

## Revisión de conceptos básicos del aprendizaje automático y sus algoritmos de implementación

pp. 136 - 152

Review of the basic concepts of Machine Learning and its implementation algorithms

Jair Abadía Correa <sup>1</sup>  
Gustavo Castaño Soto <sup>2</sup>  
Bryan García <sup>3</sup>

### RESUMEN

El artículo presenta una introducción al aprendizaje automático, definiéndolo como un subcampo de la inteligencia artificial que permite a los sistemas aprender y mejorar automáticamente a partir de datos y experiencias previas. Se explican los conceptos básicos del aprendizaje automático, incluyendo el uso de algoritmos y modelos estadísticos para analizar y predecir patrones en los datos. Se presentan los diferentes tipos de aprendizaje automático, incluyendo el aprendizaje supervisado, no supervisado y por refuerzo. El artículo enfatiza en el aprendizaje supervisado, ya que por su aplicación es el más utilizado, explicando cómo funciona y detallando algunos de los algoritmos más comunes utilizados en este enfoque, como el árbol de decisiones, la regresión lineal y las máquinas de vectores de soporte. En general, el artículo ofrece una buena introducción al aprendizaje automático para aquellos que buscan entender mejor este campo en crecimiento y sus aplicaciones en el mundo real.

**Palabras Clave:** Machine Learning, Aprendizaje

Automático, Aprendizaje Supervisado, Aprendizaje No Supervisado

### Abstract

The article presents an introduction to machine learning, defining it as a subfield of artificial intelligence that enables systems to automatically learn and improve based on data and previous experience. The basics of machine learning are explained, including the use of algorithms and statistical models to analyze and predict patterns in data. The different types of machine learning are introduced, including supervised, unsupervised, and reinforcement learning. The article emphasizes supervised learning, since its application is the most widely used, explaining how it works and detailing some of the most common algorithms used in this approach, such as decision tree, linear regression, and support vector machines. Overall, the article provides a good introduction to machine learning for those looking to better understand this growing field and its real-world applications.

<sup>1</sup> Ingeniero Electrónico, Universidad del Cauca. Magister en Gestión Informática y Telecomunicaciones, Universidad ICESI. Docente Investigador, Unicatólica Cali. Correo electrónico: jabadia@unicatolica.edu.co.

<sup>2</sup> Ingeniero de Sistemas, Universidad San Buenaventura Cali. Magister en Gestión Informática y Telecomunicaciones, Universidad ICESI. Docente Investigador, Unicatólica Cali. Correo electrónico: gcastano@unicatolica.edu.co.

<sup>3</sup> Ingeniero electrónico, Universidad del Valle. Magister en Ingeniería, Universidad del Valle. Docente Investigador, Unicatólica Cali. Correo electrónico: bgarcia@unicatolica.edu.co.

**Keywords:** Machine Learning, Automatic Learning, Supervised Learning, No Supervised Learning

## Introducción

Machine Learning (ML), también conocido como Aprendizaje Automático es una rama de la Inteligencia artificial (IA). Se trata de un tipo de IA que no depende 100% de unas reglas y un programador, sino que la maquina va a adquirir un aprendizaje, pudiendo ella misma establecer sus propias reglas. Las computadoras aprenden a través de comandos proporcionados por programadores, estos comandos contienen una serie de pasos, condiciones y reglas que la computadora debe seguir y ejecutar, los algoritmos son el núcleo de todo proceso de Aprendizaje Automático.

Cada día se ven los beneficios del aprendizaje automático. En los últimos años la tecnología de reconocimiento facial permite que las plataformas de redes sociales ayuden a los usuarios a etiquetar y compartir fotos de amigos, los motores de recomendación, impulsados por el aprendizaje automático, sugieren qué películas o programas de televisión ver a continuación según las preferencias del usuario.

El aprendizaje automático es un campo en continuo desarrollo. Debido a esto, hay algunas consideraciones a tener en cuenta al trabajar con metodologías de aprendizaje automático o analizar el impacto de los procesos de aprendizaje automático.

En este documento, se mostrará la evolución en el tiempo del concepto de Aprendizaje Automático visto por diferentes autores, los conceptos más importantes relacionados con el aprendizaje automático que ayudaran a entender su funcionamiento, los métodos de Aprendizaje Automático tanto aprendizaje supervisado como no supervisado, y los enfoques algorítmicos comunes en el Aprendizaje Automático, En general, este artículo tiene como objetivo proporcionar una guía útil y accesible para entender el Aprendizaje Automático y su importancia en la actualidad, tomando referencias de fuentes

escritas por importantes autores que protagonizaron con sus aportes el desarrollo del Aprendizaje Automático.

## Metodología

Al igual que con cualquier concepto, una búsqueda en un motor de búsqueda como Google sobre aprendizaje automático abre una serie de foros, investigaciones académicas e inclusive información errada. El propósito de este documento fue identificar fuentes confiables que ayuden a simplificar y obtener una comprensión fiable de su definición, tener una comprensión de los conceptos asociados al aprendizaje automático: tipos de aprendizaje, métodos y algoritmos de aprendizaje. Este artículo se desarrolla estructurando el documento en las siguientes secciones:

- ¿Qué es el aprendizaje automático?
- Conceptos básicos de aprendizaje automático.
- Tipos de aprendizaje automático.
- Como funciona el aprendizaje automático supervisado.
- Algoritmos de aprendizaje automático de uso común

## ¿Qué es el Aprendizaje Automático o Machine Learning?

El término Machine Learning fue utilizado por primera vez por Arthur Samuel, considerado por muchos como el padre del aprendizaje automático, usó algoritmos para jugar ajedrez y backgammon, lo definió como el campo de estudio que le da a las computadoras la habilidad de aprender sin ser explícitamente programadas (1959). El termino ha venido adquiriendo gran importancia gracias al aumento de la capacidad de computación y al boom de los datos.

Tom M. Mitchell (1997) afirmó que el aprendizaje

automático es un área que estudia cómo construir programas de computadoras que mejoren su desempeño en alguna tarea gracias a la experiencia (p.15). Ésta se basa en las ideas de diversas disciplinas, como inteligencia artificial, estadística y probabilidad, teoría de la información, psicología y neurobiología, teoría de control y complejidad computacional. En su libro sobre la inteligencia, Jeff Hawkins creador del teléfono inteligente y de la Palm Pilot, y Sandra Blakeslee definían la inteligencia como la capacidad del cerebro de realizar predicciones precisas en función de patrones previamente observados, por ejemplo, el peso de un vaso que vamos a levantar o la reacción de los demás a nuestros actos, en base a los patrones almacenados en la memoria (el marco memoria-predicción) (2004). Ese mismo principio está detrás del Machine Learning, también conocido como aprendizaje automático.

138

Pedro Domingos: en su libro “The Master Algorithm” (2015), define el Machine Learning como la habilidad de un sistema para aprender a partir de datos, sin ser explícitamente programado. Ian Goodfellow, Yoshua Bengio, y Aaron Courville: En su libro “Deep Learning” (2016), definen el Machine Learning como un subcampo de la inteligencia artificial que se enfoca en el diseño y desarrollo de algoritmos y modelos que permiten a las computadoras aprender a partir de datos.

