

UNIVERSIDAD CENTRAL
MAESTRÍA EN ANALÍTICA DE DATOS
SEMINARIO DE ANALÍTICA DE DATOS
SESIÓN 1: INTRODUCCIÓN A LA ANALÍTICA

UC-Analítica-Seminario

Henry Mauricio Sánchez

Sesión 1

CONCEPTOS FUNDAMENTALES

Dato-Información-Conocimiento

Dato-Información-Conocimiento

Datos

Datos

- ☞ Los datos son el conjunto de valores cuantitativos o cualitativos de variables que describen los atributos o estados de un sistema. Esta representación se puede dar por medio de números, letras del alfabeto, una representación algorítmica, entre otras.
- ☞ Los datos se caracterizan por no contener ninguna información.
- ☞ Pueden ser letras, palabras, descripciones.

Datos

- ➡ Los datos son el conjunto de valores cuantitativos o cualitativos de variables que describen los atributos o estados de un sistema. Esta representación se puede dar por medio de números, letras del alfabeto, una representación algorítmica, entre otras.
- ➡ Los datos se caracterizan por no contener información alguna.
- ➡ Pueden ser letras, palabras, descripciones.

Ejemplos de datos



Dato-Información-Conocimiento

Datos

- ➡ Los datos son el conjunto de valores cuantitativos o cualitativos de variables que describen los atributos o estados de un sistema. Esta representación se puede dar por medio de números, letras del alfabeto, una representación algorítmica, entre otras.
- ➡ Los datos se caracterizan por no contener información alguna.
- ➡ Pueden ser letras, palabras, descripciones.

Ejemplos de datos



Ejemplos de datos

45	M	1,76	MAESTRÍA
56	F	1,79	DOCTORADO
47	F	1,54	MAESTRÍA
37	F	1,78	MAESTRÍA
39	M	1,59	PROFESIONAL
28	M	1,57	PROFESIONAL
29	M	1,6	PROFESIONAL
61	F	1,54	PROFESIONAL
38	M	1,52	PROFESIONAL

Datos Estructurados Vs Datos No Estructurados

Datos Estructurados Vs Datos No Estructurados

Datos Estructurados

Los datos estructurados son los datos que se encuentran en la mayoría de bases de datos. Son archivos de tipo texto que se suelen mostrar en filas y columnas

Datos Estructurados Vs Datos No Estructurados

Datos Estructurados

Los datos estructurados son los datos que se encuentran en la mayoría de bases de datos. Son archivos de tipo texto que se suelen mostrar en fila y columnas

Ejemplo

EDAD	SEXO	ESTATURA	NIVEL ESCOLAR
21	M	1,54	MAESTRÍA
26	F	1,55	PROFESIONAL
30	F	1,6	DOCTORADO
31	F	1,7	PROFESIONAL
35	M	1,71	MAESTRÍA

Datos Estructurados Vs Datos No Estructurados

Datos Estructurados

Los datos estructurados son los datos que se encuentran en la mayoría de bases de datos. Son archivos de tipo texto que se suelen mostrar en filas y columnas

Datos No Estructurados

Es cualquier forma de dato que no encaja en el modelo relacional o conjunto de tablas de bases de datos: estructurados.

Ejemplo

EDAD	SEXO	ESTATURA	NIVEL ESCOLAR
21	M	1,54	MAESTRÍA
26	F	1,55	PROFESIONAL
30	F	1,6	DOCTORADO
31	F	1,7	PROFESIONAL
35	M	1,71	MAESTRÍA

Datos Estructurados Vs Datos No Estructurados

Datos Estructurados

Los datos estructurados son los datos que se encuentran en la mayoría de bases de datos. Son archivos de tipo texto que se suelen mostrar en filas y columnas

Ejemplo

EDAD	SEXO	ESTATURA	NIVEL ESCOLAR
21	M	1,54	MAESTRÍA
26	F	1,55	PROFESIONAL
30	F	1,6	DOCTORADO
31	F	1,7	PROFESIONAL
35	M	1,71	MAESTRÍA

Datos No Estructurados

Es cualquier forma de dato que no encaja en el modelo relacional o conjunto de tablas de bases de datos: estructurados.

Ejemplo



Dato-Información-Conocimiento

Información

Información

- ☞ La información corresponde a una contextualización de los datos desde el punto de vista del observador, es decir, es el proceso en el cual los datos adquieren significado, establecen relaciones y se asignan patrones.
- ☞ Es un conjunto organizado de datos que constituye un mensaje.
- ☞ Datos con significado.

