Introducción al *Machine Learning*

Breve descripción:

Este componente formativo aborda aspectos generales y claves sobre el "Machine Learning", la forma en que las máquinas de aprendizaje son un campo de la inteligencia artificial que se enfoca, principalmente, en el uso de datos y algoritmos. Con el estudio responsable de estos contenidos, el aprendiz se afianzará en el manejo de máquinas y uso de algoritmos de aprendizaje supervisado, no supervisado y semisupervisado.

Mayo 2023

Tabla de contenido

Introducción

Le damos la bienvenida al componente formativo denominado “Introducción al “Machine Learning””, el cual hace parte del programa de formación complementaria “Algoritmo de agrupamiento no supervisado K-means con Python”.

Para comenzar, preste atención al video que se muestra enseguida:

Video 1. Introducción al “Machine Learning”

Enlace de reproducción del video

Síntesis del video: Introducción al “Machine Learning”
Los datos siempre han gobernado el mundo, y hoy en día mucho más con el avance de las tecnologías computacionales, proveedores de aplicaciones en la nube y

grandes bases de datos o “Big data”. El aprendizaje supervisado, el algoritmo de “Machine Learning”, busca un conjunto de reglas que le permiten deducir características generales de los elementos del grupo, con el objetivo de aplicar una misma etiqueta a elementos similares.

El aprendizaje automático no es nuevo, desde casi un siglo estamos interesados en sacarle provecho a los datos. Podemos decir que los padres de la inteligencia artificial no son de este siglo. En 1950, Alan Turing, matemático británico, publica el artículo Máquinas de computar e inteligencia.

En 1952, Arthur Lee Samuel, pionero en inteligencia artificial y videojuegos, usó el término aprendizaje automático, creó el primer juego “checkers” o damas, basado en aprendizaje automático, el cual mejoraba después de cada partida.

Frank Rosenblatt, en 1957, desarrolla el perceptrón Mark 1, máquina capaz de aprender mediante un sistema de red nerviosa que simula procesos del cerebro humano.

En 1967, se escribió el algoritmo “nearest neighbor”, con el cual nacieron los algoritmos de reconocimiento de patrones.

Actualmente, grandes empresas presentan sus plataformas de aprendizaje automático. Uno de los pilares de la industria 4.0 son la inteligencia artificial y las máquinas de aprendizaje.

1. “Machine Learning”

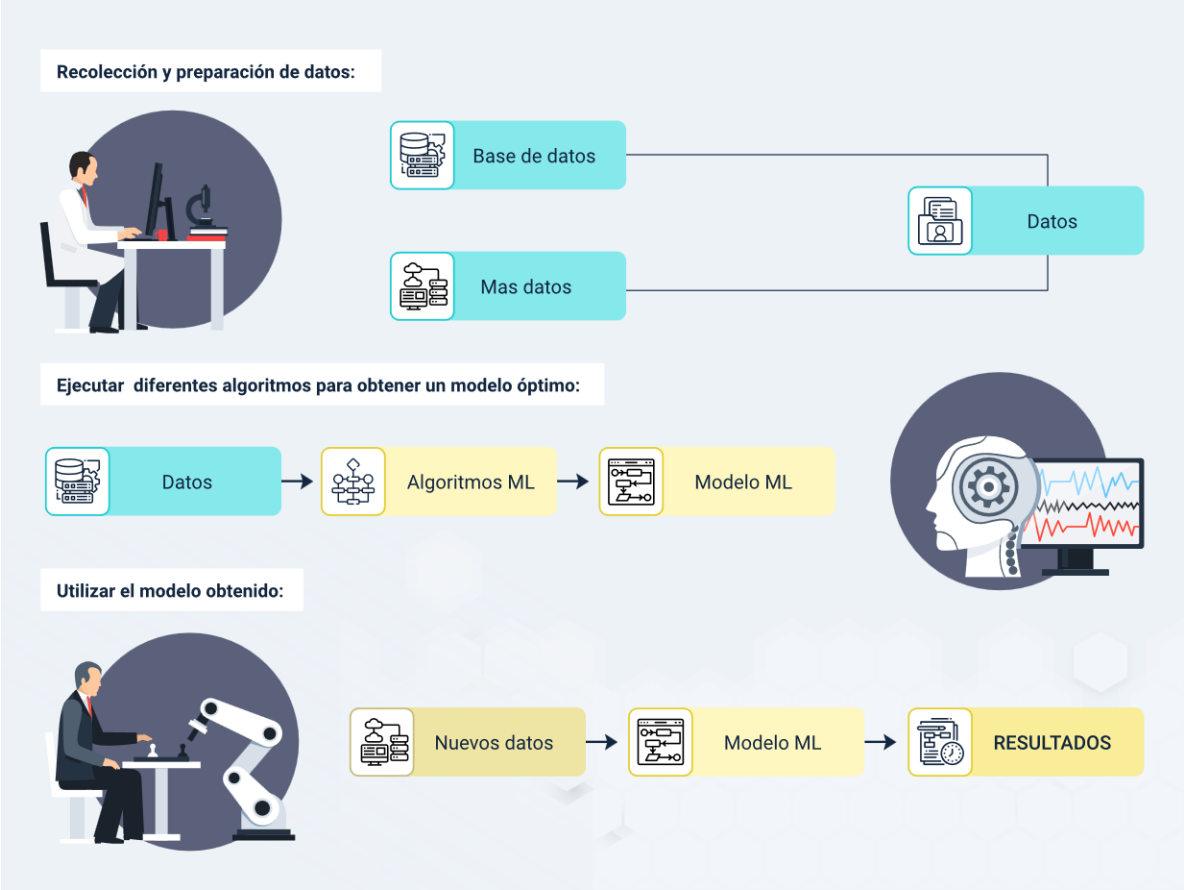
Las máquinas de aprendizaje son un campo de la inteligencia artificial que se enfoca, principalmente, en el uso de datos y algoritmos: con ellos, imitan la forma en la que las personas aprenden y, de manera progresiva, van mejorando su precisión.

Sobre la implementación de sistemas de aprendizaje automático, es posible afirmar que:

- **Escritura de programas.** Las personas que los implementan, no escriben programas por sí mismos.
- **Recopilación de datos.** Se recopilan datos y, luego, se da la orden al computador para buscar un programa que calcule una salida para cada valor de entrada.
- **Generación de modelos.** Como resultado de esto, se genera un modelo, que no es más que un archivo que se ha entrenado para reconocer determinados tipos de patrones.
- **Entrenamiento del modelo.** Es posible entrenar un modelo con un conjunto de datos y proporcionar un algoritmo que puede usar para averiguar y obtener información de esos datos.
- **Alimentación del modelo.** Una vez el modelo ya está entrenado, se puede usar para ser alimentado con los datos que no ha visto antes y realizar predicciones sobre estos.

El esquema que se muestra a continuación, expone el modelo “Machine Learning”:

Figura 1. Modelo “Machine Learning”



Esquema que presenta el modelo “Machine Learning”, donde se puede observar:

- Recolección y preparación de datos.
- Ejecutar diferentes algoritmos para obtener un modelo óptimo.
- Utilizar el modelo obtenido.

Las máquinas de aprendizaje se dividen en tres diferentes tipos de aprendizaje:

- Aprendizaje no supervisado.
- Aprendizaje supervisado.

- Aprendizaje semisupervisado.

Algunas generalidades importantes de tener en cuenta, sobre el "machine learning", son:

- **Aprendizaje supervisado.** Si se le enseña a un algoritmo el resultado que se desea obtener para un determinado valor, basado en la relación existente entre muchas observaciones y entre variables de entrada y salida, a esto se le denomina generalizar el conocimiento y, por tanto, se estaría hablando de aprendizaje supervisado.

- **Otros algoritmos.** También podría darse entre otros algoritmos basados en aprendizaje supervisado, donde existe la regresión lineal y logística, máquinas de vectores de soporte.

- **Aprendizaje no supervisado.** Si se busca conocimiento o patrones entre un montón de datos, de los cuales no se conocen relaciones entre sus variables, no hay datos de referencia, entonces se está hablando de aprendizaje no supervisado.