El Aprendizaje Automático puede ser definido como: La ciencia que permite que las computadoras aprendan y actúen como lo hacen los humanos, mejorando su aprendizaje a lo largo del tiempo de una forma autónoma, alimentándolas con datos e información en forma de observaciones e interacciones con el mundo real. (Faggella, 2020)

Para Alpaydin (2010, pág. 3) el objetivo del aprendizaje automático generalmente es comprender la estructura de los datos y ajustarlos en modelos que las personas puedan entender y utilizar. En la informática tradicional, los algoritmos son conjuntos de instrucciones explícitamente programadas que

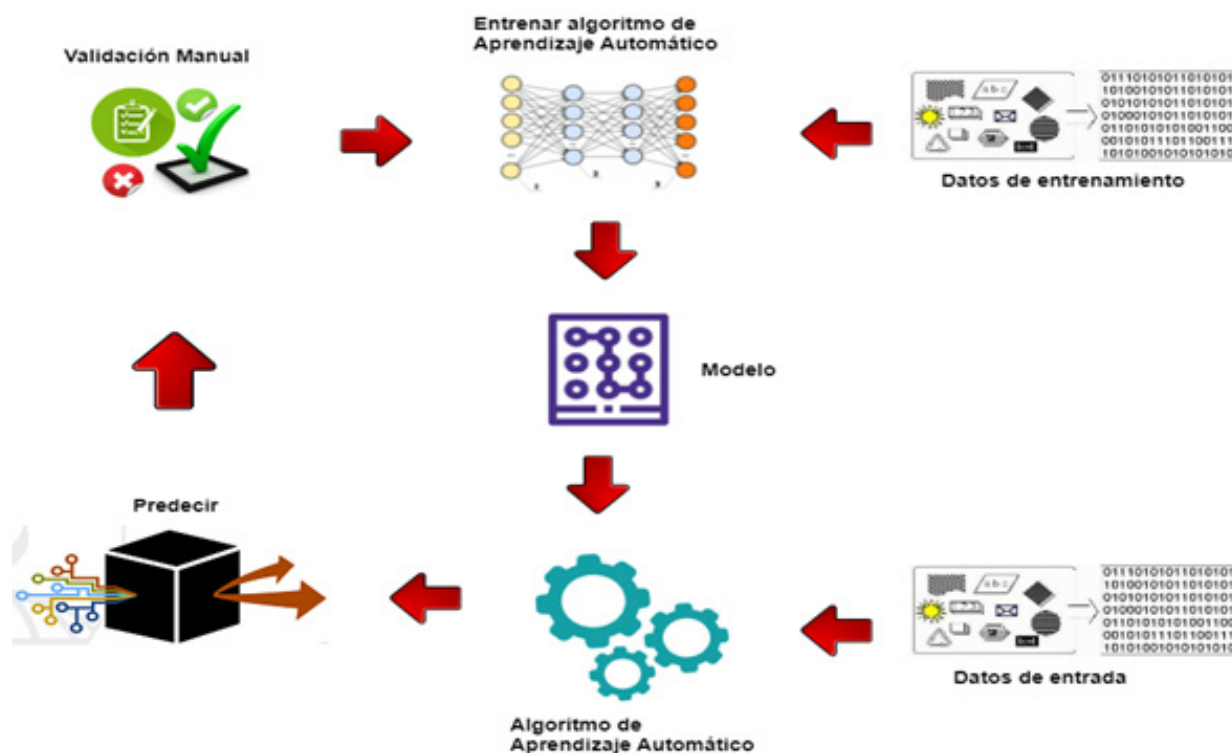
utilizan las computadoras para calcular o resolver problemas. En cambio, los algoritmos de aprendizaje automático permiten que las computadoras se entrenen en las entradas de datos y utilicen análisis estadísticos para generar valores que se encuentren dentro de un rango específico. Debido a esto, el aprendizaje automático facilita que las computadoras construyan modelos a partir de datos de muestra para automatizar los procesos de toma de decisiones en función de las entradas de datos.

El aprendizaje automático comienza con datos: números, fotos o texto, como transacciones, datos de series temporales de sensores o informes de ventas. Los datos se recopilan y preparan para usarse como datos de entrenamiento, o la información sobre la que se entrenará el modelo de aprendizaje automático (Raschka & Mirjalili, 2017, pág. 7). Cuantos más datos, mejor es el programa (Figura 1). Posteriormente los programadores eligen un modelo de aprendizaje automático, proporcionan los datos y dejan que el modelo se entrene para encontrar patrones o hacer predicciones. Con el tiempo, el programador humano también puede modificar el modelo, incluido el cambio de sus parámetros, para ayudar a impulsarlo hacia resultados más precisos (Brownlee, 2016, pág. 25). Algunos datos se extraen de los datos de entrenamiento para usarlos como datos de evaluación, lo que prueba qué tan preciso es el modelo de aprendizaje automático cuando se muestran nuevos datos. El resultado es un modelo que se puede utilizar en el futuro con diferentes conjuntos de datos (Geron, 2019, pág. 47).

### Conceptos básicos de aprendizaje automático.

**Algoritmo:** Son los procedimientos computacionales utilizados para aprender a partir de los datos y construir modelos de aprendizaje automático. Según Goodfellow, Bengio y Courville (2016), un algoritmo de aprendizaje automático es un procedimiento que toma un conjunto de datos como entrada y pro-

Figura 1. Algoritmo de Aprendizaje Automático



Fuente: Elaborado con base en Goel Ayush (2021)

duce un modelo como salida (p. 5). También podemos decir que es un conjunto de instrucciones precisas y bien definidas que describen cómo resolver un problema o realizar una tarea en un número finito de pasos (Cormen, Leiserson, C, Rivest, & Stein, 2009)

**Dato:** Los datos son un elemento fundamental del aprendizaje automático, ya que se utilizan para entrenar y evaluar los modelos de aprendizaje automático. Según Domingos (2015), los datos son el combustible del aprendizaje automático (p. 8). El aprendizaje automático se basa en el análisis de grandes cantidades de datos, los cuales son utilizados para entrenar modelos de aprendizaje automático. Los datos pueden ser estructurados o no estructurados, y pueden provenir de una variedad de fuentes. (Goodfellow, Bengio, & Courville, 2016), (Hastie, Tibshirani, & Friedman, 2009)

**Evaluación:** La evaluación es el proceso de medir

el rendimiento del modelo de aprendizaje automático en datos no vistos. Según James, Witten, Hastie y Tibshirani (2017), la evaluación es crucial en el aprendizaje automático, ya que permite seleccionar el modelo que mejor generaliza a datos no vistos (p. 7)