Dato-Información-Conocimiento

Información

- ☞ La información corresponde a una contextualización de los datos desde el punto de vista del observador, es decir, es el proceso en el cual los datos adquieren significado, establecen relaciones y se asignan patrones.
- ☞ Es un conjunto organizado de datos que constituye un mensaje.
- ☞ Datos con significado.

Ejemplo de información

Se indagó a cinco personas que estaban haciendo fila en un banco por su edad, sexo, estatura y nivel escolar.

EDAD	SEXO	ESTATURA	NIVEL ESCOLAR
21	M	1,54	MAESTRÍA
26	F	1,55	PROFESIONAL
30	F	1,6	DOCTORADO
31	F	1,7	PROFESIONAL
35	M	1,71	MAESTRÍA

Dato-Información-Conocimiento

Información

- ☞ La información corresponde a una contextualización de los datos desde el punto de vista del observador, es decir, es el proceso en el cual los datos adquieren significado, establecen relaciones y se asignan patrones.
- ☞ Es un conjunto organizado de datos que constituye un mensaje.
- ☞ Datos con significado.

Ejemplo de información

Se indago a cinco personas que estaban haciendo fila en un banco por su edad, sexo, estatura y nivel escolar.

EDAD	SEXO	ESTATURA	NIVEL ESCOLAR
21	M	1,54	MAESTRÍA
26	F	1,55	PROFESIONAL
30	F	1,6	DOCTORADO
31	F	1,7	PROFESIONAL
35	M	1,71	MAESTRÍA

Ejemplo de información

Resumen de data de un estudio en una empresa



Dato-Información-Conocimiento

Conocimiento

Conocimiento

☞ El conocimiento hace referencia a la consolidación de múltiples datos organizados e interrelacionados en un marco conceptual que permite comprender y modificar el entorno, mejorar la capacidad para resolver problemas y sustentar la toma de decisiones, ya sean de carácter personal o desde el ámbito de las organizaciones.

☞ Apoyo a toma de decisiones.

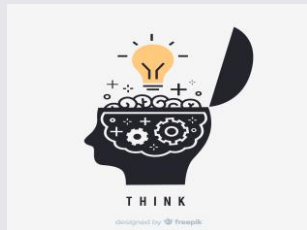
Dato-Información-Conocimiento

Conocimiento

☞ El conocimiento hace referencia a la consolidación de múltiples datos organizados e interrelacionados en un marco conceptual que permite comprender y modificar el entorno, mejorar la capacidad para resolver problemas y sustentar la toma de decisiones ya sean de carácter personal o desde el ámbito de las organizaciones.

☞ Apoyo a toma de decisiones.

Generación de Conocimiento

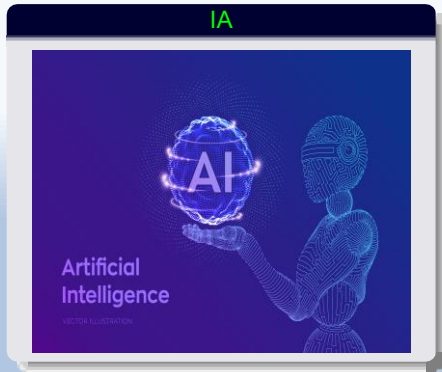


Decisiones



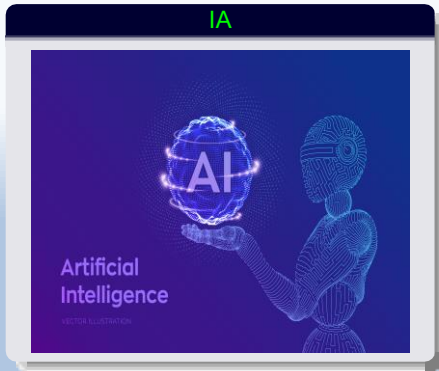
Inteligencia artificial

Inteligencia artificial



Inteligencia artificial

Inteligencia Artificial



Inteligencia artificial

IA

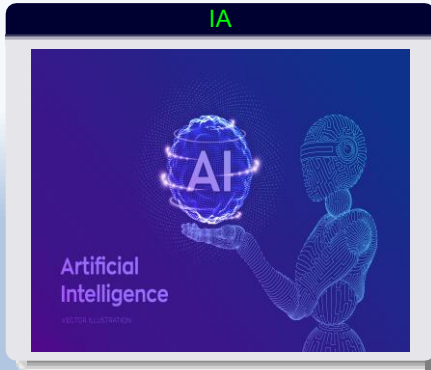


Inteligencia Artificial

- ☞ La creación de dispositivos o algoritmos que imiten o reemplacen al ser humano emulando sus funciones cognitivas.

