

Machine Learning

Ing. Remigio Hurtado Ortiz, PhD.

Presentación

- Ingeniería de Sistemas, Universidad Politécnica Salesiana, Ecuador, 2012
- Maestría en Software y Tecnologías de la Información, Instituto Tecnológico y de Estudios Superiores de Monterrey, México, 2014
- Máster Universitario en Ciencias y Tecnologías de la Computación, Universidad Politécnica de Madrid, España, 2017
- Doctorado en Ciencias y Tecnologías de la Computación para Smart Cities, Universidad Politécnica de Madrid, España, 2020



Cisco Networking Academy
Instructor Excellence Advanced 2018



Cisco Networking Academy
Instructor Excellence Expert 2019

- Miembro del Grupo de Investigación en Inteligencia Artificial y Tecnologías de Asistencia (GI-IATa)
- Áreas y líneas de investigación: Inteligencia Artificial, Ciencia de Datos, Machine Learning, Deep Learning, Sistemas de Recomendación, Procesamiento de Lenguaje Natural.
- Correo institucional: rhurtadoo@ups.edu.ec
- Whatsapp: 0999922672

Objetivos

Objetivo general

- Investigar, experimentar y aplicar modelos computacionales de aprendizaje automático con el objetivo de predecir o identificar comportamientos sobre un conjunto de datos o ejemplos de entrada y para una mejor toma de decisiones.

Objetivos específicos

- Conocer los fundamentos de ML: definiciones, tipos de aprendizaje, aplicaciones, clasificación de algoritmos, proceso de evaluación y medidas de calidad estándares
- Aplicar varios clasificadores a un conjunto de datos
- Evaluar los clasificadores y elegir el más adecuado para el conjunto de datos
- Descubrir las variables o características más relevantes de los datos para alcanzar una mejor clasificación y reducir las dimensiones en que se busca una solución.
- Investigar, experimentar y aplicar otras técnicas alternas de aprendizaje automático como aprendizaje no supervisado, aprendizaje profundo, aprendizaje por refuerzo, aprendizaje por métodos de conjunto, filtrado colaborativo, entre otras.

Otros objetivos específicos

- Identificar métodos y pipelines de análisis siguiendo procesos del ciclo de vida de ciencia de datos
- Identificar análisis de datos con métodos basados en probabilidad y estadística
- Identificar métodos de machine learning con aprendizaje supervisado y no supervisado
- Identificar métodos con procesamiento de lenguaje natural
- Identificar métodos con Redes Neuronales y *Deep Learning*
- Identificar métodos con aprendizaje moderno: métodos ensamble y aprendizaje por refuerzo
- Identificar métodos para sistemas de recomendación

Salidas profesionales

Dentro de la Inteligencia Artificial como:

- Ingeniero de Machine Learning
- Ingeniero especialista en Visión Artificial
- Desarrollador de algoritmos y servicios de IA
- Desarrollador de software inteligente

Dentro de Ciencia de Datos y Big Data como:

- Administrador y analista de datos masivos
- Científico de datos Junior

En combinación con la experiencia profesional le permitirán desempeñarse como:

- Científico de datos Senior
- Ingeniero de datos

- Fundamentos de Inteligencia Artificial
- Análisis de Datos, Ciencia de Datos y Machine Learning
- Selección de Variables y Preprocesamiento de Datos
- Procesamiento de Lenguaje Natural y Procesamiento de Imágenes
- IA basada en técnicas estadísticas y teoría de la probabilidad: Regresión Lineal, Cadenas de Markov y Redes

Tópicos de Inteligencia Artificial

Tópicos de Aprendizaje Automático

Machine Learning

1. Definiciones
2. Tipos de aprendizaje
3. Aplicaciones
4. Clasificación de algoritmos

Definiciones

"La inteligencia artificial es una herramienta poderosa para ayudar a las personas a resolver problemas, pero siempre será necesario el juicio humano para evaluar la solución final" - Elon Musk

"La inteligencia artificial es la capacidad de una máquina para imitar el comportamiento humano" - Marvin Minsky

"El aprendizaje automático es una disciplina que desarrolla algoritmos que permiten a las computadoras aprender. Más específicamente, el aprendizaje automático se ocupa del desarrollo de técnicas que pueden descubrir automáticamente patrones en los datos y luego utilizar estos patrones para hacer predicciones o tomar decisiones. Estas técnicas pueden incluir algoritmos de clasificación, regresión, clustering y otros métodos que son fundamentales en el campo del machine learning" (Bishop C, 2006).