**Modelos:** Los modelos son la representación matemática del conocimiento adquirido a través de los datos. Según Hastie, Tibshirani y Friedman (2009), un modelo es una función matemática que toma las características de entrada y produce una salida” (p. 12). Los modelos de Machine Learning son algoritmos que utilizan los datos de entrenamiento para aprender patrones y hacer predicciones sobre datos nuevos. Existen muchos tipos diferentes de modelos, incluyendo modelos lineales, árboles de decisión, redes neuronales, y más. (Alpaydin E. , 2010), (Jordan & Mitchell, 2015)

**Entrenamiento y validación:** Para entrenar un modelo de Machine Learning, se utiliza un conjunto de datos de entrenamiento. Luego, se prueba el modelo utilizando un conjunto de datos de validación, para evaluar su capacidad para hacer predicciones precisas. (Murphy, 2012)

**Supervisado vs. no supervisado:** Los modelos de Machine Learning pueden ser supervisados o no supervisados. En el aprendizaje supervisado, se proporcionan datos etiquetados (con resultados esperados) para entrenar el modelo. En el aprendizaje no supervisado, el modelo debe encontrar patrones por sí mismo, sin etiquetas predefinidas. (Bishop, 2006), (Hastie, Tibshirani, & Friedman, 2009)

**Componentes:** Pedro Domingos, ha propuesto tres componentes clave del aprendizaje automático en su libro “The Master Algorithm” (2015) y manifiesta que todos los algoritmos de aprendizaje automático son combinaciones de estos tres componentes. (Figura 2)

**Representación:** La representación se refiere a cómo se representan los datos en el modelo

de aprendizaje automático. Según Domingos, “la representación es la clave del éxito del aprendizaje automático” (p. 69). Una buena representación de los datos puede hacer que el modelo de aprendizaje automático sea más eficiente y preciso.

**Evaluación:** La evaluación se refiere a cómo se mide el rendimiento del modelo de aprendizaje automático. Según Domingos, “la evaluación es fundamental para determinar si un modelo de aprendizaje automático es útil” (p. 74). La evaluación puede incluir métricas como la precisión, la sensibilidad y la especificidad.

**Optimización:** La optimización se refiere a cómo se ajustan los parámetros del modelo de aprendizaje automático para mejorar su rendimiento. Según Domingos, “la optimización es el proceso de ajuste de los parámetros de un modelo de aprendizaje automático para minimizar su error” (p. 81). La optimización puede ser realizada mediante algoritmos como el descenso de gradiente.

Figura 2. . Elementos Claves del Aprendizaje Automático

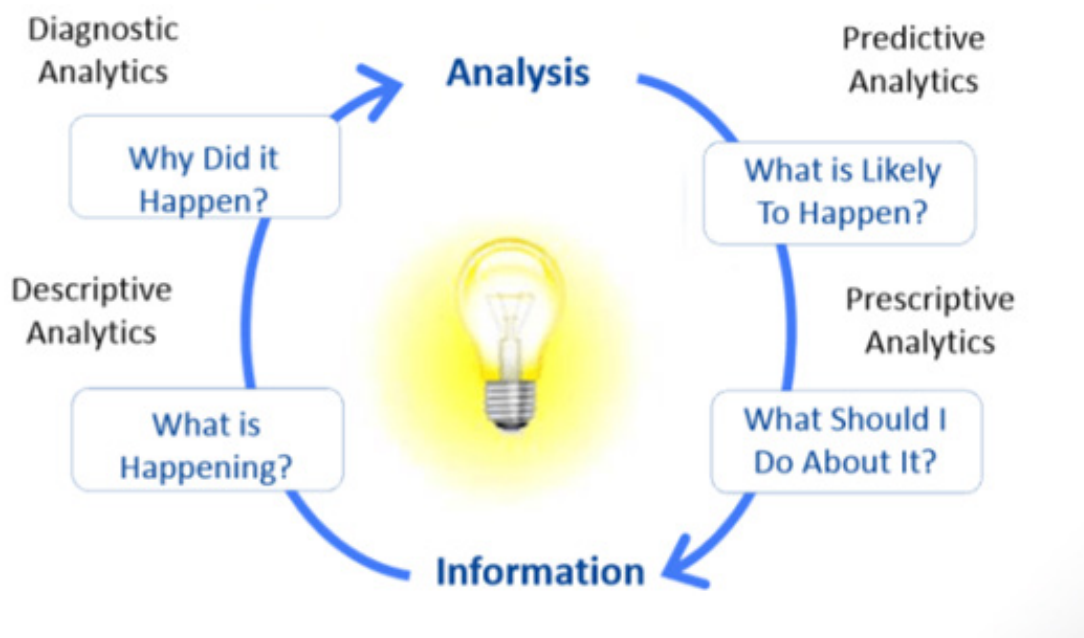
Representation	Evaluation	Optimization
Instances	Accuracy/Error rate	Combinatorial optimization
<i>K</i> -nearest neighbor	Precision and recall	Greedy search
Support vector machines	Squared error	Beam search
Hyperplanes	Likelihood	Branch-and-bound
Naive Bayes	Posterior probability	Continuous optimization
Logistic regression	Information gain	Unconstrained
Decision trees	K-L divergence	Gradient descent
Sets of rules	Cost/Utility	Conjugate gradient
Propositional rules	Margin	Quasi-Newton methods
Logic programs		Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian networks		
Conditional random fields		

Fuente: Dr. Pedro Domingo, University of Washington.

**Tipos de Análisis:** Los algoritmos exitosos de aprendizaje automático pueden hacer cosas diferentes, escribió Malone (2017) en un informe de investigación reciente sobre la IA y el futuro del trabajo que fue coescrito por la profesora del MIT y directora de CSAIL Daniela Rus y Robert Lau-

bacher, director asociado del Centro de Inteligencia Colectiva del MIT. Un sistema de aprendizaje automático puede buscar un Análisis Descriptivo, Análisis Diagnostico, Análisis Predictivo, Análisis Prescriptivo:

Figura 3. Análisis Descriptivo, Diagnostico, Predictivo y Pre-



Fuente: <https://radacad.com/business-understanding-for-machine-learning-descriptive-analysis>

**Análisis Descriptivo:** donde se busca resumir, clasificar, extraer reglas para para explicar lo que sucedió, Según Hair et al. (2019), el análisis descriptivo es “el proceso de resumir los datos que han sido recolectados y presentarlos de manera informativa y significativa” (p. 29). El análisis descriptivo puede incluir técnicas como la estadística descriptiva, la visualización de datos y el análisis de frecuencia.

**Análisis Diagnostico:** Según Montgomery y Runger (2018), es el proceso de determinar la causa raíz de un problema identificando las variables que tienen una relación significativa con la respuesta del problema” (p. 693). El análisis diagnóstico puede incluir técnicas como el análisis de correlación y la regresión.

**Análisis Predictivo:** Trata de enfoques para el objetivo de pronosticar datos futuros basados en datos pasados, Según James et al. (2017), el análisis predictivo es “el proceso de usar modelos estadísticos y algoritmos de aprendizaje automático para analizar datos y hacer predicciones sobre eventos futuros” (p. 4). El análisis predictivo puede incluir técnicas como la regresión, la clasificación y el agrupamiento. Y, por último, el análisis prescriptivo utiliza algoritmos de optimización y recomendación para responder a los posibles resultados.

