ya que los mismos no dependen solo del modelo si no de la más acertada representación de acuerdo al dominio del problema.

Por eso antes de pasar a producción hay un ciclo de pruebas, ajustes y contraste del comportamiento del modelo de acuerdo a valores definidos como potencialmente aceptables. El rigor en el registro de cada una de las pruebas es altamente recomendable donde se detallan cantidad de datos de entrenamiento, cantidad de datos de prueba, valores en los parámetros del modelo, tiempo de entrenamiento, error general obtenido, entre otros datos para determinar cuál es el conjunto de parametrización del modelo con el cual se logra el mejor comportamiento posible.

A partir de este punto es cuando se puede decir que pasamos a producción el modelo y empezamos a evaluar su comportamiento en el ámbito de explotación exponiéndose a nuevas situaciones.

Para que el modelo cumpla con la idea de aprendizaje continuo se debería pasar por procesos de corrección de errores tanto de predicción como de comportamiento en general, incorporando nuevos datos y pasando por procesos de entrenamiento y prueba. Se podría decir que cuanto más datos tenemos mejor y más preciso será el modelo (más adelante veremos los problemas de sobre entrenamiento que también existen)

## Cuestionario guía de lectura del material

1. ¿Qué es la inteligencia artificial?
2. ¿Qué es el machine learning?
3. ¿Qué es el deep learning?
4. ¿Cual es la clasificación de los tipos de aprendizaje automático? Es decir, las formas de aprender que tienen las máquinas hoy.
5. ¿Qué ámbitos de aplicación de la inteligencia artificial en general y del machine learning en particular se te ocurre que podrían ser de beneficio?
6. ¿Que es importante tener para poder avanzar sobre un proyecto de machine learning?

## Bibliografía:

1. Bagnato, J. i., (2020). Aprende Machine Learning en Español: Teoría + Práctica Python. Editorial Leanpub.
2. Britos, P. V., & García Martínez, R. (2009). Propuesta de Procesos de Explotación de Información. In *XV Congreso Argentino de Ciencias de la Computación*.
3. Chazallet, S. (2016). Python 3: los fundamentos del lenguaje. Ediciones ENI.
4. Geron, A., (2020). Aprende Machine Learning con Scikit-Learn, Keras y Tensor Flow: Conceptos, herramientas y técnicas para construir sistemas inteligentes. Editorial O'Reilly y Anaya



5. Hilera, J. R. y Martinez, V. J. (2000) Redes Neuronales Artificiales. Fundamentos. modelos y aplicaciones. Alfaomega Ed
6. JIMÉNEZ, R. O., (2021). Python a fondo. Editorial Marcombo
7. Kuna, H. D., Caballero, S., Rambo, A., Meinel, E., Steinhilber, A., Pautsch, J., ... & Villatoro, F. (2010). Avance en procedimientos de la explotación de información para la identificación de datos faltantes, con ruido e inconsistentes. In *XII Workshop de Investigadores en Ciencias de la Computación*.
8. Kuna, H., Pautsch, G., Rey, M., Cuba, C., Rambo, A., ... & Villatoro, F. (2012). Obtenido de COMPARACIÓN DE LA EFECTIVIDAD DE PROCEDIMIENTOS DE LA EXPLOTACIÓN DE INFORMACIÓN PARA LA IDENTIFICACIÓN DE OUTLIERS EN BASES DE DATOS:
9. Matthes, E. (2021) Curso intensivo de Python, 2ª edición: Introducción práctica a la programación basada en proyectos. Editorial Anaya Multimedia
10. Ochoa, M. A. (2004). Herramientas inteligentes para explotación de información. Trabajo Final Especialidad en Ingeniería de Sistemas Expertos url: <https://ri.itba.edu.ar/server/api/core/bitstreams/a848d640-0277-459d-9104-b37017309d31/content>