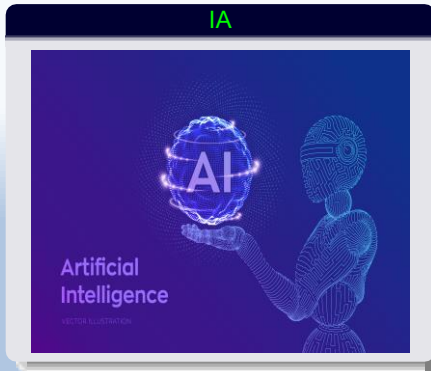
Inteligencia artificial

Inteligencia Artificial



- ☞ La creación de dispositivos o algoritmos que imiten o reemplacen al ser humano emulando sus funciones cognitivas.
- ☞ La IA nació en 1950 con los trabajos de Alan Turing. Inicio de las ciencias de la computación.

Inteligencia artificial



Inteligencia Artificial

- ☞ La creación de dispositivos o algoritmos que imiten o reemplacen al ser humano emulando sus funciones cognitivas.
- ☞ La IA nació en 1950 con los trabajos de Alan Turing. Inicio de las ciencias de la computación.
- ☞ En 1956 se dio el nombre de inteligencia artificial: Dartmouth

Inteligencia artificial

IA



Inteligencia Artificial

- ➡ La creación de dispositivos o algoritmos que imiten o reemplacen al ser humano emulando sus funciones cognitivas.
- ➡ La IA nació en 1950 con los trabajos de Alan Turing. Inicio de las ciencias de la computación.
- ➡ En 1956 se dio el nombre de inteligencia artificial: Dartmouth
- ➡ En 1997 aparece Deep Blue. IBM

IA

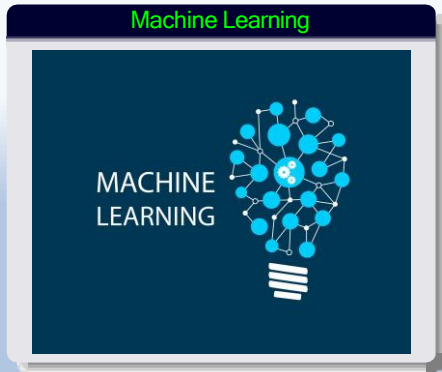


Inteligencia Artificial

- ➡ La creación de dispositivos o algoritmos que imiten o reemplacen al ser humano emulando sus funciones cognitivas.
- ➡ La IA nació en 1950 con los trabajos de Alan Turing. Inicio de las ciencias de la computación.
- ➡ En 1956 se dio el nombre de inteligencia artificial: Dartmouth
- ➡ En 1997 aparece Deep Blue. IBM
- ➡ En los últimos años...

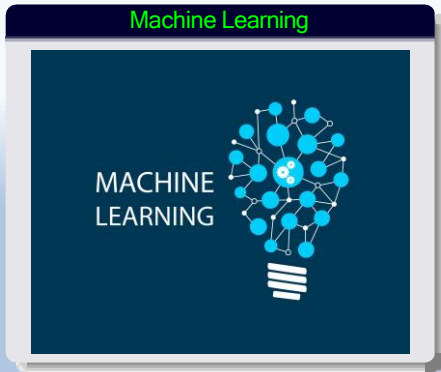
Machine Learning

Machine Learning



Machine Learning

Aprendizaje de Máquina



Machine Learning

Machine Learning

MACHINE
LEARNING



Aprendizaje de Máquina

- ➡ También llamado aprendizaje automático. Es una rama de la Inteligencia Artificial.

Machine Learning

Machine Learning

MACHINE
LEARNING



Aprendizaje de Máquina

- ➡ También llamado aprendizaje automático. Es una rama de la Inteligencia Artificial.
- ➡ Crea sistemas que aprenden de manera automatizada, es decir, identifica patrones complejos en millones de datos y predice comportamientos mediante un algoritmo. Además, los algoritmos son capaces de mejorarse de manera independiente con el tiempo.

Machine Learning

Machine Learning

MACHINE
LEARNING



Aprendizaje de Máquina

- ➡ También llamado aprendizaje automático. Es una rama de la Inteligencia Artificial.
- ➡ Crea sistemas que aprenden de manera automatizada, es decir, identifica patrones complejos en millones de datos y predice comportamientos mediante un algoritmo. Además, los algoritmos son capaces de mejorarse de manera independiente con el tiempo.
- ➡ En 1959 se acuña el término: Arthur Samuel.