**Análisis Prescriptivo:** Es una técnica de análisis de datos que se utiliza para recomendar acciones óptimas para lograr un resultado deseado. Según Davenport y Harris (2017), el análisis prescriptivo es “el uso de mod-

elos analíticos para recomendar acciones que maximicen o minimicen alguna función de negocio” (p. 4). El análisis prescriptivo puede incluir técnicas como la optimización, la simulación y el análisis de escenarios.

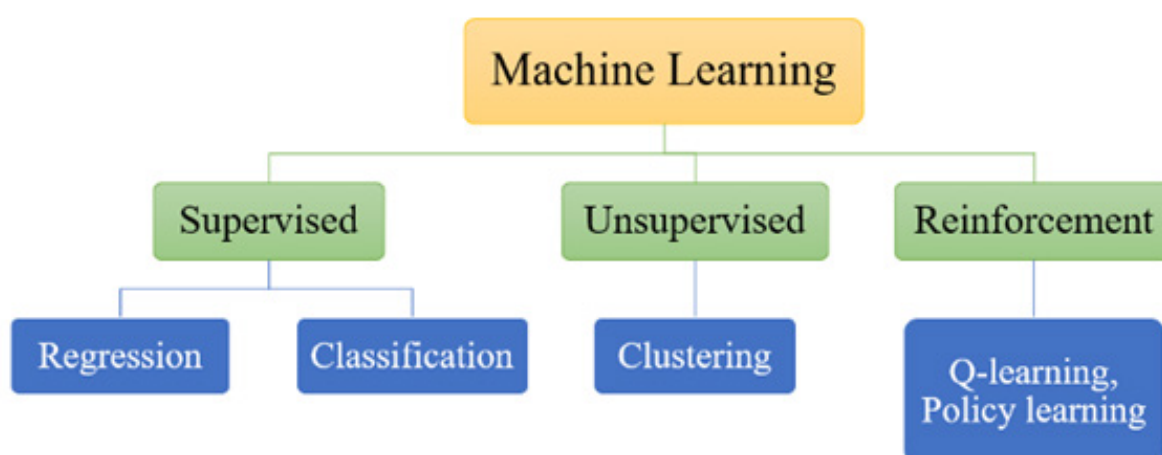
### Tipos de algoritmos de aprendizaje automático

Existen algunas variaciones de cómo definir los tipos de algoritmos de aprendizaje automático, pero comúnmente se pueden dividir en categorías según

su propósito y las categorías principales son las que aparecen en la figura 4 y se relacionan a continuación:

- Algoritmos de aprendizaje supervisado
- Algoritmos de Aprendizaje sin supervisión o no supervisado
- Algoritmos de Aprendizaje semisupervisado
- Algoritmos de Aprendizaje reforzado o por refuerzo

Figura 4. Tipos de Algoritmos de Aprendizaje Automático



Fuente: : [https://medium.com/@novus\\_afk/intro-to-ml-479f63452426](https://medium.com/@novus_afk/intro-to-ml-479f63452426)

### Algoritmos de Aprendizaje Automático Supervisado

La mayoría del aprendizaje automático práctico utiliza el aprendizaje supervisado, los algoritmos de aprendizaje supervisado son aquellos que aprenden a partir de un conjunto de datos de entrenamiento etiquetados y luego utilizan ese conocimiento para hacer predicciones sobre nuevos datos. Frank Rosenblatt fue el creador del Perceptrón, uno de los primeros algoritmos de aprendizaje supervisado utilizado para el reconocimiento de patrones. Según Alpaydin (2010), el aprendizaje supervisado se define como “la tarea de inferir una función a partir

de datos de entrenamiento etiquetados” (p. 9). El aprendizaje supervisado es donde tiene variables de entrada (x) y una variable de salida (Y) y usa un algoritmo para aprender la función de mapeo desde la entrada hasta la salida,  $Y = f(X)$ , El objetivo es aproximar la función de mapeo tan bien que cuando tenga nuevos datos de entrada (x) pueda predecir las variables de salida (Y) para esos datos (Hastie, Tibshirani, & Friedman, 2009).

Se llama aprendizaje supervisado porque el proceso de aprendizaje de un algoritmo a partir del conjunto de datos de entrenamiento puede considerarse como un maestro que supervisa el proceso de aprendizaje.

.Conocemos las respuestas correctas, el algoritmo iterativamente hace predicciones sobre los datos de entrenamiento y es corregido por el profesor. El aprendizaje se detiene cuando el algoritmo alcanza un nivel aceptable de rendimiento. (Alpaydin E. , 2010, pág. 10)

Algunos ejemplos populares de algoritmos de aprendizaje automático supervisado son: Regresión lineal para problemas de regresión, random forest para problemas de clasificación y regresión, y las máquinas de vectores (SVM) de soporte para problemas de clasificación (Muller & Guido, 2016, pág. 21). Los problemas de aprendizaje supervisado se pueden agrupar en problemas de clasificación y regresión (Burkov, 2019).

**Clasificación:** Un problema de clasificación es cuando la variable de salida es una categoría, como “rojo” o “azul” o “enfermedad” y “sin enfermedad”, los algoritmos de clasificación se pueden dividir en los siguientes tipos: (Geron, 2019, pág. 120).

- Regresión logística (Logistic Regression)
- K-vecinos más cercanos (K-Nearest Neighbours)
- Máquinas de vectores de soporte (Support Vector Machines)
- Núcleo SVM (Kernel SVM)
- Bayesiana ingenua (Naïve Bayes)
- Clasificación del árbol de decisión (Decision Tree Classification)
- Clasificación aleatoria de bosques (Random Forest Classification)

**Regresión:** Un problema de regresión es cuando la variable de salida es un valor real, como “dólares” o “peso”. A parte de la regresión lineal, los algoritmos de regresión se pueden dividir en los siguientes tipos: (Muller & Guido, 2016).

- Regresión lineal
- Regresión no lineal.
- Árboles de regresión (Ej: Random Forest).
- Máquinas de Vectores de Soporte para Regresión (SVR).
- Ridge Regression

Según Alpaydin (2010, pág. 49) un algoritmo de aprendizaje automático supervisado puede describirse a través de los siguientes pasos:

- Definir los objetivos del problema a resolver.
- Preprocesar datos: recopilar y etiquetar datos: Los conjuntos de datos deben recopilarse y organizarse, y luego etiquetarse con el resultado correcto (esto generalmente lo hacen humanos).
- Dividir los datos en un conjunto de entrenamiento y otro de validación (prueba).
- Entrenar el modelo.
- Evaluar el modelo, con la ayuda del conjunto de validación.
- Mejorar el modelo ajustando varios parámetros del modelo hasta que el modelo funcione bien en los datos de entrenamiento.
- Implementar el modelo y supervisararlo en tiempo real integrándolo en un sistema o producto más grande.