- **Algoritmos del aprendizaje no supervisado.** Entre algoritmos de aprendizaje no supervisado, existen el "clustering" o agrupamiento, "K-means" o K-medias y reglas de asociación.

- **Modelos predictivos.** En general, los modelos predictivos se aprenden de manera supervisada, mientras que los modelos descriptivos son producidos por técnicas de aprendizaje no supervisado.

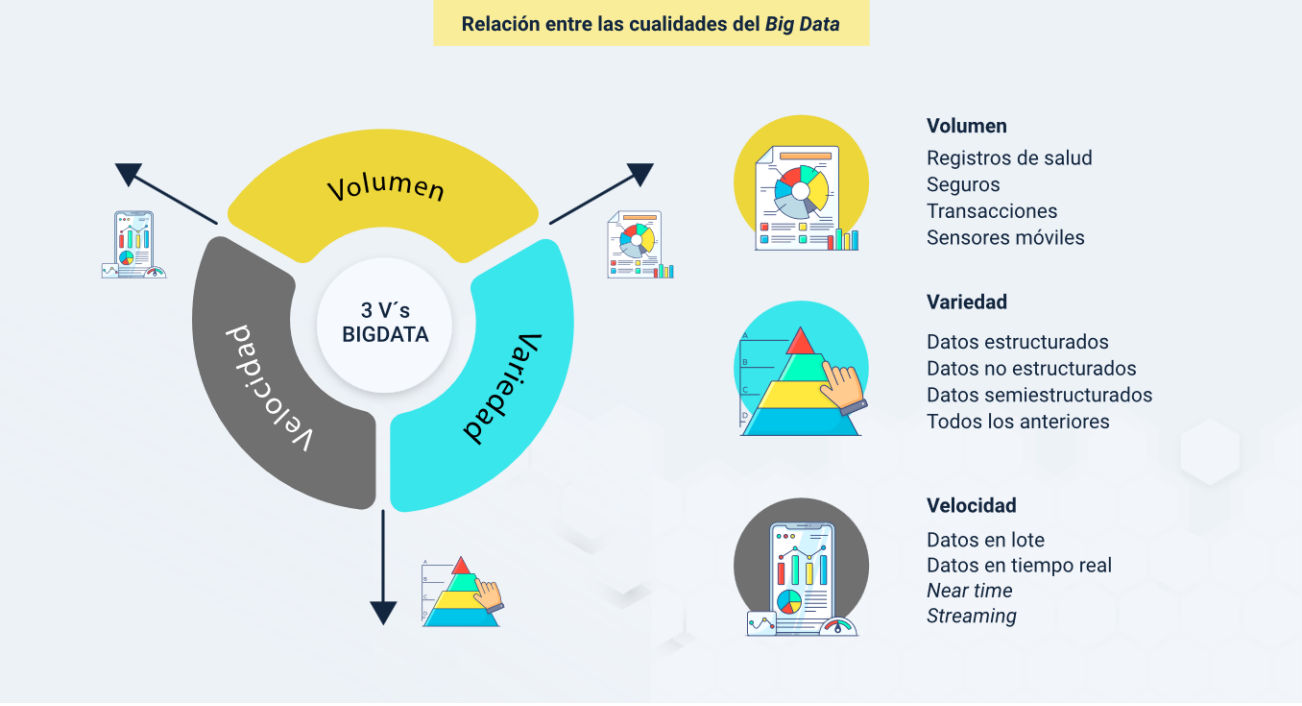
1.1. Conjuntos de datos

Los datos son el insumo más valioso cuando se habla de ciencia de datos y un término muy usado en la actualidad es el "Big Data" que, como su nombre lo indica, hace referencia a conjuntos de datos muy grandes.

Se han definido tres cualidades para "Big Data": volumen, velocidad y variedad. Entonces es posible ver que este implica grandes conjuntos de datos (volumen), de diversos tipos (variedad) y que se generan muy rápidamente (velocidad).

La siguiente infografía muestra la relación entre cada concepto, preste atención:

Figura 2. Relación de las cualidades del "Big Data".



Se presenta la relación entre las cualidades del "Big Data" de la siguiente manera:

- En primer lugar, se aborda el aspecto del volumen mediante el registro de diversas fuentes como información de salud, transacciones de seguros, datos de sensores móviles y otros.
- En segundo lugar, se explora la cualidad de la variedad, incluyendo datos tanto estructurados como no estructurados, que abarcan una amplia gama de formatos y tipos de información.
- Por último, se analiza la importancia de la velocidad en el manejo de los datos, desde el procesamiento en lotes hasta la capacidad de trabajar con datos en tiempo real, "near time" y mediante "streaming".

A continuación, se mencionan eventos sobre volumen, velocidad y variedad:

- **Volumen y velocidad.** Como ejemplo de volumen, se puede observar que YouTube sube más de quinientas (500) horas de videos por minuto.
- **Volumen.** En Twitter, se publican por día 500 millones de tuits; y si se mira en otros ámbitos laborales, se podría tomar el área de la salud, donde se cuenta con millones de historias clínicas de pacientes en los centros médicos y hospitales.
- **Velocidad.** La velocidad con que se suben videos a YouTube o la velocidad a la que un tuit se expande en la red es enorme; el solo hecho de viajar con aplicaciones como Waze proporciona, en tiempo real, la ubicación a millones y millones de usuarios.
- **Variedad.** Los datos son de muchas variedades, pueden ser no estructurados, como los videos y el audio, o muy estructurados, como los

tamaños de los videos subidos a YouTube, las distancias recorridas por transportadores obtenidas por GPS, o datos de pacientes en centros médicos.

- **Volumen, velocidad y variedad.** Lo anterior demuestra que existen montañas de información, que está ahí lista para ser explorada y puede ser usada para contestar preguntas, realizar predicciones con ella y aportar en la toma de decisiones en las empresas, usando estadística, matemáticas, computación, transformación, visualización de datos.

Habiendo hablado de ciencia de datos, se debe conectar con el concepto de datos en informática, los cuales son representaciones simbólicas (vale decir: numéricas, alfabéticas, algorítmicas, etc.) de un determinado atributo o variable cualitativa o cuantitativa, o sea: la descripción codificada de un hecho empírico, un suceso, una entidad.

En general, existe un conjunto de valores de variables cualitativas o cuantitativas, los cuales se pueden detallar a continuación:

a. **Cualitativas.** Brindan información sobre las cualidades, tales como:

   - País de nacimiento.
   - Ciudad de nacimiento.
   - Sexo.

   Estas variables se describen con textos y no necesariamente los datos deben estar ordenados. Se clasifican en ordinales, cuando representan

algún valor ordenado según una escala establecida. Y en cuasi cuantitativa o nominales si no representan un orden.

b. **Cuantitativas.** Están relacionadas con cantidades, se representan con números y se miden en una escala ordenada o continua, tales como:

    - Altura.
    - Peso.
    - Presión arterial.

    Pueden ser discretas, si se representan por números enteros, o continuas, si se representan con números reales.

En relación con la "Big Data", se destacan algunos aspectos como:

- En la actualidad existen millones de datos que, quizá, no están tan ordenados como se debería.
- Tales datos podrían ser, por ejemplo, el censo de población o la gran cantidad de historias clínicas; y es en este aspecto donde es importante la exploración de la información y la respuesta a preguntas formuladas sobre esa información.
- El manejo de datos puede ser mucho más complejo de lo que se imagina, ya que va desde usar los datos inmersos en las imágenes para reconocer las mismas, hasta llegar al reconocimiento de dígitos o caracteres.

- Por ejemplo, esta situación se da en el caso de que se quiera obtener información de historias clínicas no automatizadas, reconocimiento de patrones en la información, reconocimiento automático de voz, etc.
- Los datos no estructurados tienen estructura interna, pero no están predefinidos por modelos de datos. Pueden ser generados por personas o una máquina, en formato textual o no textual.