Machine Learning

Machine Learning

MACHINE
LEARNING



Aprendizaje de Máquina

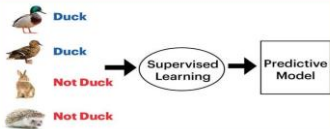
- ➡ También llamado aprendizaje automático. Es una rama de la Inteligencia Artificial.
- ➡ Crea sistemas que aprenden de manera automatizada, es decir, identifica patrones complejos en millones de datos y predice comportamientos mediante un algoritmo. Además, los algoritmos son capaces de mejorarse de manera independiente con el tiempo.
- ➡ En 1959 se acuña el término: Arthur Samuel.
- ➡ Dos grandes grupos del ML: Aprendizaje Supervisado y Aprendizaje no Supervisado

Aprendizaje Supervisado

Aprendizaje Supervisado

Aprendizaje Supervisado

Supervised Learning (Classification Algorithm)

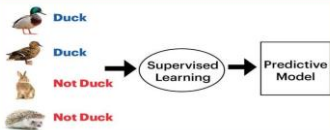


Aprendizaje Supervisado

Definición

Aprendizaje Supervisado

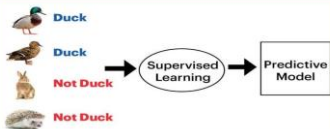
Supervised Learning (Classification Algorithm)



Aprendizaje Supervisado

Aprendizaje Supervisado

Supervised Learning (Classification Algorithm)

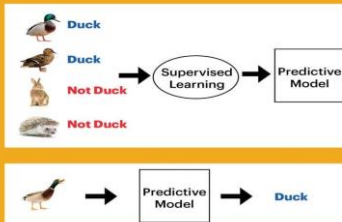


Definición

Aprendizaje Supervisado

Aprendizaje Supervisado

Supervised Learning (Classification Algorithm)



Definición

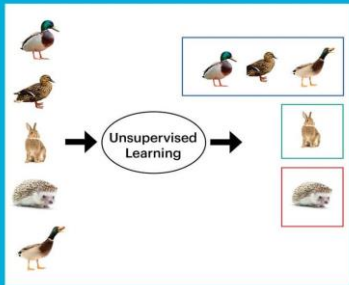
- Los algoritmos de aprendizaje supervisado intentan modelar la relación y las dependencias entre la salida de predicción objetivo y las características de entrada, de manera que podemos predecir los valores de salida para los nuevos datos basados en esas relaciones que ha aprendido de conjuntos de datos anteriores alimentados.
- Existen unas reglas determinadas.

Aprendizaje No Supervisado

Aprendizaje No Supervisado

Machine Learning

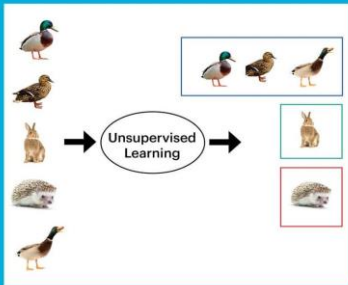
Unsupervised Learning (Clustering Algorithm)



Aprendizaje No Supervisado

Machine Learning

Unsupervised Learning (Clustering Algorithm)

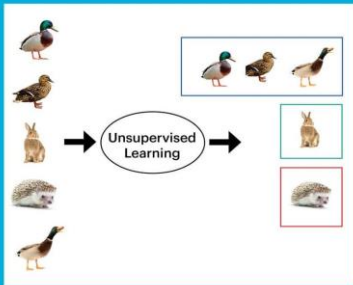


Definición

Aprendizaje No Supervisado

Machine Learning

Unsupervised Learning (Clustering Algorithm)



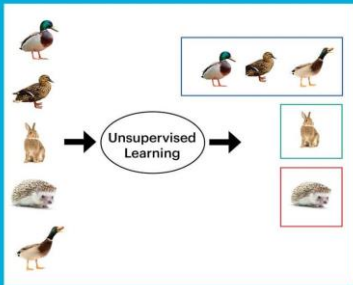
Definición

👉 Familia de algoritmos de aprendizaje automático, que se utilizan principalmente en la detección de patrones y el modelado descriptivo. Estos algoritmos no tienen categorías de salida o etiquetas en los datos (el modelo está entrenado con datos sin etiquetar).

Aprendizaje No Supervisado

Machine Learning

Unsupervised Learning (Clustering Algorithm)



Definición

- ➡ Familia de algoritmos de aprendizaje automático, que se utilizan principalmente en la detección de patrones y el modelado descriptivo. Estos algoritmos no tienen categorías de salida o etiquetas en los datos (el modelo está entrenado con datos sin etiquetar).
- ➡ No existen reglas predeterminadas.