### Algoritmos de Aprendizaje Automático No Supervisado

Los algoritmos de aprendizaje no supervisado son aquellos que aprenden a partir de un conjunto de datos de entrenamiento no etiquetados y buscan patrones o estructuras en los datos. Según Bish-

op (2006), el aprendizaje no supervisado se define como la tarea de descubrir la estructura inherente en los datos no etiquetados (p. 485). Ejemplos de algoritmos de aprendizaje no supervisado incluyen el clustering, la reducción de la dimensionalidad y el análisis de componentes principales. El aprendizaje no supervisado es donde solo tiene datos de entrada (X) y ninguna variable de salida correspondiente (Geron, 2019). El objetivo del aprendizaje no supervisado es descubrir estructuras en los datos que permitan una representación más compacta o una mejor comprensión de los mismos (Bishop, 2006, pág. 495)

Se denomina aprendizaje no supervisado porque, a diferencia del aprendizaje supervisado anterior, no hay respuestas correctas y no hay un maestro. El aprendizaje no supervisado es una forma de aprender patrones en los datos sin tener que especificar de antemano cuáles son esos patrones. Es la forma más natural de aprender, ya que es lo que hacen los seres humanos y muchos animales: aprenden de lo que sucede a su alrededor, sin necesidad de recibir retroalimentación explícita. (Hinton, 2015). Cada vez que construimos un modelo que trata de aprender algo de los datos sin que se le haya proporcionado una respuesta correcta explícita, se trata de aprendizaje no supervisado (Bengio, 2018, pág. 1). Los problemas de aprendizaje no supervisados se pueden agrupar en problemas de agrupación y asociación.

**Agrupamiento:** un problema de agrupamiento es donde desea descubrir los agrupamientos inherentes en los datos, como agrupar clientes por comportamiento de compra (Alpaydin E. , 2010, pág. 218).

**Asociación:** un problema de aprendizaje de reglas de asociación es donde desea descubrir reglas que describan grandes

porciones de sus datos, como que las personas que compran X también tienden a comprar Y. Algunos ejemplos populares de algoritmos de aprendizaje no supervisados son: k-medias para problemas de agrupamiento y Algoritmo a priori para problemas de aprendizaje de reglas de asociación (Tang, Steinbach, & Kumar, 2005).

### Algoritmos de Aprendizaje Automático Reforzado o Por Refuerzo

Según Sutton y Barto (2018), referentes en el campo del aprendizaje por refuerzo definen este como la tarea de aprender qué acciones tomar en qué situaciones para maximizar una señal de recompensa a largo plazo (p. 1). El método tiene como objetivo utilizar las observaciones recopiladas de la interacción con el entorno para tomar acciones que maximicen la recompensa o minimicen el riesgo. El algoritmo de aprendizaje por refuerzo (llamado el agente) aprende continuamente del entorno de forma iterativa. En el proceso, el agente aprende de sus experiencias del entorno hasta que explora la gama completa de estados posibles. Ejemplos de algoritmos de aprendizaje por refuerzo incluyen el Q-learning y la programación dinámica. Los algoritmos de aprendizaje por refuerzo son aquellos que aprenden a partir de la retroalimentación que reciben en forma de recompensas o castigos.

Según Kaelbling (1996), el aprendizaje por refuerzo es un tipo de aprendizaje automático y, por lo tanto, también una rama de la inteligencia artificial. Permite que las máquinas y los agentes de software determinen automáticamente el comportamiento ideal dentro de un contexto específico, para maximizar su rendimiento. Se requiere una retroalimentación de recompensa simple para que el agente aprenda su comportamiento; esto se conoce como la señal de refuerzo.

De acuerdo a Sutton y Barto (2018) en el problema, se supone que un agente decide la mejor acción para seleccionar en función de su estado actual. Cuando se repite este paso, el problema se conoce como Proceso de Decisión de Markov. Para producir programas inteligentes (también llamados agentes), el aprendizaje por refuerzo pasa por los siguientes pasos:

- El agente observa el estado de entrada.
- La función de toma de decisiones se utiliza para hacer que el agente realice una acción.
- Una vez realizada la acción, el agente recibe una recompensa o refuerzo del entorno.
- Se almacena la información del par estado-acción sobre la recompensa.

Entre la lista de algoritmos comunes se encuentran: Q-aprendizaje, Diferencia Temporal (TD) y Redes antagonicas profundas

### Algoritmos de Aprendizaje Automático semisupervisados

En este tipo especial los algoritmos de aprendizaje semisupervisado, a diferencia de los supervisados donde las etiquetas están presentes para todas las observaciones, y los no supervisados donde no hay etiquetas para todas las observaciones en el conjunto de datos, el aprendizaje semisupervisado se encuentra entre estos dos. Según Smith & Brown (2010) en muchas situaciones prácticas, el costo de etiquetar es bastante alto, ya que requiere expertos humanos calificados para hacerlo. Entonces, en ausencia de etiquetas en la mayoría de las observaciones, pero presentes en pocas, los algoritmos semisupervisados son los mejores candidatos para la construcción del modelo. Estos métodos explotan la idea de que, aunque se desconocen las pertenencias a grupos de los datos no etiquetados, estos datos contienen infor-

mación importante sobre los parámetros del grupo.

El aprendizaje semisupervisado ofrece un término medio entre el aprendizaje supervisado y el no supervisado. Durante el entrenamiento, utiliza un conjunto de datos etiquetados más pequeño para guiar la clasificación y la extracción de características de un conjunto de datos más grande sin etiquetar. El aprendizaje semisupervisado puede resolver el problema de no tener suficientes datos etiquetados para un algoritmo de aprendizaje supervisado. También ayuda si es demasiado costoso etiquetar suficientes datos (Chapelle, Scholkopf, & Zien, 2006).

Entonces se pueden utilizar técnicas de aprendizaje no supervisado para descubrir y aprender la estructura de las variables de entrada. También se pueden usar técnicas de aprendizaje supervisado para hacer las mejores predicciones para los datos no etiquetados, devolver esos datos al algoritmo de aprendizaje supervisado como datos de entrenamiento y usar el modelo para hacer predicciones sobre nuevos datos no vistos (Alpaydin E. , 2010).

### ¿Cómo funciona el algoritmo de aprendizaje supervisado?

En el aprendizaje supervisado, los modelos se entrenan utilizando conjuntos de datos etiquetados, donde el modelo aprende sobre cada tipo de datos. Cuando se finaliza el proceso de entrenamiento, el modelo se prueba sobre la base de datos de prueba, que es un subconjunto de los datos de entrenamiento, para luego tratar de predecir el resultado.