A través de las siguientes figuras, conozca un ejemplo de identificación de tipos de datos en el Python. Se trata de un Dataset cuyos datos contienen cuatro características (longitud y anchura de sépalos y pétalos) de 50 muestras de tres especies de iris (iris setosa, iris virginica e iris versicolor).

Figura 3. Extracto de datos de "dataset"

En Python, usando Jupyter Lab, se pueden ver algunos tipos de variables usando el siguiente código:

Figura 4. Variables

```
[36]: import pandas as pd

[38]: import os

[41]: cache_path = "data/iris.csv"

[44]: iris = pd.read_csv("https://raw.githubusercontent.com/toneloy/data/master/iris.csv")

[46]: print(iris.dtypes)

sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
species          object
dtype: object
```

Se observan, después de imprimir los tipos de datos de iris, las variables sepal_length (longitud del sépalo), sepal_width (ancho de sépalo), petal_length (longitud de pétalo), petal_width (ancho de pétalo), que son de tipo float64 y, por tanto, se habla de variables cuantitativas.

La variable "Species" (especie) es de tipo "object", que puede ser de cualquier tipo, por ejemplo, texto; en este caso, representa una variable cualitativa.

1.2. "Clustering"

En el aprendizaje supervisado el objetivo es mapear una entrada con una salida, con valores correctos suministrados por un supervisor. Por su parte, en el aprendizaje no supervisado no existe ese supervisor y solo existen datos de entrada; aquí el objetivo es describir similitudes en los datos de entrada y así observar patrones que aparecen con más frecuencia que otros, entonces la idea es investigar qué es lo que sucede con estos datos y qué no.

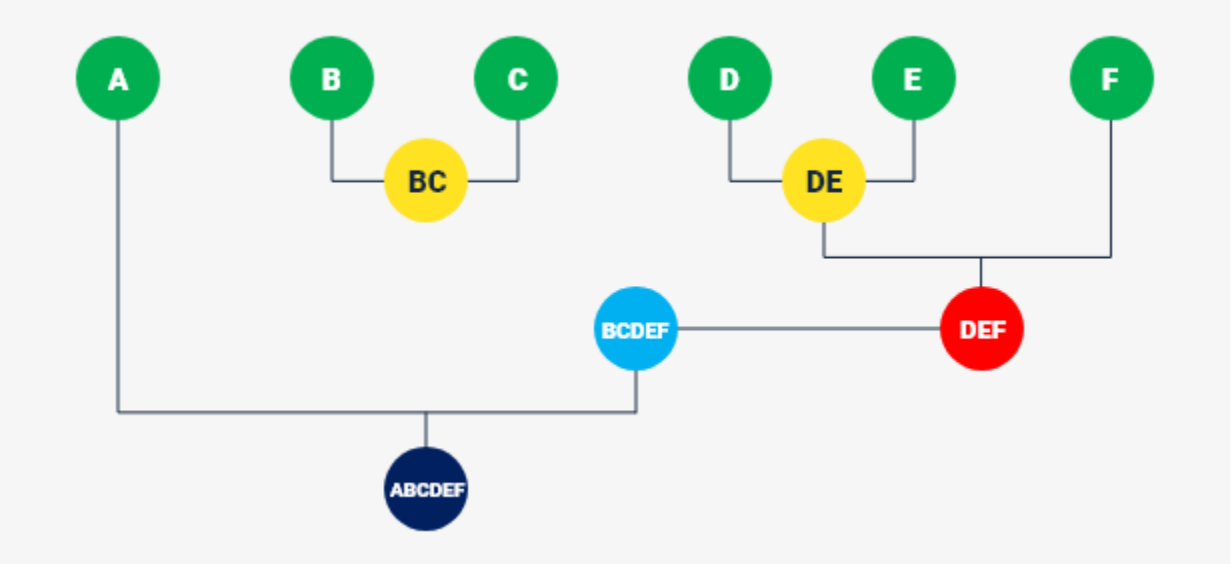
En términos estadísticos, lo que se hace es estudiar la densidad de los datos, agrupándolos en clústeres; todos los patrones, asociaciones, relaciones; y los clústeres son extraídos de los mismos datos.

Existen varias técnicas para realizar “clustering” basadas en agrupamiento con “K-Means”. A continuación, se presenta un “clustering” jerárquico:

Agrupamiento jerárquico

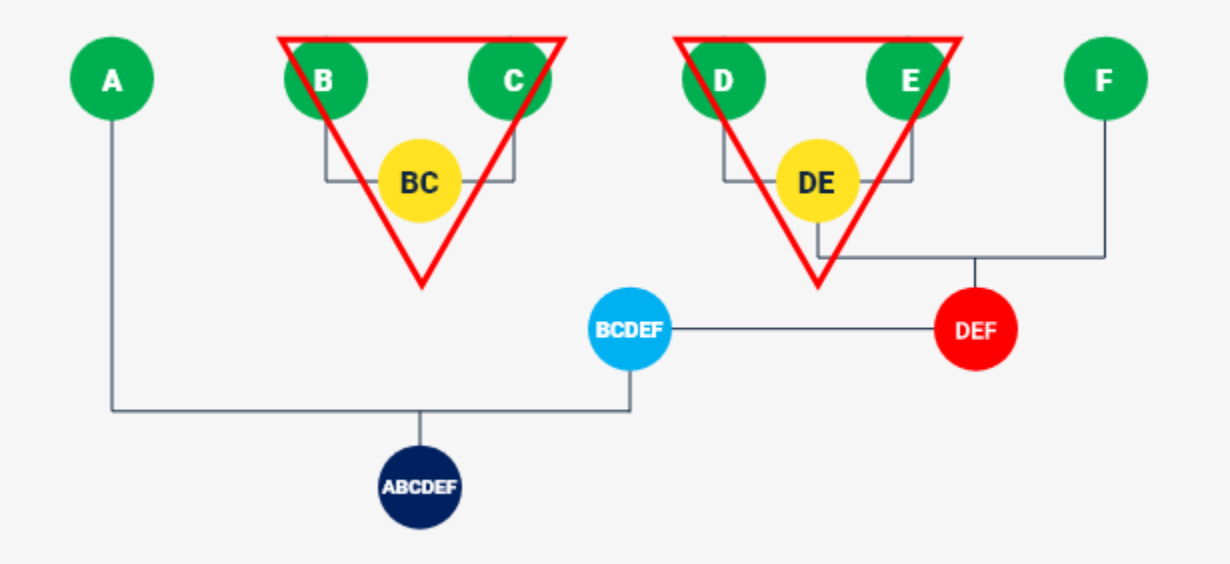
Es un método para agrupar datos basándose en la distancia entre ellos, buscando que los datos que queden en un clúster sean lo más parecidos posibles. Este tipo de agrupaciones ayuda a las organizaciones, por ejemplo, a que sus precios, bienes, servicios y cualquier aspecto de sus negocios estén bien dirigidos a sus clientes.

Esta jerarquía muestra los datos similares a una estructura de árbol llamada dendrograma. A su vez, existen dos formas de agrupar los datos: aglomerante y divisivo.