El aprendizaje automático aborda los desafíos y las oportunidades presentados por el **análisis de datos masivos** para **modelar datos existentes y predecir los resultados futuros**.

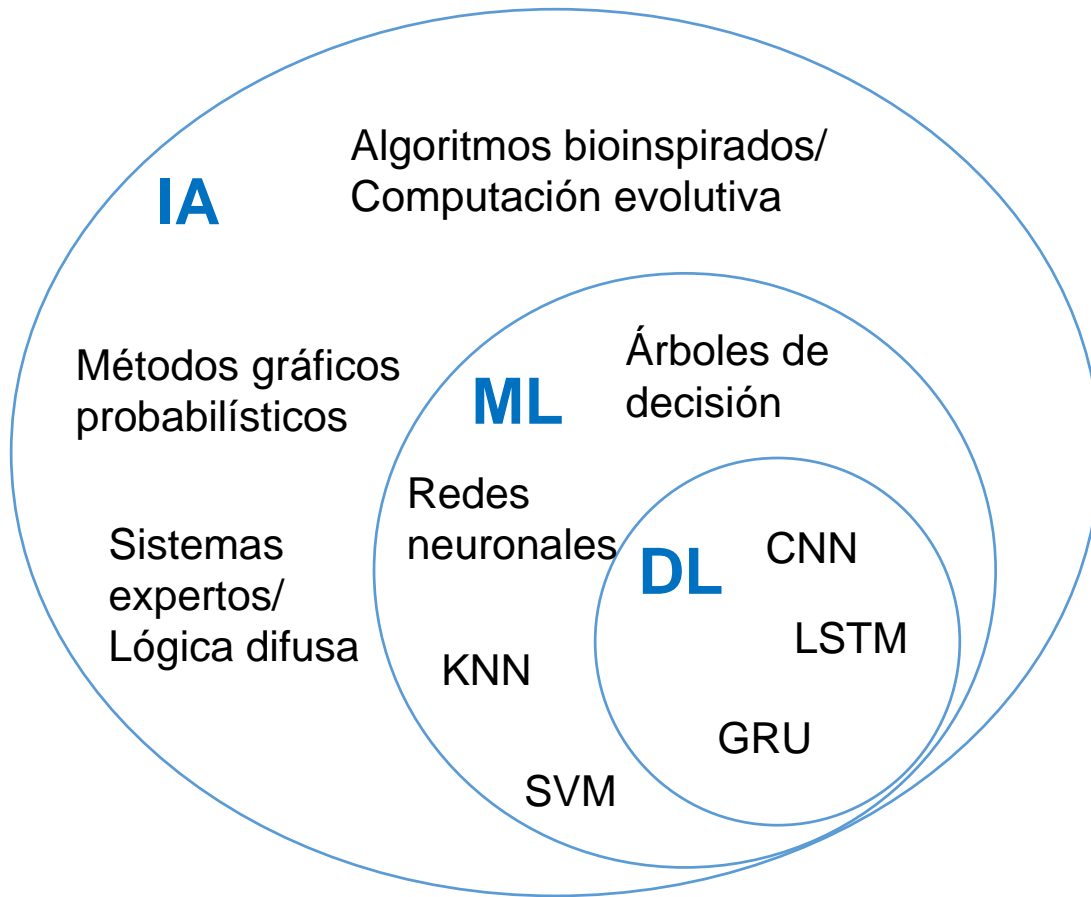
Por **ejemplo**, un programa informático es diseñado por un servicio de video para recomendar películas que podrían gustarles a los usuarios individuales. El algoritmo analiza las películas que los espectadores han visto ya y las películas que las personas con preferencias similares de visualización calificaron con buena puntuación. El objetivo es mejorar la satisfacción del cliente con el servicio de video.