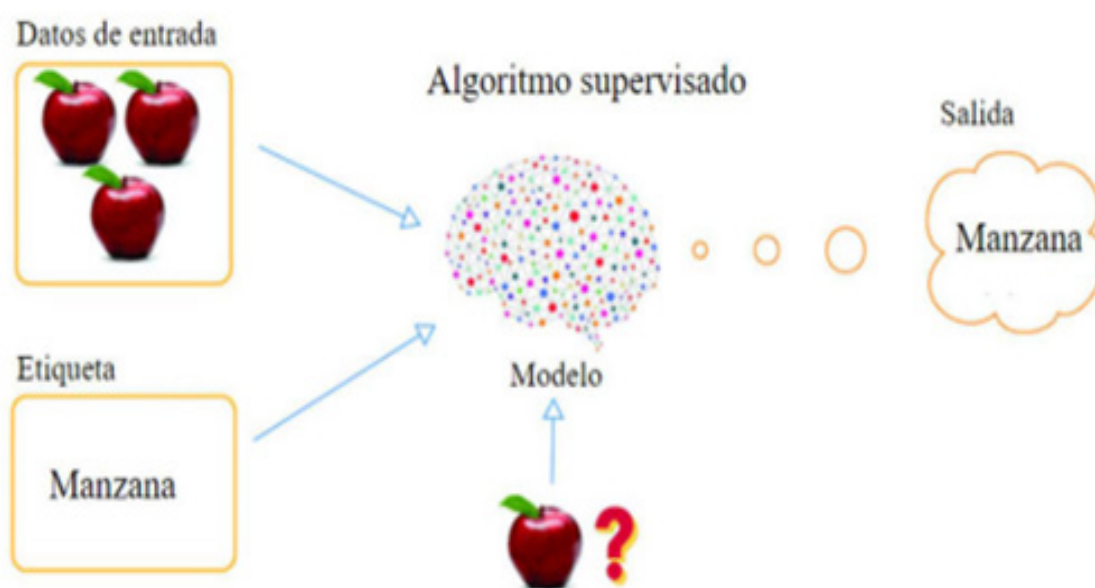
Según Alpaydin (2010, pág. 47) el funcionamiento del aprendizaje supervisado se puede entender a través de las siguientes situaciones: Si se desea identificar frutas, se pueden entregar un conjunto de imágenes de frutas, debidamente etiquetadas con sus nombres. Por ejemplo, se tendría que las imágenes de plátanos poseen las siguientes características: son amarillos, largos y curvos, las manzanas son rojas y

redondas (figura 5), las naranjas son anaranjadas y redondas.

De acuerdo con lo anterior, el modelo, basado en las características de una fruta, estará entrenado para identificar cualquier fruta basado en sus descriptores. Si se presenta una manzana, con seguridad el

modelo sabrá que no es de color anaranjado y tampoco amarillo y además que no es larga y curva, por lo que descarta las naranjas y los plátanos. Entonces la salida es una manzana cuyos descriptores son redonda y de color rojo (Muller & Guido, 2016, pág. 14).

Figura 5. Ejemplo 1 Algoritmo Supervisado



Fuente: López, B. Estudio de Ideología Política en Redes Sociales a través de Machine Learning.

Raschka y Mirjalili explican: (2017) tenemos los datos de una serie de figuras geométricas, en la que tenemos: cuadrados, rectángulos, triángulos y polígonos. Lo primero que se debe realizar es entrenar el modelo, especificando las características de cada figura:

Las figuras con cuatro lados, y todos de dimensión igual, se marcarán con Cuadrado.

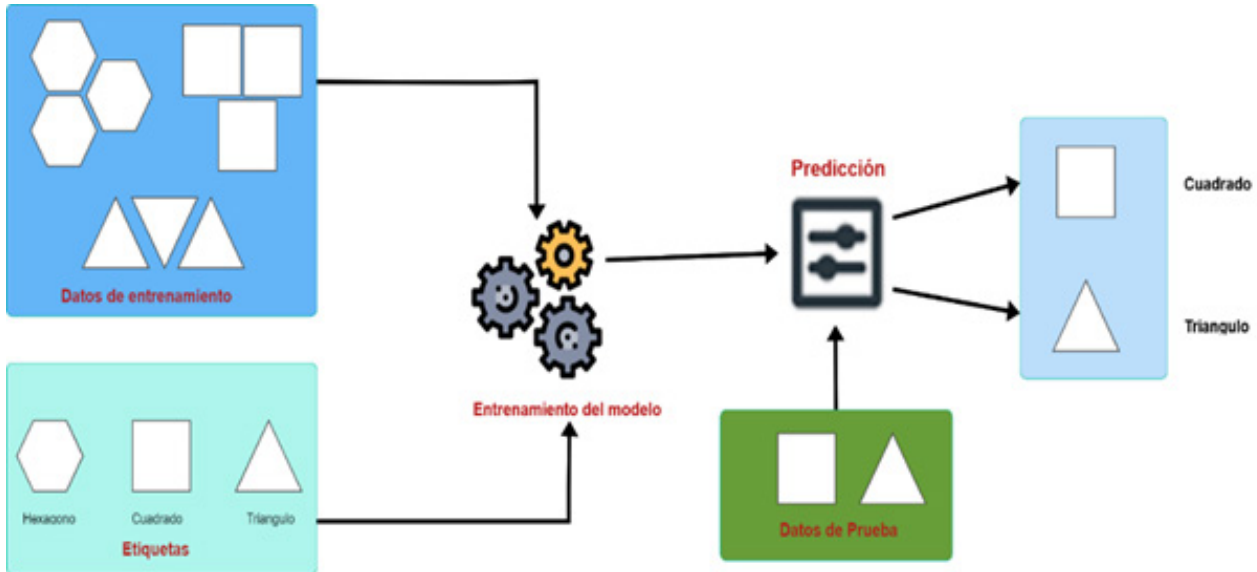
Las figuras con cuatro lados, y de dimensión diferente, se marcarán como rectángulo.

Las figuras con tres lados, se marcarán con Cuadrado.

Las figuras con seis lados, y todos de dimensión igual, se marcarán como Hexágono.

El siguiente paso es probar el modelo usando unos datos de prueba. El objetivo del modelo es determinar que figura geométrica es. Con esto la Máquina ya está entrenada para la identificación de las figuras geométricas, y cuando se le introduzcan los valores de una figura, predice el nombre de la figura (Figura 6).

Figura 6. Ejemplo 2 Algoritmo Supervisado



Fuente: López, B. Estudio de Ideología Política en Redes Sociales a través de Machine Learning.

De acuerdo con varias fuentes se pueden considerar los siguientes pasos para aplicar un modelo de aprendizaje supervisado:

- Determinar el tipo de conjunto de datos de entrenamiento
- Reunir los datos de entrenamiento etiquetados.
- Dividir los datos de entrenamiento: datos de entrenamiento, datos de prueba y datos de validación.
- Definir las características (preguntas) de entrada de los datos de entrenamiento, para que el modelo pueda predecir con alta precisión la salida.
- Determinar el algoritmo adecuado para el modelo, como puede ser SVM (Máquinas de vectores de soporte), Árbol de decisión, etc.
- Ejecutar el algoritmo sobre los datos de entrenamiento. A veces se necesitan datos de validación como parámetros de control.

- Evaluar la precisión del modelo. Si el modelo predice el resultado correcto, significa que nuestro modelo es exacto.