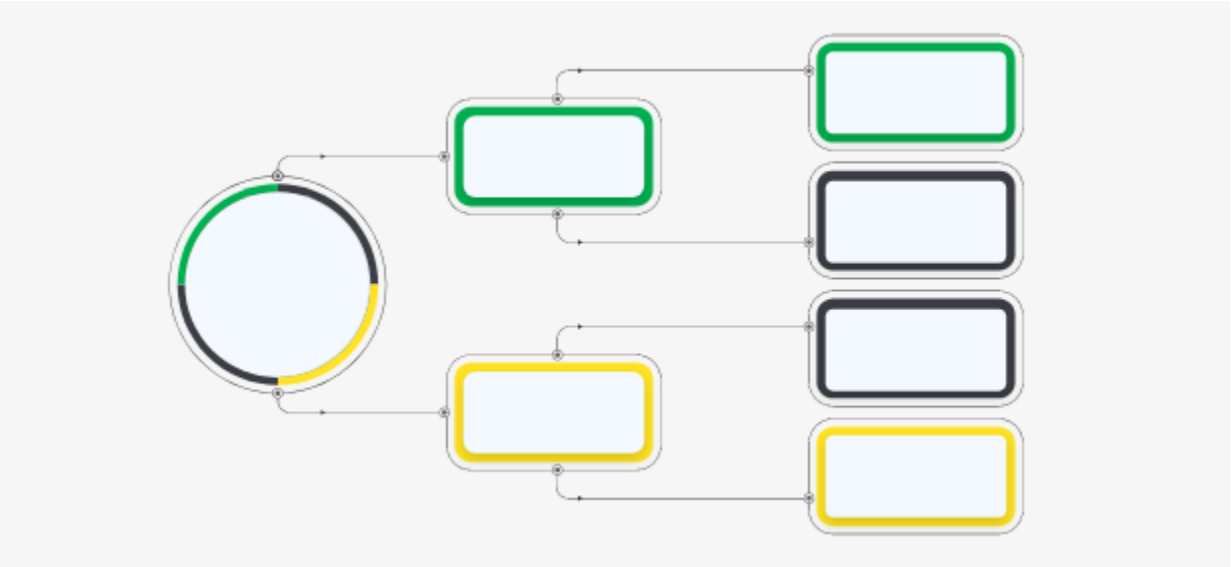
Enfoque ascendente

En el caso de aglomerante, es un enfoque ascendente, cada elemento se piensa primero como un clúster de un solo elemento. Los dos clústeres más parecidos se unen en un nuevo clúster más grande en cada fase del método (nodos); este método se repite hasta que todos los puntos pertenezcan a un único clúster grande (raíz).



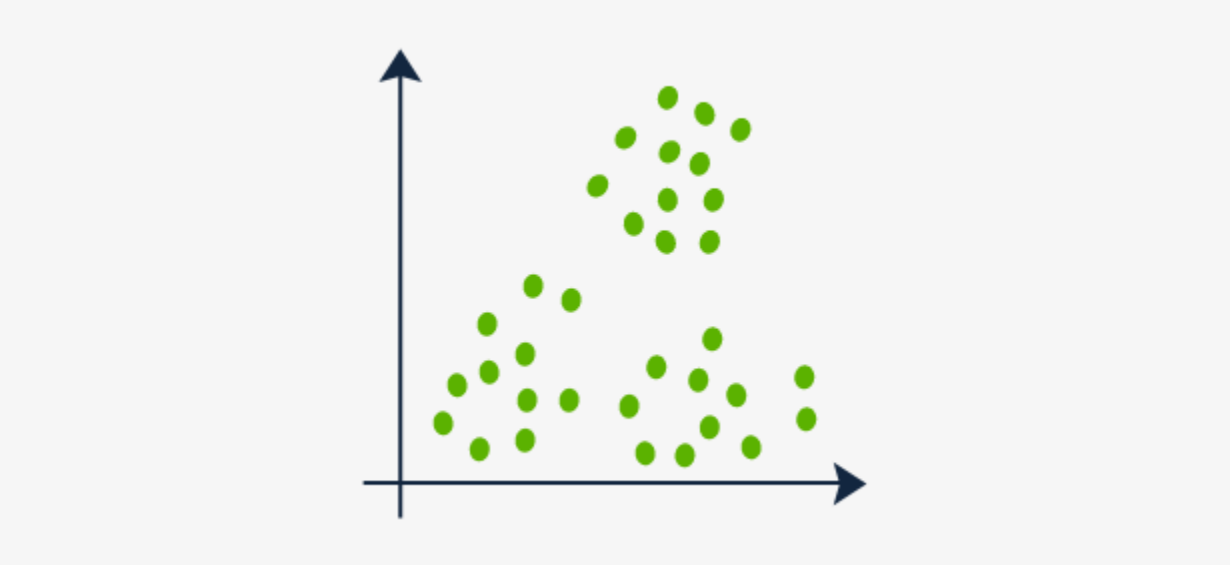
“Clustering” divisivo

En el “clustering” divisivo, funciona en forma descendente: comienza en la raíz, donde todos los elementos están en un solo clúster y luego separa los más diversos en dos en cada fase de iteración. Se itera el procedimiento hasta que todos los elementos estén en su grupo.



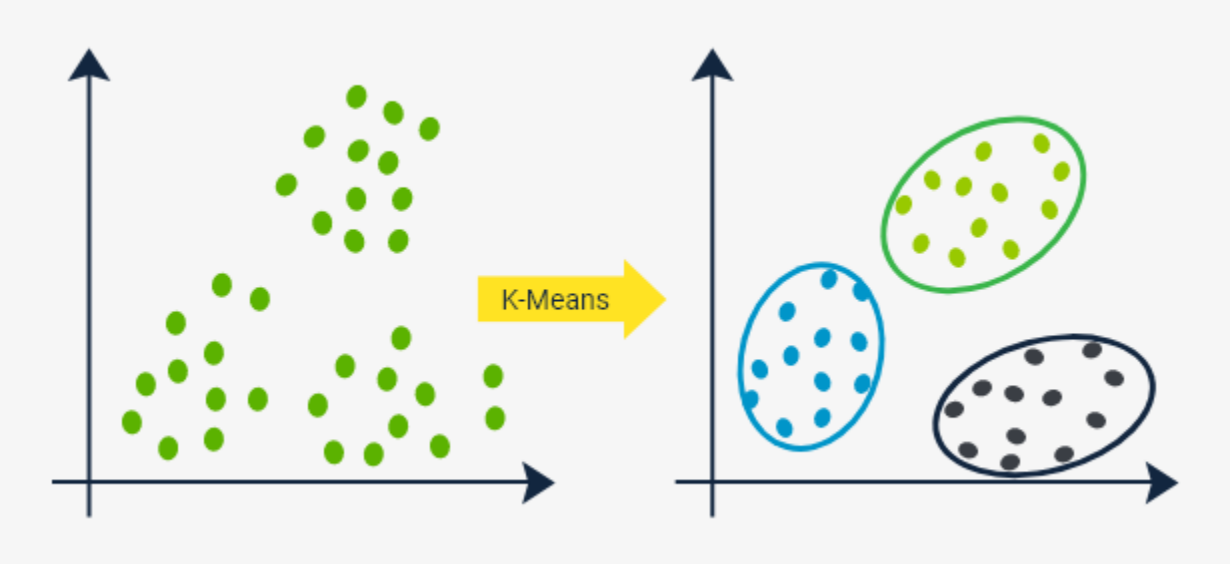
“K-means”

Es uno de los algoritmos de máquinas de aprendizaje no supervisado más utilizados en ciencia de datos, el objetivo es, igual que en el caso anterior, agrupar datos con características similares e identificar patrones que muchas veces no se pueden observar a simple vista.



Antes y después del “K-means”

En el caso de “K-means” o K-medias, el algoritmo elige aleatoriamente una cantidad k de centroides iniciales que marcan el centro de cada clúster. Cada punto se ubica con su centroide más cercano, usando cualquier medida de distancia, tal como la distancia euclídea.



Ventajas

- Es un algoritmo veloz y eficiente, en términos de costo computacional, para segmentar datos.
- Es sencillo de implementar y aplicar.
- Produce clústeres más definidos que el clustering jerárquico.
- Puede manejar grandes datos.


Unlabelled Data
Labellerd Cñusters
K-Means
= Centroid

2. Aplicaciones de “clustering” de datos

La clusterización usando “K-means” tiene un espectro bastante amplio de aplicaciones muy útiles en todas las ramas de la ciencia, tecnología, salud, transporte, astronomía, industria, educación, etc.