IA vs Machine Learning vs Deep Learning

El aprendizaje automático (*Machine Learning* o ML) es un subconjunto de la inteligencia artificial y el aprendizaje profundo (*Deep Learning* o DL) es un subconjunto del aprendizaje automático.

Los modelos de **Machine Learning** no requieren necesariamente grandes sumas de datos, por lo que pueden resolver problemas menos complicados en un corto período de tiempo.

El **Deep learning** no se centra solo en un conjunto único de problemas, sino que brinda un amplio nivel intelectual e inteligencia para las computadoras. Las redes neuronales se adaptan y aprenden de grandes cantidades de datos. Un problema se resuelve de todas las formas posibles (ejemplos: en un juego de ajedrez se manipula al oponente a un escenario donde la computadora gana, analizar los ADN humanos para conocer los patrones que nos conducen a enfermedades terminales).



Tipos de Aprendizaje

Generalmente puede dividirse en tres categorías:

1. **Supervisado** (Supervised Learning): requieren etiquetas
2. **No supervisado** (Unsupervised Learning): no requieren etiquetas
3. **Por refuerzo** (Reinforced Learning)

1. Aprendizaje Supervisado: son los algoritmos de aprendizaje automático más utilizados para el análisis predictivo. Estos algoritmos dependen de conjuntos de datos que fueron procesados por los expertos humanos (por lo tanto, se usa la palabra “supervisión”). Los algoritmos luego aprenden cómo realizar las mismas tareas de procesamiento de forma independiente en los nuevos conjuntos de datos.

En particular, los métodos supervisados se utilizan para resolver problemas de regresión y clasificación.