Algunos métodos supervisados

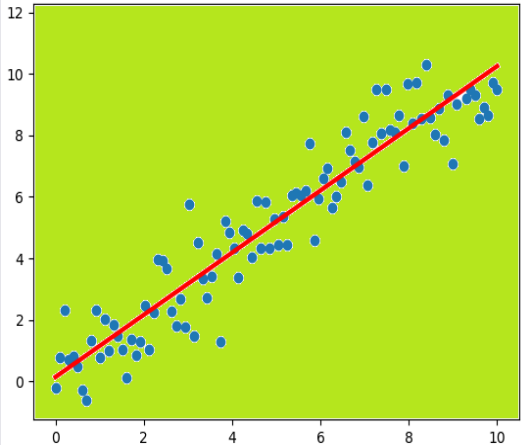
Métodos supervisados

Algunos métodos supervisados

Métodos supervisados

➡ Regresiones: lineal, polinomial.

Regresion lineal



Algunos métodos supervisados

Métodos supervisados

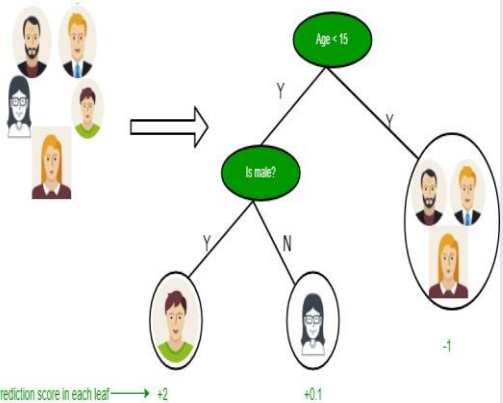
➡ Regresiones: lineal, polinomial.

➡ Árboles de decisión.

Árboles de Decisión

Input: Age, Gender, Occupation, ...

Does the person likes computer games

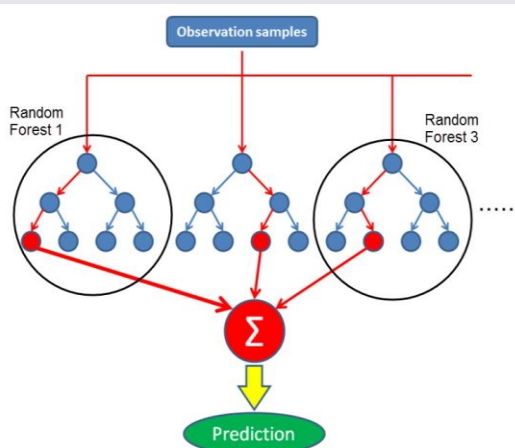


Algunos métodos supervisados

Métodos supervisados

- ☞ Regresiones: lineal, polinomial.
- ☞ Árboles de decisión.
- ☞ **Random Forests**

Random Forests



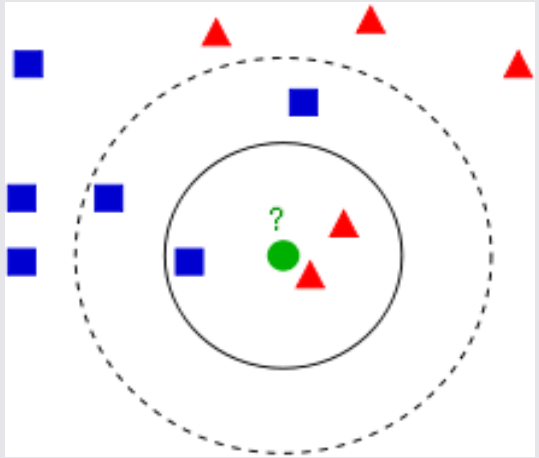
<https://images.app.goo.gl/YDLCg5ZJebrvtbPR6>

Algunos métodos supervisados

Métodos supervisados

- ☞ Regresiones: lineal, polinomial.
- ☞ Árboles de decisión.
- ☞ Random Forests
- ☞ **KNN**: k-nearest neighbors.