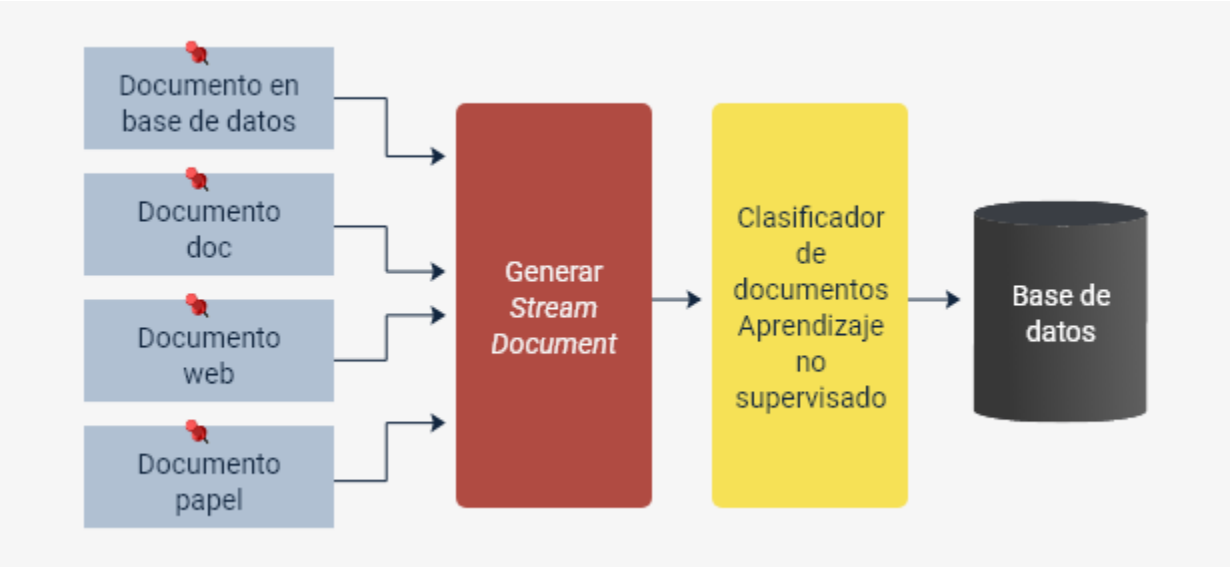
Clasificación de documentos

Se pueden usar aplicaciones con “K-means” para clasificar documentos en múltiples categorías, basadas en temas del texto o documentos, dependiendo del contenido de las cadenas.

A continuación, se presenta la clasificación:

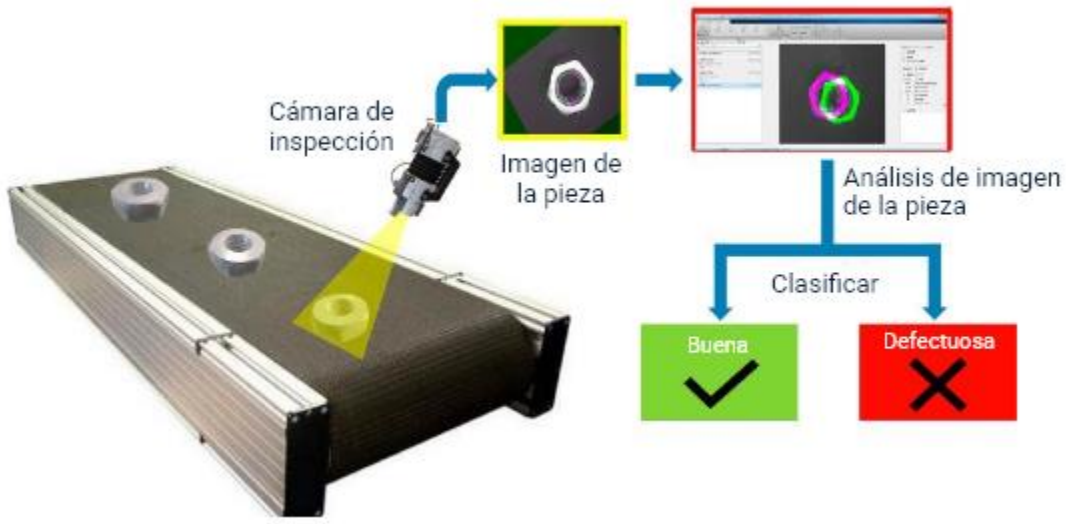
Clasificación de textos

Tiene una variedad de aplicaciones; por ejemplo, en salud, sería bastante importante obtener textos útiles de historias clínicas y clasificar pacientes por diagnósticos, procedimientos aplicados, medicamentos formulados, etc. Clasificación de email como spam o no spam, etiquetar automáticamente las consultas de los consumidores, detectar estados de ánimo de usuarios desde un tuit.



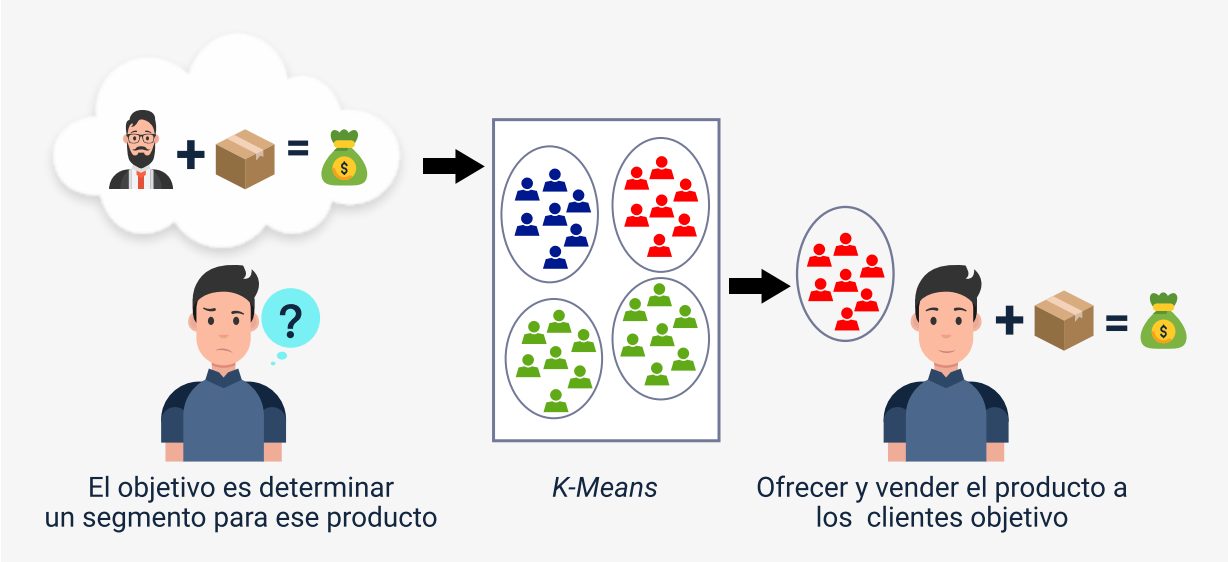
Detección de defectos de fábrica

Una aplicación de clusterización con "K-means" es la detección de defectos de fábrica, reconociendo patrones. La figura muestra la captura de imágenes, procesamiento de imágenes, modelado y clasificación del objeto como bueno o defectuoso.