IA vs Machine Learning vs Deep Learning

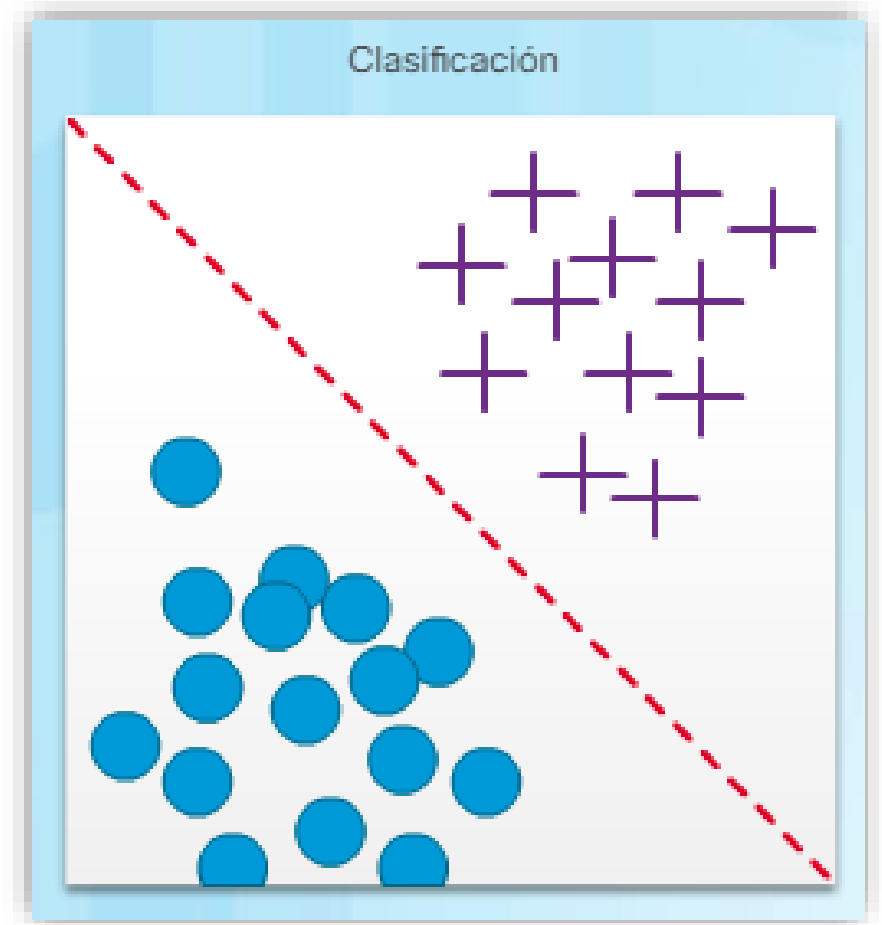
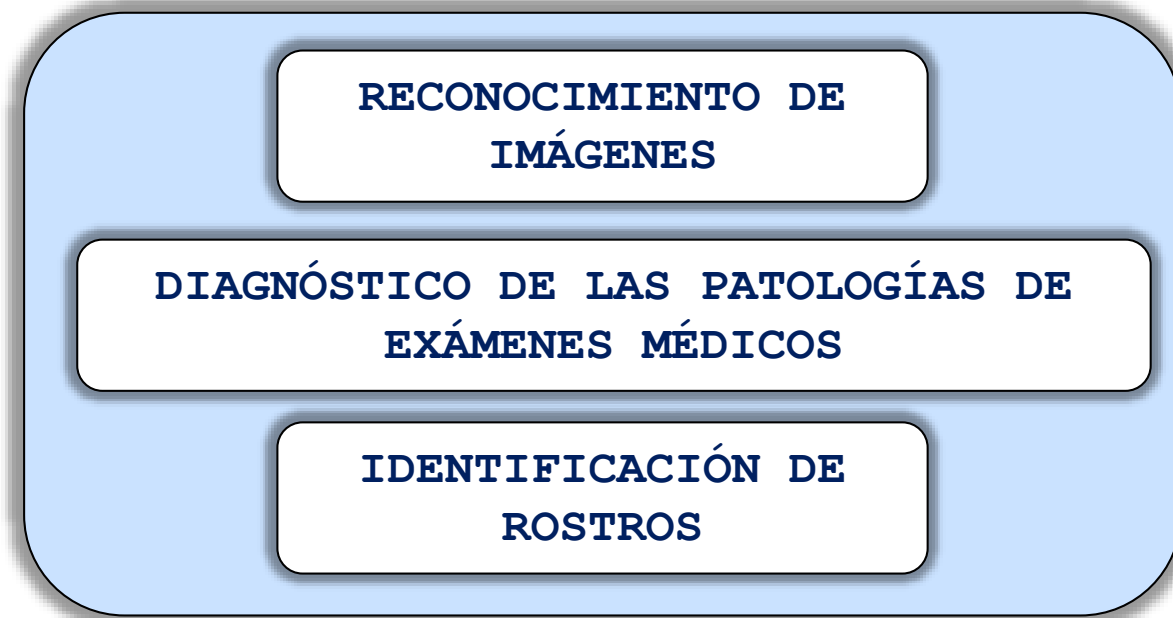
2. Aprendizaje No Supervisado: Los algoritmos de aprendizaje automático no supervisados no requieren expertos humanos de los que aprender, sino que descubren patrones en los datos de forma autónoma. Algunos ejemplos de problemas resueltos con métodos no supervisados son el agrupamiento y la asociación.

3. Aprendizaje por Refuerzo: por cada paso incorrecto el sistema es castigado y por cada paso correcto el sistema es recompensado.