KNN



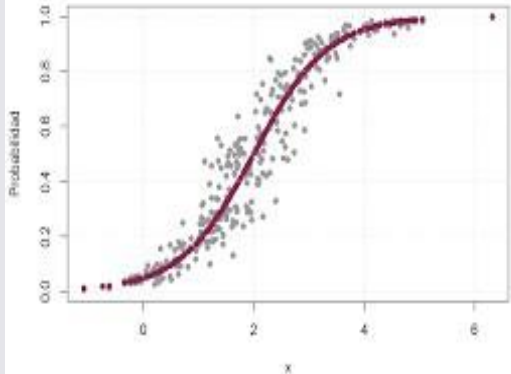
<https://images.app.goo.gl/zcELXkiQCq72QX8b8>

Algunos métodos supervisados

Métodos supervisados

- ➡ Regresiones: lineal, polinomial.
- ➡ Árboles de decisión.
- ➡ Random Forests
- ➡ **KNN**: k-nearest neighbors.
- ➡ **Regresión Logística**

Regresión Logística



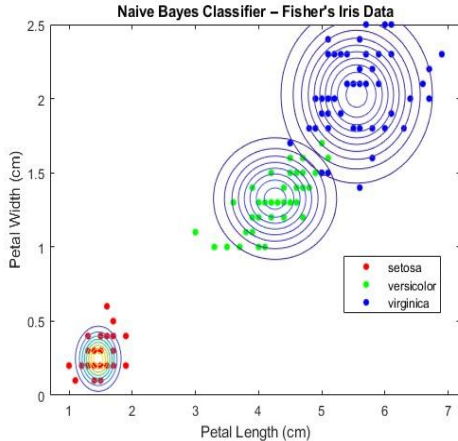
<https://images.app.goo.gl/BrNcLNbvKScokEFy5>

Algunos métodos supervisados

Métodos supervisados

- ➡ Regresiones: lineal, polinomial.
- ➡ Árboles de decisión.
- ➡ Random Forests
- ➡ **KNN**: k-nearest neighbors.
- ➡ Regresión Logística
- ➡ **Naive-Bayes**

Naive-Bayes



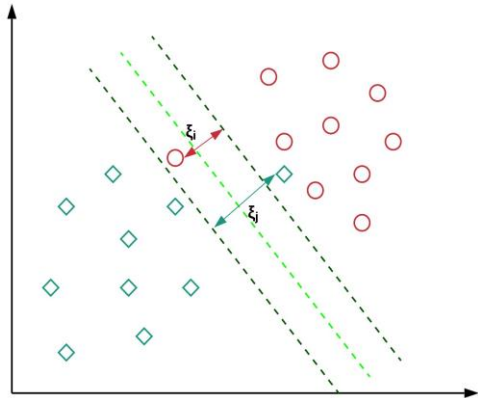
<https://images.app.goo.gl/4udvd9Bmck1h3TQ79>

Algunos métodos supervisados

Métodos supervisados

- ☞ Regresiones: lineal, polinomial.
- ☞ Árboles de decisión.
- ☞ Random Forests
- ☞ **KNN**: k-nearest neighbors.
- ☞ Regresión Logística
- ☞ Naive-Bayes
- ☞ **SVM**: Support Vector Machine

SVM



<https://images.app.goo.gl/8qtcdF1YtX5iYwoAA>

Algunos métodos no supervisados



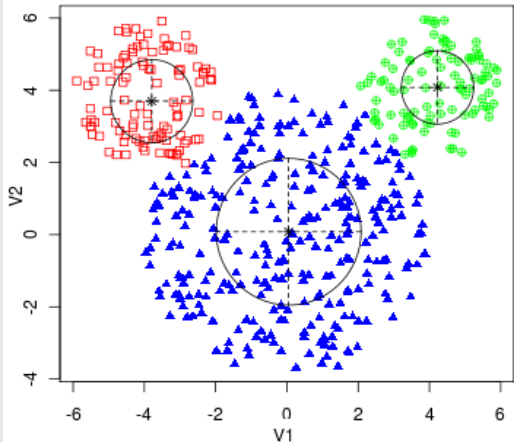
Métodos no supervisados

Algunos métodos no supervisados

Métodos no supervisados

👉 K- means Clustering

K- means Clustering



<https://images.app.goo.gl/eDkZbks2PiTLDvk49>

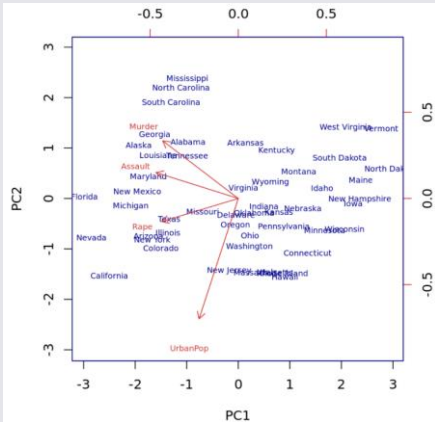
Algunos métodos no supervisados

Métodos no supervisados

☞ K- means Clustering

☞ Reducción de dimensionalidad.