Segmentación de clientes

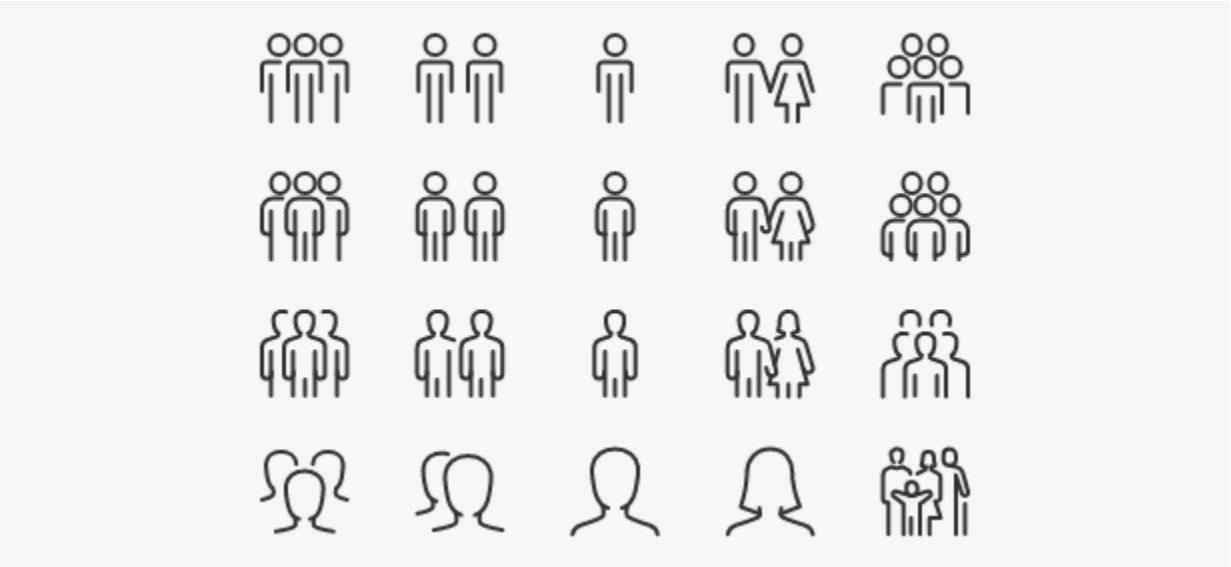
La segmentación de clientes es una de las aplicaciones más importantes del aprendizaje no supervisado, las empresas pueden identificar varios segmentos de clientes y luego dirigirse a grupos de usuarios potenciales. Lo que se hace es dividir una base de datos de clientes en grupos de personas que tienen semejanzas en género, edad, hábitos de consumo, intereses personales.



Beneficios de la segmentación

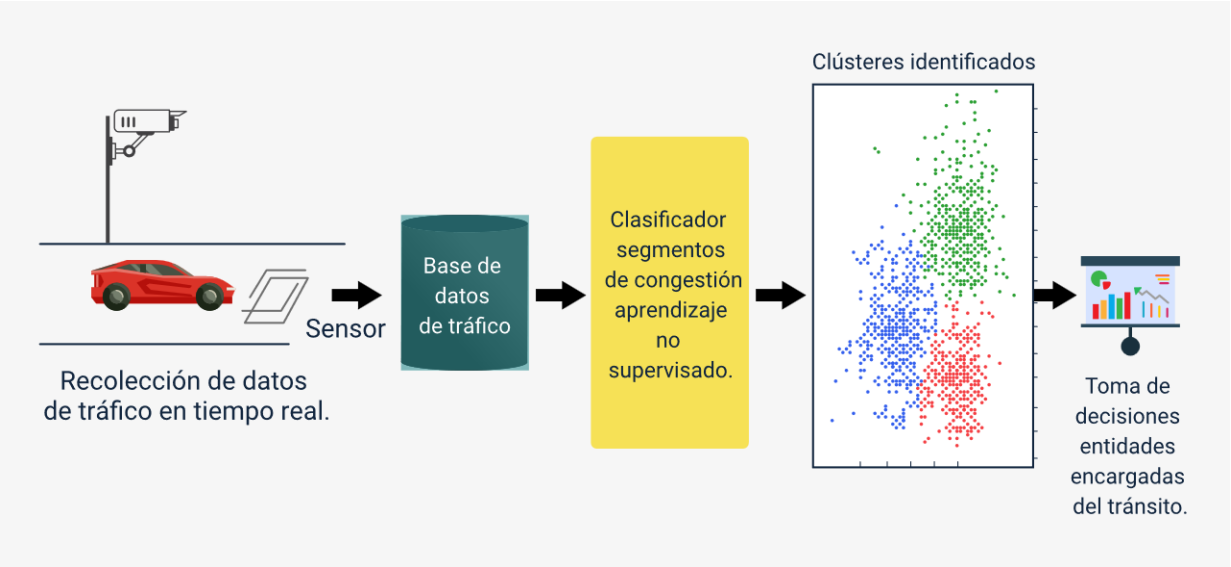
El área comercial de la empresa formula esfuerzos de mercadeo diferentes y se adapta a las necesidades de cada cliente de acuerdo con el clúster en el que queda cada persona.

La empresa puede identificar, también, preferencias de clientes y conocer necesidades nuevas de mercado y así tomar decisiones y nuevas estrategias de mercadeo en forma efectiva.



Congestión de tráfico

El número de vehículos en las ciudades se incrementa de forma abrumadora, originando con esto congestión vehicular y, por tanto, aumentando la contaminación del medio ambiente, pérdidas de tiempo y costos en combustible, entonces es esencial monitorear el tráfico en las autopistas, identificando las más congestionadas, identificando patrones de congestión y creando una clasificación en segmentos de tráfico; esto ayuda a entidades estatales de tránsito a optimizar el tránsito en dichas zonas.

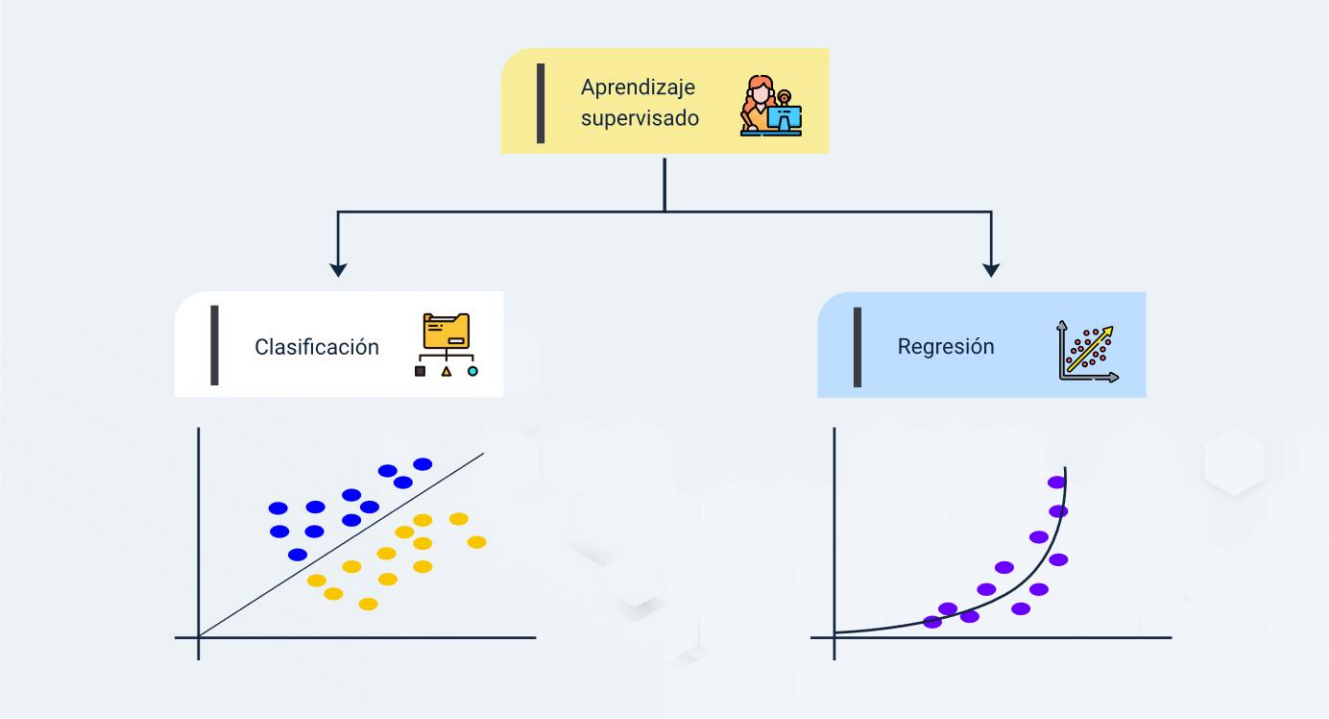


2.1. Modelos de clasificación

A continuación, una breve introducción a algoritmos de aprendizaje supervisado. Por su importancia en la clasificación de datos se puede encontrar que en este tipo de aprendizaje los algoritmos trabajan con observaciones y contienen variables de entrada y variables de salida o etiquetas relacionadas con las variables de entrada.