Aprendizaje Supervisado

Problemas de clasificación

- Variable desconocida es discreta
- Grupos predefinidos



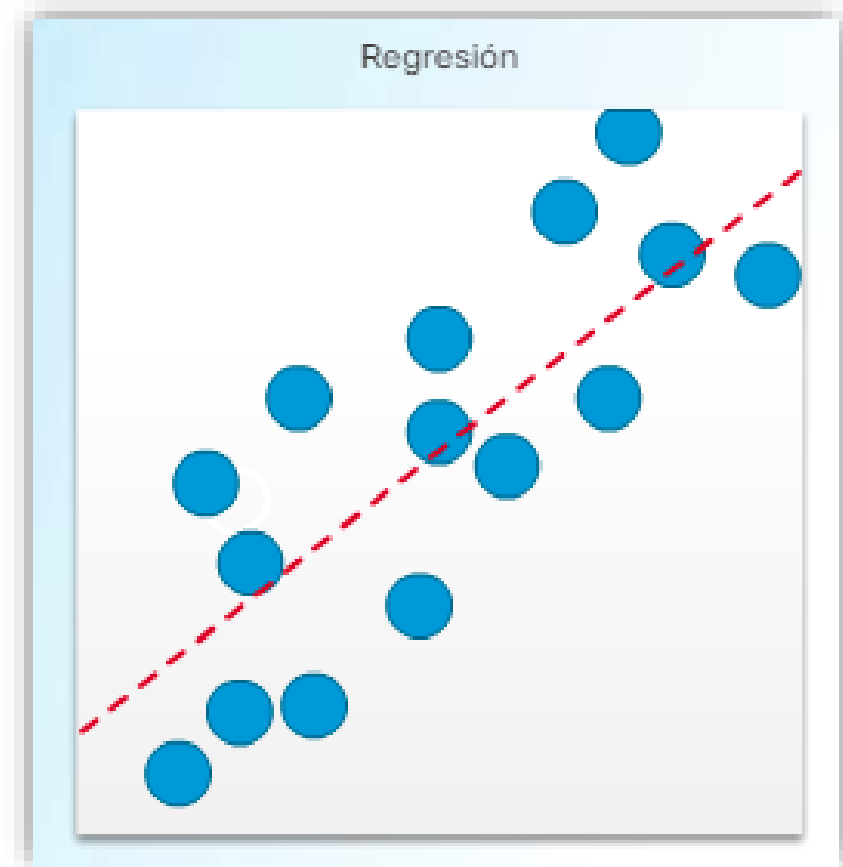
Aprendizaje Supervisado

Problemas de regresión: son el cálculo de las relaciones matemáticas entre una variable continua y una o más variables. La ejecución de la regresión significa encontrar la FUNCIÓN que interpola mejor los valores. Una función de regresión **permite estimar el valor de una variable dado el valor de otra, para los valores que no se han obtenido antes.**

POSICIÓN DE UN AUTOMÓVIL Y SU
VELOCIDAD

PREDICCIÓN DE LA TRAYECTORIA DE UN
TORNADO

ESTIMACIÓN DEL COSTO
DE UNA CASA



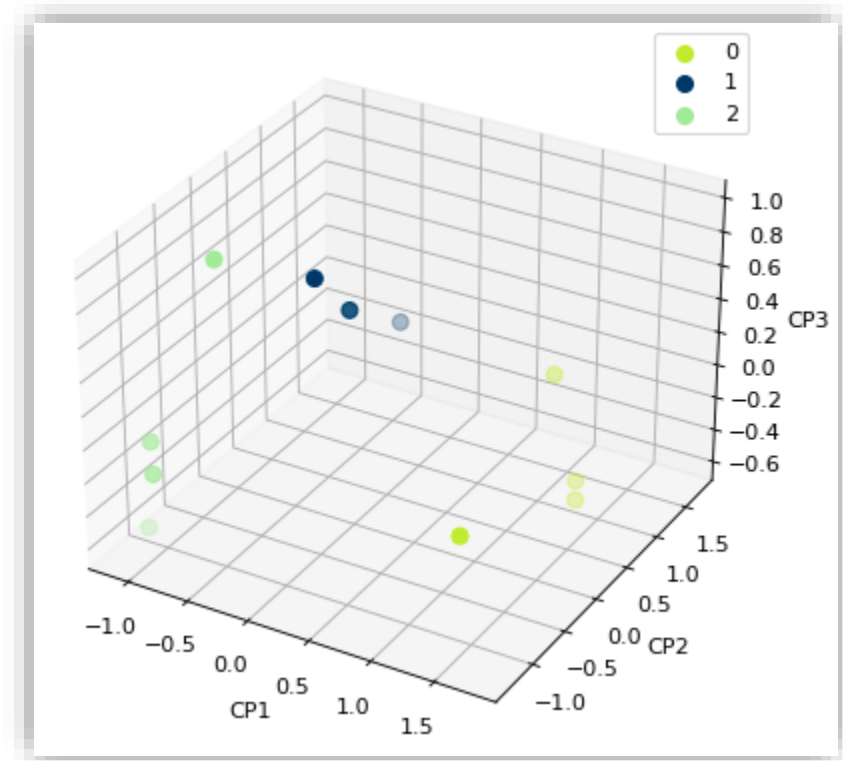
Aprendizaje No Supervisado

Métodos de agrupamiento: estos se pueden ver como la detección automática de grupos de ejemplos que tienen características similares, que pueden indicar posiblemente el hecho de que un miembro del grupo pertenece a una clase bien definida. Por ejemplo, los algoritmos de agrupamiento se utilizan para identificar grupos de usuarios basados en su historial de compras en línea, y luego envían avisos dirigidos a cada miembro.