Reduccion de Dimensionalidad



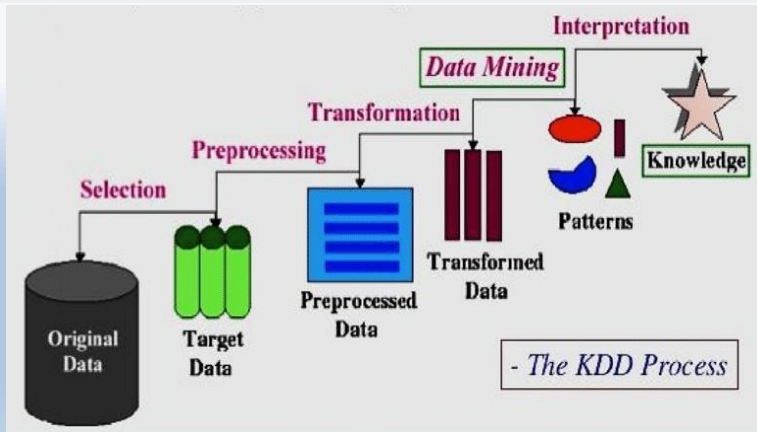
https://www.cienciadedatos.net/documentos/35_principal_component_analysis

Proceso KDD : Knowledge Discovery in Databases

Knowledge Discovery in Databases

Proceso: KDD

KDD o Descubrimiento de Conocimiento a partir de Bases de Datos, es un proceso metodológico y además secuencial, que se sigue para encontrar o extraer conocimiento de un conjunto de datos en bruto.



<https://images.app.goo.gl/iWyzvLVQA3oQVB7q9>

Etapas del proceso KDD

Etapas del proceso KDD

1. Abstracción del Escenario

Entender la problemática a la que nos vamos a enfrentar y tener contexto para proponer soluciones viables y reales. Es importante entender las propiedades, limitaciones y reglas del escenario de estudio para definir las metas a alcanzar.

Etapas del proceso KDD

1. Abstracción del Escenario

Entender la problemática a la que nos vamos a enfrentar y tener contexto para proponer soluciones viables y reales. Es importante entender las propiedades, limitaciones y reglas del escenario de estudio para definir las metas a alcanzar.

2. Selección de datos

Una vez definidos los objetivos del estudio, se deben elegir los datos disponibles para realizar dicho estudio. Luego, se integran los datos de diferentes fuentes que permitan llevar a cabo el estudio.

Etapas del proceso KDD

1. Abstracción del Escenario

Entender la problemática a la que nos vamos a enfrentar y tener contexto para proponer soluciones viables y reales. Es importante entender las propiedades, limitaciones y reglas del escenario de estudio para definir las metas a alcanzar.

2. Selección de datos

Una vez definidos los objetivos del estudio, se deben elegir los datos disponibles para realizar dicho estudio. Luego, se integran los datos de diferentes fuentes que permitan llevar a cabo el estudio.

3. Limpieza y pre-procesamiento

Se determina la confiabilidad de la información. Se debe hacer limpieza de los datos, por ejemplo, datos perdidos, o remover datos atípicos. Se debe considerar la eliminación de variables o atributos con datos faltantes o eliminar información no útil para el problema a resolver.

Etapas del proceso KDD

4. Transformación de los datos

Se debe mejorar la calidad de los datos con algunas transformaciones, por ejemplo, reducción de dimensionalidad o discretización, es decir, convertir los valores que son números a datos categóricos.

Etapas del proceso KDD

4. Transformación de los datos

Se debe mejorar la calidad de los datos con algunas transformaciones, por ejemplo, reducción de dimensionalidad o discretización, es decir, convertir los valores que son números a datos categóricos.

5. Selección de la tarea de Minería de Datos

En esta fase se debe elegir el paradigma que se usará, ya sea de regresión, clasificación o agrupación, de acuerdo a los objetivos del estudio; si es predictivo o descriptivo.

Predictiva: Clasificación y Regresión

Descriptiva: Agrupamiento, Reglas de Asociación

Etapas del proceso KDD

4. Transformación de los datos

Se debe mejorar la calidad de los datos con algunas transformaciones, por ejemplo, reducción de dimensionalidad o discretización, es decir, convertir los valores que son números a datos categóricos.

5. Selección de la tarea de Minería de Datos

En esta fase se debe elegir el paradigma que se usará, ya sea de regresión, clasificación o agrupación, de acuerdo a los objetivos del estudio; si es predictivo o descriptivo.