## Algoritmos de aprendizaje automático de uso común

### Regresión lineal

La regresión lineal es uno de los algoritmos de aprendizaje automático más simples y populares recomendados por un científico de datos. Se utiliza para el análisis predictivo al hacer predicciones para variables reales como experiencia, salario, costo, etc. Es un enfoque estadístico que representa la relación lineal entre dos o más variables, ya sean dependientes o independientes, por lo que se denomina regresión lineal. Muestra los cambios de valor de la variable dependiente con respecto a la variable independiente, y la pendiente de este gráfico se denomina Línea de regresión. (Muller & Guido, 2016, pág. 35). Los tipos de regresión lineal son: Regresión lineal simple, Regresión lineal múltiple (Weisberg, 2014, pág. 1)

Aplicaciones de la regresión lineal: La regresión lineal es útil para evaluar las tendencias comerciales y los pronósticos, como la predicción del salario de una persona en función de su experiencia, la predicción de la producción de cultivos en función de la cantidad de lluvia (Montgomery, Peck, & Vining, 2012, pág. 2).

## Regresión logística

La regresión logística es un subconjunto de la técnica de aprendizaje supervisado. La regresión logística se utiliza cuando la variable de respuesta es binaria o categórica, nominal u ordinal. Nos ayuda a predecir la salida de variables dependientes categóricas usando un conjunto dado de variables independientes (Hosmer, Lemeshow, & Sturdivant, 2013, pág. 1). Sin embargo, puede ser Binario (0 o 1) así como Booleano (verdadero/falso), pero en lugar de dar un valor exacto, da un valor probabilístico entre 0 o 1. Es muy similar a la Regresión Lineal, dependiendo de su uso en el modelo de aprendizaje automático. Así como la regresión lineal se usa para resolver problemas de regresión, la regresión logística es útil para resolver problemas de clasificación (Raschka & Mirjalili, 2020).

La regresión logística se puede expresar como una curva en forma de S denominada funciones sigmoideas. Predice dos valores máximos (0 o 1).

Tipos de regresión logística: Binomial, Multinomial, Ordinal.

## K Vecino más cercano (KNN)

Para Bishop (2006) es uno de los algoritmos de aprendizaje automático más simples y efectivos que se incluyen en las técnicas de aprendizaje supervisado. Es útil para resolver problemas de regresión y de clasificación. Asume la similitud entre los nuevos datos y los datos disponibles y coloca los nuevos

datos en la categoría que es más similar a las categorías disponibles. También se conoce como algoritmos de aprendizaje perezoso porque no aprende del conjunto de entrenamiento inmediatamente; en cambio, almacena el conjunto de datos y, en el momento de la clasificación, realiza una acción en el conjunto de datos.

Suponiendo que se tienen algunos conjuntos de imágenes de gatos y perros y se quiere identificar si una nueva imagen es de un gato o un perro. Entonces, el algoritmo KNN es la mejor manera de identificar al gato a partir de los conjuntos de datos disponibles porque funciona con medidas de similitud. Por lo tanto, el modelo KNN comparará la nueva imagen con las imágenes disponibles y colocará la salida en la categoría del gato. Los algoritmos KNN se utilizan en los siguientes campos: Diagnóstico médico, Comprobación de puntuación de crédito, Edición de texto, Reserva de hotel, etc. (Raschka & Mirjalili, 2017, pág. 89)

## Árbol de decisión

Árbol de Decision (Decision Tree) es otro tipo de técnica de aprendizaje automático que se incluye en el aprendizaje supervisado. Similar a KNN, el árbol de decisión también nos ayuda a resolver problemas de clasificación y regresión, pero se prefiere principalmente para resolver problemas de clasificación (James, Witten, Hastie, & Tibshirani, 2017).

El nombre de árbol de decisión se debe a que consta de un clasificador con estructura de árbol en el que los atributos están representados por nodos internos, las reglas de decisión están representadas por ramas y el resultado del modelo está representado por cada hoja de un árbol. El árbol comienza en el nodo de decisión, también conocido como nodo raíz, y termina en el nodo hoja (Murphy, 2012). Los nodos de decisión ayudan a tomar cualquier decisión, mientras que las hojas se utilizan para determinar el resultado de esas decisiones. Un Árbol de Decisión

es una representación gráfica para obtener todos los resultados posibles de un problema o decisión dependiendo de ciertas condiciones dadas (Tang, Steinbach, & Kumar, 2005).

### Naive Bayes (bayesiana ingenua)

El algoritmo naive Bayes es uno de los algoritmos de aprendizaje automático más simples y efectivos que se encuentran bajo la técnica de aprendizaje supervisado. Se basa en el concepto del Teorema de Bayes, utilizado para resolver problemas relacionados con la clasificación. Ayuda a construir modelos de aprendizaje automático rápidos que pueden hacer predicciones rápidas con mayor precisión y rendimiento. Se prefiere principalmente para la clasificación de texto que tiene conjuntos de datos de entrenamiento de alta dimensión (Mehta, 2019, pág. 96).

Se utiliza como clasificador probabilístico, lo que significa que predice sobre la base de la probabilidad de un objeto. La filtración de spam, el análisis sentimental y la clasificación de artículos son algunas aplicaciones importantes del algoritmo Naive Bayes (Jurafsky & Martin, 2020, pág. 375). Este método se basa en el teorema de Bayes, también conocido como Regla de Bayes o Ley de Bayes (Bishop, 2006, pág. 181).

### Conclusiones

La definición de Machine Learning o Aprendizaje Automático ha ido evolucionando de la mano con la evolución de la tecnología y su aplicación en diferentes campos, como la robótica, la inteligencia artificial y la ciencia de datos, a medida que se han desarrollado nuevas técnicas y algoritmos y se ha logrado una mayor comprensión de cómo funcionan los modelos y cómo pueden ser aplicados en diferentes campos. En sus comienzos la definición se enfocaba en la capacidad de las máquinas de aprender de los datos sin ser programadas explíci-

tamente, sin embargo, con el tiempo, se han ido incorporando nuevas técnicas, como el aprendizaje profundo y el aprendizaje por refuerzo, lo que ha llevado a una definición más amplia y compleja del concepto.

La comprensión de conceptos fundamentales en el Aprendizaje Automático como: algoritmo, datos, evaluación, modelos, optimización es esencial para desarrollar modelos precisos y generalizables, lo que puede tener un gran impacto en una amplia gama de campos, desde la medicina y la ingeniería hasta el comercio y la tecnología.

el Aprendizaje Automático es una disciplina que se enfoca en el desarrollo de algoritmos y modelos que pueden aprender de los datos para realizar tareas específicas sin ser programados explícitamente. La calidad de los modelos de Machine Learning depende en gran medida de los datos de entrenamiento utilizados, por lo que la selección y preprocesamiento adecuado de los datos es un factor crítico en el éxito de los modelos. Además, la evaluación rigurosa de los modelos es esencial para garantizar que sean efectivos y generalizables en datos no vistos.

La optimización de los modelos es un proceso crítico que implica ajustar los parámetros del modelo para maximizar su precisión y minimizar el error en la predicción. Además, la representación adecuada de los datos es fundamental para garantizar que el modelo pueda aprender de manera efectiva a partir de los datos.