DESCUBRIMIENTO DE CLIENTES POTENCIALES
SEGÚN EL HISTORIAL DE COMPRAS

SEGMENTACIÓN DE IMÁGENES EN REGIONES CON
CARACTERÍSTICAS SIMILARES

ANÁLISIS DE DATOS ESPACIALES: AGRUPAR
UBICACIONES GEORGRÁFICAS PARA PLANIFICACIÓN
URBANA



Aprendizaje No Supervisado

- **Problemas de asociación**

- Detección de grupos de elementos que se observan con frecuencia en conjunto
- Patrones o secuencias de elementos con mayor frecuencia

SUGERENCIA DE PRODUCTOS QUE SE COMPRAN CON MAYOR FRECUENCIA CON UN PRODUCTO ESPECÍFICO

ANÁLISIS DE NAVEGACIÓN EN SITIOS DE WEB PARA COLOCACIÓN DE CONTENIDO

DETECCIÓN DE PATRONES INUSUALES EN TRANSACCIONES FINANCIERAS



Análisis multivariante

El **análisis multivariante** es el estudio estadístico de varias variables medidas en elementos (observaciones) de una población. Proporciona **métodos objetivos (técnicas estadísticas)** para conocer cuantas variables indicadoras (a veces se denominan factores) son necesarias para describir una realidad compleja y determinar su estructura. Es importante entender la estructura de dependencia entre variables, ya que las relaciones entre variables son las que permiten resumirlas en variables indicadoras, encontrar grupos no aparentes por las variables individuales o clasificar casos complejos.

Objetivos específicos:

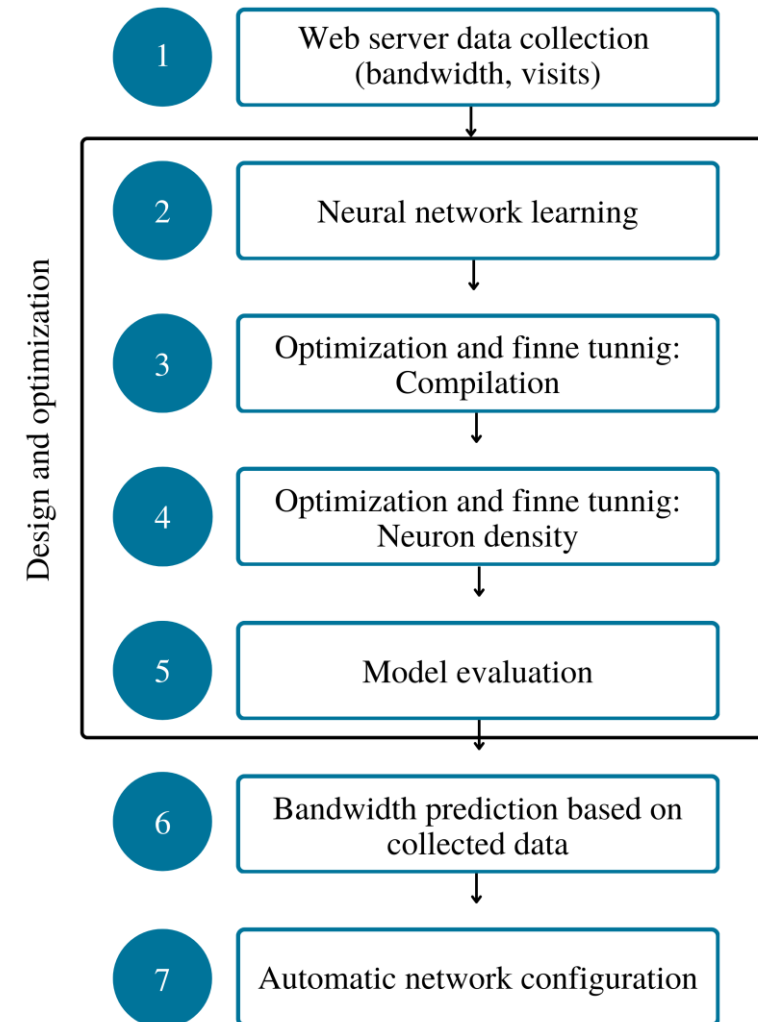
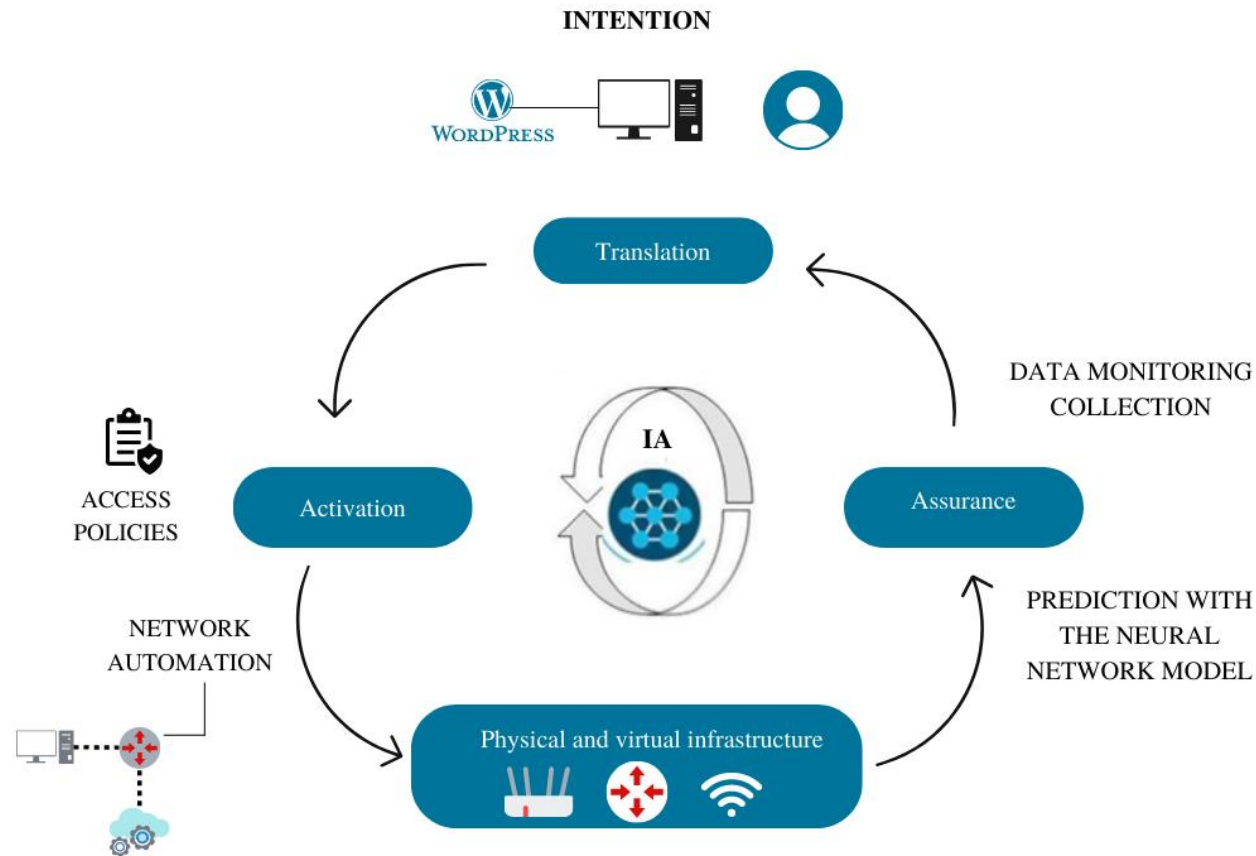
1. **Resumir/reducir** el conjunto de **variables (dimensiones)** en pocas nuevas variables construidas como **transformaciones** de las originales, con la mínima pérdida de información. Al tener menos variables (factores/indicadores) estas se pueden representar gráficamente y comparar. Los indicadores pueden llegar a interpretarse y mejorar el conocimiento de la realidad.
2. **Descubrir grupos** en los datos (si existieran). Se descubren grupos sin conocer previamente la cantidad de grupos.
3. **Clasificar nuevas observaciones** en grupos definidos.
4. **Relacionar variables**