Predictiva: Clasificación y Regresión

Descriptiva: Agrupamiento, Reglas de Asociación

6. Elección del algoritmo de MD

Una vez seleccionado lo anterior, se debe escoger el algoritmo o los algoritmos para la búsqueda de patrones y obtener conocimiento de los datos. Cada algoritmo tiene su propia esencia, su propia manera de trabajar y de obtener los resultados, por lo que es recomendable conocer las propiedades de aquellos candidatos a utilizar y ver cuál se ajusta mejor a los datos.

Etapas del proceso KDD

7. Aplicación del Algoritmo

En esta etapa se aplica el algoritmo y se van ajustando los parámetros de tal forma que se vayan optimizando los resultados.

Etapas del proceso KDD

7. Aplicación del Algoritmo

En esta etapa se aplica el algoritmo y se van ajustando los parámetros de tal forma que se vayan optimizando los resultados.

8. Evaluación

Una vez se aplican los algoritmos al conjunto de datos, procedemos a evaluar los patrones que se generaron y el rendimiento que se obtuvo, para verificar que cumpla con las metas planteadas en las primeras fases. Se puede usar validación cruzada: parte los datos en entrenamiento y prueba.

Etapas del proceso KDD

7. Aplicación del Algoritmo

En esta etapa se aplica el algoritmo y se van ajustando los parámetros de tal forma que se vaya optimizando los resultados.

8. Evaluación

Una vez se aplican los algoritmos al conjunto de datos, procedemos a evaluar los patrones que se generaron y el rendimiento que se obtuvo, para verificar que cumpla con las metas planteadas en las primeras fases. Se puede usar validación cruzada: parte los datos en entrenamiento y prueba.

9. Aplicación

Finalmente, una vez validado el algoritmo y los resultados, se procede a aplicar el conocimiento adquirido al contexto de estudio con el fin de resolver la(s) problemática(s).

Minería de Datos

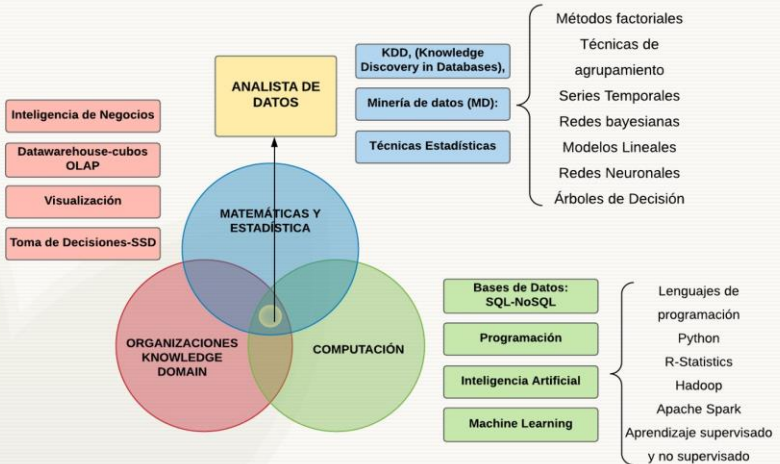
Minería de Datos

La **Minería de Datos** o el Data Mining (DM) es una etapa del proceso KDD. Además, es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos, entre otros.

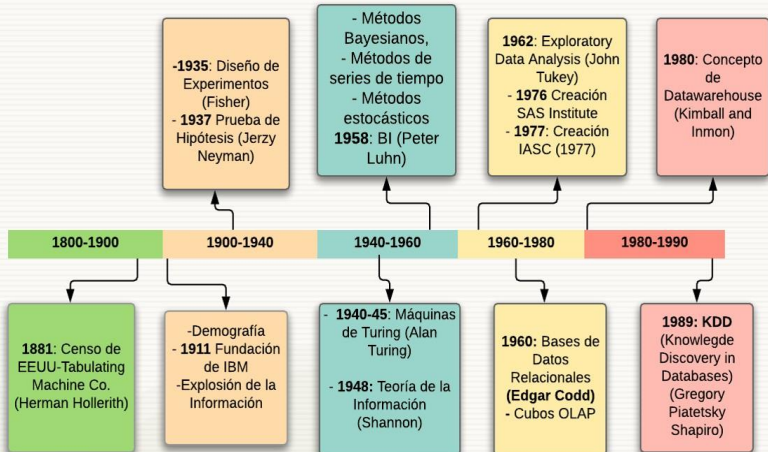


CONVERGENCIA DE LAS DISCIPLINAS Y UN CONTEXTO HISTÓRICO

Convergencia de Disciplinas



Línea de Tiempo



Línea de Tiempo