Las técnicas de aprendizaje supervisado consisten en dos clases principales: clasificación y regresión, dependiendo del tipo de problema de aprendizaje automático a resolver.

Figura 5. Aprendizaje supervisado



El modelo entrenado

Puede ser utilizado para predecir salidas de cualquier conjunto nuevo de datos de entrada. Estas técnicas se definen como supervisados puesto que el modelo aprende de muestras de datos y sus salidas en la fase de entrenamiento.

A continuación, se muestran más generalidades y aspectos destacados sobre las técnicas del aprendizaje supervisado:

- **Regresión.** Consiste en la estimación de un valor numérico con base en los datos de entrada; es una de las técnicas más usadas en "Machine Learning". Por ejemplo, se podría generar un modelo de regresión cuya variable Y dependiente será el ingreso del empleado y X sería la variable independiente o explicativa; X puede estar relacionada con la educación de

ese empleado, su experiencia, estrato, etc. En la fórmula presentada, se pretende estimar el valor m, el cual es la pendiente de la regresión, y B, que es el intercepto en eje Y.

- **Clasificación.** Su objetivo es predecir etiquetas de salida de naturaleza categórica, por tanto, cada salida es una variable discreta entre un número limitado de resultados. Algunos ejemplos de aplicaciones de algoritmos de clasificación son la detección de fraudes, la detección de "spam", clasificación de enfermedades o diagnósticos basados en edad, sexo, azúcar en la sangre, etc., clasificación de imágenes, identificación de caracteres, clasificación de posibles clientes.

2.2. Algoritmos no supervisados

El aprendizaje no supervisado tiene la misión de descubrir similitudes, patrones o uniformidades dentro de los datos de entrada; en este caso, no existe un supervisor que etiquete los datos. Los algoritmos de este tipo de aprendizaje forman clústeres en forma autónoma y asignan observaciones a estos clústeres.

Por ejemplo, si los datos son miles de fotos de leones y tigres, en el aprendizaje no supervisado el programa ordena las fotos de los leones en un clúster y la de los tigres en otro. Este algoritmo toma decisiones de ordenamiento de forma independiente, de acuerdo con características similares y características diferentes.

Un algoritmo no supervisado aprende de un modelo de clústeres a partir de los datos de entrenamiento que se pueden utilizar más tarde para asignar nuevos datos a los clústeres.

Los algoritmos no supervisados se agrupan en problemas de asociación y agrupación, así:

- **Asociación.** Trabaja en función de reglas de asociación que permiten establecer asociaciones entre los datos dentro de bases de datos grandes. Por ejemplo, si un usuario compra carro nuevo, tiene probabilidades de comprar un seguro contra accidentes.

  Así, los algoritmos combinan los datos basándose en los atributos que se comparten, aquí la idea no es encontrar semejanzas entre ellos, sino encontrar relaciones entre los datos.

- **Reglas de asociación.** Un concepto en este tipo de aprendizaje son las reglas de asociación que se usan para extraer conocimiento; estas reglas se obtienen de los datos históricos identificando relaciones entre los datos.

  Una regla se define como X -> Y, donde X y Y son conjuntos de datos distintos.

- **Agrupamiento.** Se trata de identificar un patrón en datos no categorizados y agrupados en clústeres o grupos. Se supone que los datos tienen similitudes identificadas por métricas de distancia, tales como la distancia euclídea. Así, dos registros que tengan una distancia mínima, en comparación con otras distancias, podrían pertenecer al mismo clúster. Es bien importante, entonces, conocer cómo se calcula la distancia.

Por medio de un ejemplo en Python y distancia euclídea, se identificarán los estados más similares en Estados Unidos, en cuanto a porcentaje de asaltos, asesinatos

y secuestros por cada 100.000 habitantes, para cada uno de los 50 estados, usando el código fuente en Python. Este código fuente se puede ejecutar usando Jupyter lab.

Los resultados son los siguientes, ordenados en forma descendente y ascendente:

Figura 6. Resultado 1 de ejemplo

	estado_2	estado_1	distancia
444	Vermont	Florida	6.138335
1383	North Dakota	Nevada	6.113387
1394	Vermont	Nevada	6.105144
433	North Dakota	Florida	6.096939
244	Vermont	California	6.093594
...	...	...	....
425	Montana	Florida	4.312995
1547	West Virginia	New Mexico	4.311465

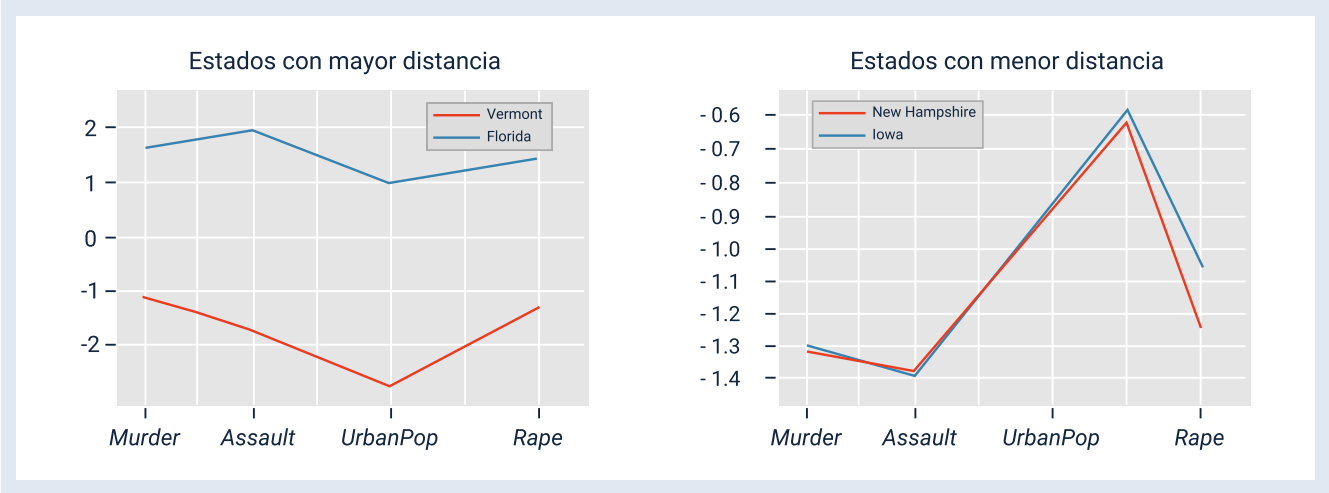
Figura 7. Resultado 2 de ejemplo

	estado_2	estado_1	distancia
728	New Hampshire	Iowa	0.207944
631	New York	Illinois	0.353774
665	Kansas	Indiana	0.433124
1148	Wisconsin	Minnesota	0.499099
928	New Hampshire	Maine	0.504669
...	...	...	....
971	Michigan	Maryland	1.091065

Los estados con mayor distancia y, por tanto, completamente diferentes en porcentaje de asaltos, asesinatos y secuestros son Vermont y Florida, que presentan una distancia euclidiana calculada de 6.14, estos estarán en clústeres separados.

Los estados con menor distancia y, por tanto, muy similares en porcentaje de asaltos, asesinatos y secuestros son New Hampshire y Iowa, que presentan una distancia euclidiana calculada de 0.21; estos podrían estar en el mismo clúster de porcentaje de asaltos, asesinatos y secuestros.