Aplicaciones de ML

- **Marketing:** Ayudan a acertar en la toma de decisiones. La demanda de una tienda en concreto, permitiendo optimizar el inventario y el personal.
- Marketing Personalizado / **Sistemas de Recomendación:** Descubrir mercados, analizar comportamientos, predecir ventas.
- **Ecommerce:** la inteligencia artificial impacta al comercio online (y el offline): mejores recomendaciones de compra y en la optimización de la logística (planificación de rutas de reparto).
- **Detección de Fraude** (transacciones ilícitas, lavado de activos)
- **Reconocimiento de Imágenes:** Digitalización de textos, Reconocimiento de Manuscritos (handwriting)
- **Reconocimiento de Rostros:** Seguridad con cámaras para mitigar la delincuencia
- **Bolsa de valores:** Predecir tendencias
- **Comercio Financiero:** Análisis financiero. Predecir stock para tomar decisiones de compra, venta
- **Atención al Cliente:** con **Procesamiento de Lenguaje Natural** (NLP) y **chatbots**
- **Predicción de Enfermedades**
- **Cybersecurity:** Detección de Malware, Spam
- **Evaluación de Riesgos:** Clasificación de usuarios - nivel de riesgo (en seguros, préstamos).
- **Tendencias en Política**
- **Redes basadas en intención (IBN):** reconocimiento de intenciones y aseguramiento del servicio

Aprendizaje Automático para Networking

Redes basadas en Intención (IBN) Tendencias empresariales y tecnológicas para la red de TI



Técnicas / Algoritmos de Aprendizaje Automático Moderno

- **SUPERVISADO:**
 - **CLASIFICACIÓN:** KNN, Naive Bayes, SVM, Árboles de decisión, Regresión Logística
 - **REGRESIÓN:** Regresión Lineal, Regresión Polinomial, Regresión Lidge/Lasso
- **NO SUPERVISADO:**
 - **CLUSTERING:** KMeans, DBSCAN, Aglomerativo, Fuzzy CMeans, Mean-Shift
 - **PATTERN SEARCH:** A priori, Euclat, FP-Growth
 - **REDUCCIÓN DE DIMENSIONALIDAD:** PCA, LDA, SVD, MF, LSA, t-SNE
- **POR REFUERZO:**
 - **BÚSQUEDA Y OPTIMIZACIÓN:** Algoritmos genéticos, Deep Q-Network (DQN), ASC, SARSA, Q-learning
- **REDES NEURONALES Y DEEP LEARNING:** Perceptrón Multicapa (MLP), Convolucionales (CNN), Recurrentes (LSM, LSTM, GRU), Adversas/Adversarias/Antagónicas Generativas (GAN), Autoencoders
- **MÉTODOS DE CONJUNTO O ENSAMBLE:**
 - **APILAMIENTO** (Stacking)
 - **EMPAQUETAMIENTO** (Bagging): Bosques Aleatorios (Random Forest)
 - **IMPULSO** (Boosting): AdaBoost, CatBoost, LightGBM, XGBoost
- **SISTEMAS DE RECOMENDACIÓN:** de filtrado colaborativo, de filtrado basado en contenidos, híbridos.

Técnicas / Algoritmos de Aprendizaje Automático



Referencias

- Bishop, C. M. (2006). Pattern recognition and machine learning. springer.
- El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In machine learning in radiation oncology (pp. 3-11). Springer, Cham.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2012). Pattern classification. John Wiley & Sons.
- P. Joshi. (2017). Artificial intelligence with python. Packt Publishing Ltd.
- Brownlee. J. (2019, Agosto 12). A Tour of Machine Learning Algorithms [Machine Learning Mastery]. Recuperado de <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>
- S. Raschka, V. Mirjalili. (2007). Python Machine Learning, Packt Publishing Ltd.