Gracias a su capacidad para aprender de los datos, el Aprendizaje Automático ha demostrado ser una herramienta poderosa en la toma de decisiones, la identificación de patrones y la resolución de problemas complejos en una amplia variedad de campos, desde la medicina y la ciencia de datos hasta la robótica y la inteligencia artificial. Es un campo que ha experimentado un rápido crecimiento y evolución en las últimas décadas. Además, la evolución constante de las técnicas y algoritmos de Aprendizaje Automáti-

co promete desbloquear aún más su potencial en el futuro, lo que hace que sea un campo emocionante y lleno de posibilidades.

## Referencias

Alpaydin, E. (2010). Introduction to machine learning. MIT press.

Alpaydin, E. (2010). Supervised Machine Learning.

Ansari, Y. (10 de 2020). medium.com. Obtenido de Introduction to Machine Learning (ML): [https://medium.com/@novus\\_afk/intro-to-ml-479f63452426](https://medium.com/@novus_afk/intro-to-ml-479f63452426)

Bengio, Y. (2018). The Consciousness Prior.

Bishop, C. (2006). Pattern recognition and machine learning. Springer.

150

Brown, S. (21 de Abril de 2021). MIT . Obtenido de Machine learning, explained: 2021

Brownlee, J. (15 de Agosto de 2015). Machine Learning Mastery. Obtenido de Basic Concepts in Machine Learning: <https://machinelearningmastery.com/basic-concepts-in-machine-learning/>

Brownlee, J. (2016). Machine Learning Mastery With Python. Machine Learning Mastery LLC.

Brownlee, J. (2020 de Agosto de 2019). Machine Learning Mastery. Obtenido de A Tour of Machine Learning Algorithms: <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>

Burkov, A. (2019). The Hundred-Page Machine Learning Book. Andriy Burkov.

Chapelle, O., Scholkopf, B., & Zien, A. (2006). Semi-supervised learning (1st ed.). The MIT Press.

Cormen, T., Leiserson, C. C., Rivest, R., & Stein, C. (2009). Cormen, T. H., Leiserson, C. E.,

Rivest, R. L., & Stein, C. Introduction to algorithms. MIT press.

Davenport, T., & Harris, J. G. (2017). Competing on analytics: Updated, with a new introduction. Harvard Business Review Press.

Domingos, P. (2015). The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books.

Faggella, D. (26 de Febrero de 2020). Emerj - The AI Research and Advisory Company. Obtenido de What is Machine Learning?: <https://emerj.com/ai-glossary-terms/what-is-machine-learning/>

Fumo, D. (15 de Junio de 2017). Towards Data Science. Obtenido de Types of Machine Learning Algorithms You Should Know: <https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861>

Geron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.

Goel, A. (16 de Mayo de 2016). DEV Community . Obtenido de INTRODUCTION TO MACHINE LEARNING: <https://dev.to/1801ayush/introduction-to-machine-learning-3mof>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (Adaptive Computation and Machine Learning series). The MIT Press.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate Data Analysis. Cengage Learning.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learn-

- ing: data mining, inference, and prediction. Springer Science & Business Media. Springer.
- Hawkins, J., & Blakeslee, S. (2004). *On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*. St. Martin's Griffin.
- Hinton, G. (2015). *Deep learning*.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. (2013). *Applied logistic regression* (3rd ed.). John Wiley & Sons.
- IBM Cloud Education. (15 de Julio de 2020). IBM Cloud Learn Hub. Obtenido de What is Machine Learning?: <https://www.ibm.com/cloud/learn/machine-learning>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning: With applications in R*. Springer.
- JavatPoint. (s.f.). Javatpoint. Obtenido de Basic Concepts in Machine Learning: <https://www.javatpoint.com/basic-concepts-in-machine-learning>
- Jordan, M., & Mitchell, T. (2015). Jordan, M. I., & Mitchell, T. M. *Machine learning: Trends, perspectives, and prospects*. *Science*, 349(6245), 255-260.
- Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd edition). Pearson Education.
- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4, 237-285.
- Leila Etaati. (26 de febrero de 2018). RADACAD. Obtenido de Business Understanding for Machine Learning – Descriptive Analysis: <https://radacad.com/business-understanding-for-machine-learning-descriptive-analysis>
- Malone, T., Rus, D., & L'Aubacher, R. (2017). *Cómo cambiará la IA el futuro del trabajo*. Harvard Business Review.
- Mehta, R. (2019). *Python Machine Learning: A Guide for Beginners*. Packt Publishing.
- Mitchel, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Montgomery, D. C., & Runger, G. C. (2018). *Applied statistics and probability for engineers* (7th ed.). John Wiley & Sons.
- Montgomery, D. C., Peck, E. A., & Vining, G. (2012). *Introduction to linear regression analysis*. John Wiley & Sons.
- Muller, A. C., & Guido, S. (2016). *Introduction to Machine Learning with Python*. O'Reilly Media, Inc.
- Murphy, K. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Nabi, J. (15 de Agosto de 2018). *Towards Data Science*. Obtenido de Machine Learning — Fundamentals: <https://towardsdatascience.com/machine-learning-basics-part-1-a36d38c7916#:~:text=Machine%20Learning%20is%20an%20application,be%20at%20least%20human%20level.&text=In%20order%20to%20perform%20the,from%20the%20data%2Dset%20provided>.
- Oracle. (s.f.). oracle.com. Obtenido de What is Machine Learning?: <https://www.oracle.com/ng/artificial-intelligence/machine-learning/what-is-machine-learning/>
- Raschka, S., & Mirjalili, V. (2017). *Python Machine*

Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. Packt Publishing Ltd.

Raschka, S., & Mirjalili, V. (2020). Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), 210-229.

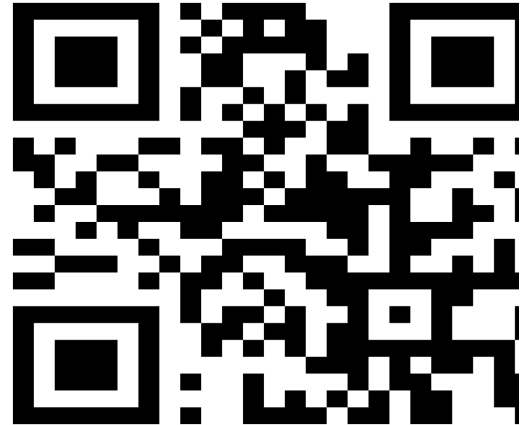
Smith, J., & Brown, A. (2010). Algoritmos de aprendizaje automático semi-supervisado. Journal of Machine Learning Research, 11, 1235-1278.

152

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

Tang, P.-N., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining. Pearson Education.

Weisberg, S. (2014). Applied Linear Regression (4th ed.). John Wiley & Sons.



### Podcast

Revisión de conceptos básicos del aprendizaje automático y sus algoritmos de implementación

Jair Abadía Correa  
Gustavo Castaño Soto  
Brayan García