Figura 8. Distancia euclídea



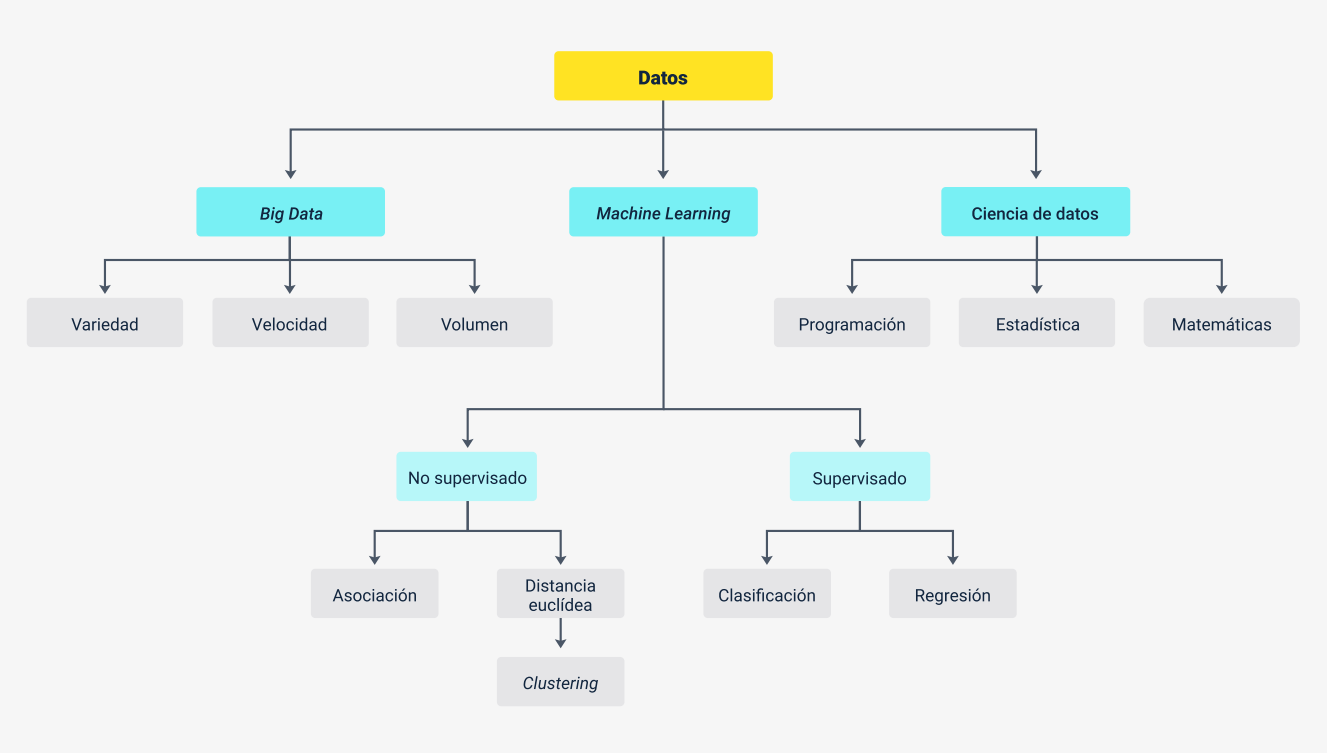
La industria 4.0 va tomando gran importancia en estos últimos años; uno de sus objetivos es que cualquier proceso de producción, sea cual sea, esté completamente automatizado, evitando en lo posible que las personas trabajen en forma manual y dejando que las máquinas o equipos trabajen por sí solos, traduciendo esto en reducción de costos, seguridad, eficiencia y productividad en los resultados.

Pero para llevar a cabo todo esto se requieren datos y estos datos no trabajan solos, existe el aprendizaje automático, que aprende de los datos que recibe

constantemente, realiza predicciones, identifica patrones y es capaz de obtener conocimiento y conclusiones.

Síntesis

Aquí finaliza el estudio de los temas de este componente formativo. En este punto, analice el esquema que se muestra enseguida y haga su propia síntesis de los temas vistos. ¡Adelante!



El esquema presenta la síntesis de la síntesis de la temática estudiada en el componente formativo, comenzando por los datos de donde se deriva el "Big Data", el "Machine Learning" y la ciencia de datos.

Material complementario

Tema	Referencia	Tipo de material	Enlace del recurso
1.1 Conjunto de datos	Anaconda. (s. f.). Installing on Windows. Anaconda Documentation.	Documento web	https://docs.anaconda.com/free/anaconda/install/windows/
2. “Clusterin g”	Singh, S. (2017). Iris.csv. Kaggle.	“Dataset”	https://www.kaggle.com/datasets/saurabh00007/iriscsv

Glosario

Aprendizaje automático: rama de la inteligencia artificial cuyo objetivo es implementar técnicas que permitan a los computadores aprender mediante un proceso de inducción del conocimiento.

Aprendizaje automático no supervisado: cuando el algoritmo identifica patrones y saca conclusiones de los datos que se le proporcionan.

Aprendizaje automático supervisado: cuando el algoritmo recibe datos de entrenamiento consistentes en datos etiquetados.

Clúster: conjunto de objetos o registros que son similares entre sí.

"Clustering": proceso de dividir un conjunto de objetos o registros en subconjuntos llamados clústeres, que tienen similaridades.

Distancia euclídea: es la longitud del segmento entre dos puntos; en el caso del "clustering", define las observaciones más cercanas para asignarlas a un clúster.

Inteligencia artificial: sistemas informáticos que pueden aprender como aprende un ser humano.

"K-means": lenguaje de alto nivel usado para construir todo tipo de aplicaciones y muy usado en ciencia de datos.

"Machine Learning": aprendizaje automático o máquinas de aprendizaje.

Python: proceso criptográfico que proporciona comunicaciones seguras a través de las redes, haciendo que la información entre extremos se transporte en forma segura, mediante uso de criptografía.

Referencias bibliográficas

Abdulhamit, S. (2020). Data analysis using python. Academic Press.

Akram. (2018). Mall-customers. Kaggle. https://www.kaggle.com/akram24/mall-customers

Anaconda. (s. f.). Installing on Windows. Anaconda Documentation. https://docs.anaconda.com/anaconda/install/windows/

Rusell, R. (2018). Machine Learning. Step-by-Step Guide to implement machine learning algorithms with python.

Severance, C. (2020). Python para todos: explorando la información con Python 3.

Singh, S. (2017). Iris.csv. Kaggle. https://www.kaggle.com/datasets/saurabh00007/iriscsv

Créditos

Nombre	Cargo	Regional y Centro de Formación
Claudia Patricia Aristizábal	Líder del Ecosistema	Dirección General
Rafael Neftalí Lizcano Reyes	Responsable de Línea de Producción	Centro Industrial del Diseño y la Manufactura - Regional Santander
Héctor Henry Jurado Soto	Experto Temático	Centro de Teleinformática y Producción Industrial - Regional Cauca
Caterine Bedoya Mejía	Diseñadora Instruccional	Centro de Gestión Industrial - Regional Distrito Capital
Carolina Coca Salazar	Metodóloga	Centro de Diseño y Metrología - Regional Distrito Capital
Darío González	Corrector de Estilo	Centro de Diseño y Metrología - Regional Distrito Capital
Fabián Leonardo Correa Díaz	Diseñador Instruccional	Centro Industrial del Diseño y la Manufactura - Regional Santander
Carmen Alicia Martínez Torres	Animador y Productor Multimedia	Centro Industrial del Diseño y la Manufactura - Regional Santander
Wilson Andrés Arenales Cáceres	Storyboard e ilustración	Centro Industrial del Diseño y la Manufactura - Regional Santander
Camilo Andrés Bolaño Rey	Locución	Centro Industrial del Diseño y la Manufactura - Regional Santander
Yerson Fabian Zarate Saavedra	Diseñador de Contenidos Digitales	Centro Industrial del Diseño y la Manufactura - Regional Santander
Andrea Paola Botello De la Rosa	Desarrollador Fullstack	Centro Industrial del Diseño y la Manufactura - Regional Santander
Emilsen Alfonso Bautista	Actividad didáctica	Centro Industrial del Diseño y la Manufactura - Regional Santander

Nombre	Cargo	Regional y Centro de Formación
Daniel Ricardo Mutis Gómez	Evaluador para Contenidos Inclusivos y Accesibles	Centro Industrial del Diseño y la Manufactura - Regional Santander
Zuleidy María Ruíz Torres	Validador de Recursos Educativos Digitales	Centro Industrial del Diseño y la Manufactura - Regional Santander
Luis Gabriel Urueta Álvarez	Validador de Recursos Educativos Digitales	Centro Industrial del Diseño y la Manufactura - Regional Santander