

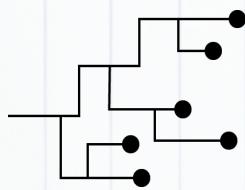
INTRODUCCIÓN A LA CIENCIA DE DATOS

ALEXANDRA POMARES QUIMBAYA

DEFINICIÓN



CONCEPTOS RELACIONADOS



EVOLUCIÓN



HERRAMIENTAS



CASOS DE ÉXITO



MADUREZ ANALÍTICA



PROCESO



CIENCIA DE DATOS

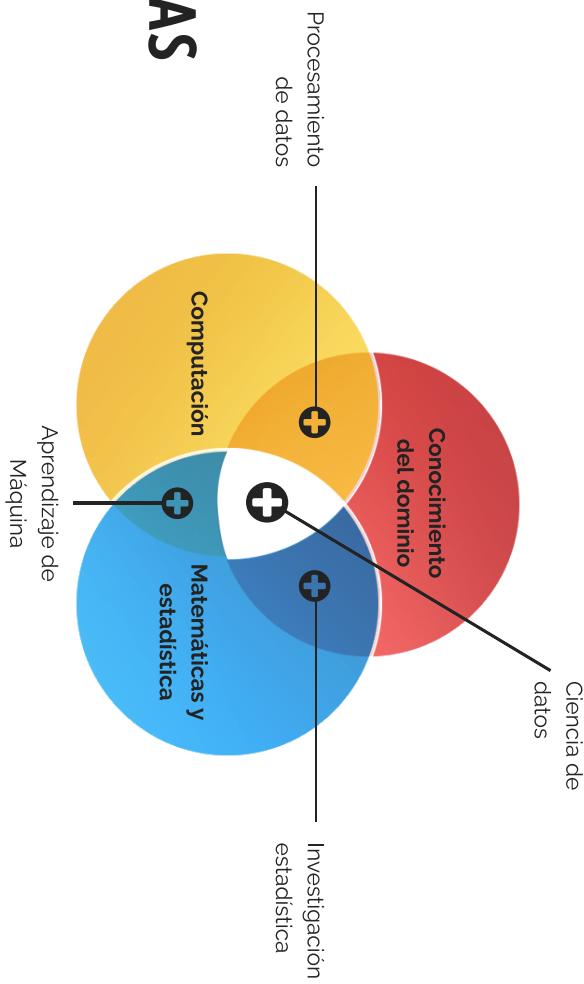
Es el conjunto de principios fundamentales que guían la extracción de conocimiento a partir de los datos.

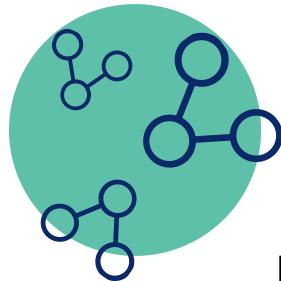
Provost, Foster; Fawcett, Tom



DE DÓNDE TOMA LOS PRINCIPIOS? —

INTEGRA LOS AVANCES DE DIVERSAS
ÁREAS





Analítica de datos

QUÉ ES?

Es el proceso de extracción
de Conocimiento



Válido



Útil y
comprendible



Previamente
desconocido

A PARTIR DE



DATOS

Almacenados

En distintos formatos

CUÁLES FORMATOS ?



Estructurados



No
Estructurados



Semi
estructurados

La analítica de datos usa los principios de la
ciencia de datos

MINERÍA DE DATOS

ES SINÓNIMO DE ANALÍTICA DE DATOS

Es el proceso de extraer conocimiento válido, útil y comprensible, previamente desconocido, desde datos

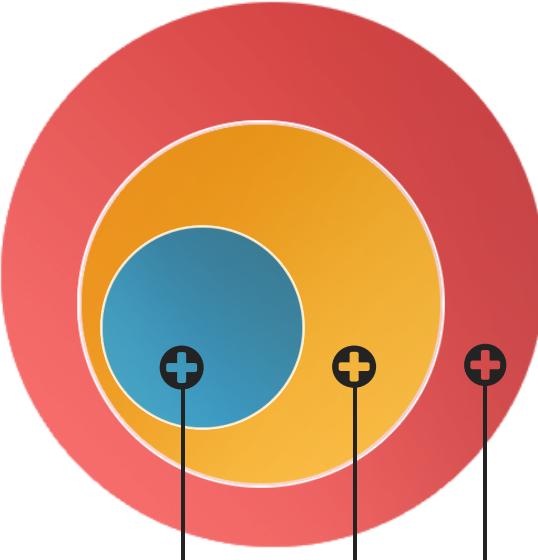


La minería de datos usa los principios de la ciencia de datos

APRENDIZAJE DE MÁQUINA

LA CIENCIA DE DATOS USA LAS TÉCNICAS DE ML

Es la ciencia de hacer que las computadoras actúen sin ser programadas explícitamente.



A diagram consisting of three concentric circles. The innermost circle is blue, the middle is yellow, and the outermost is red. Three lines extend from the right side of the diagram to the right, each ending in a small circle with a plus sign (+). The top line points to the word 'Inteligencia' above 'Artificial'. The middle line points to the words 'Aprendizaje de' above 'máquina'. The bottom line points to the words 'Deep Learning'.

Inteligencia
Artificial
Aprendizaje de
máquina
Deep Learning

BIG DATA

LA CIENCIAS DE DATOS INCORPORA TECNOLOGÍAS BIG DATA PARA GENERAR Y APLICAR MODELOS EN PROBLEMAS CON LAS TRES V

Big Data es usado para describir montos grandes de datos diversos, tanto estructurados como no estructurados, a los cuales las organizaciones pueden acceder de manera rápida, para analizarlos usando herramientas innovadoras, que, en conjunto, ayudan a determinar con precisión oportunidades de mejora en la gestión y en la generación de valor

Big Data es un paradigma que busca dar soluciones a problemas que involucren las tres V

BIG DATA / VOLUMEN



Redes sociales



E-commerce



Sensores

BIG DATA / VELOCIDAD



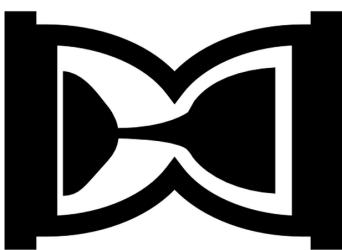
Real Time

- Recomendación a cliente geolocalizada
- Alerta de fraude



Near Real Time

- Personalización de oferta de producto
- Acción sube o baja?
- Personalizar portal



Batch

- Entender comportamiento de un producto
- Planeación de compra efectivo

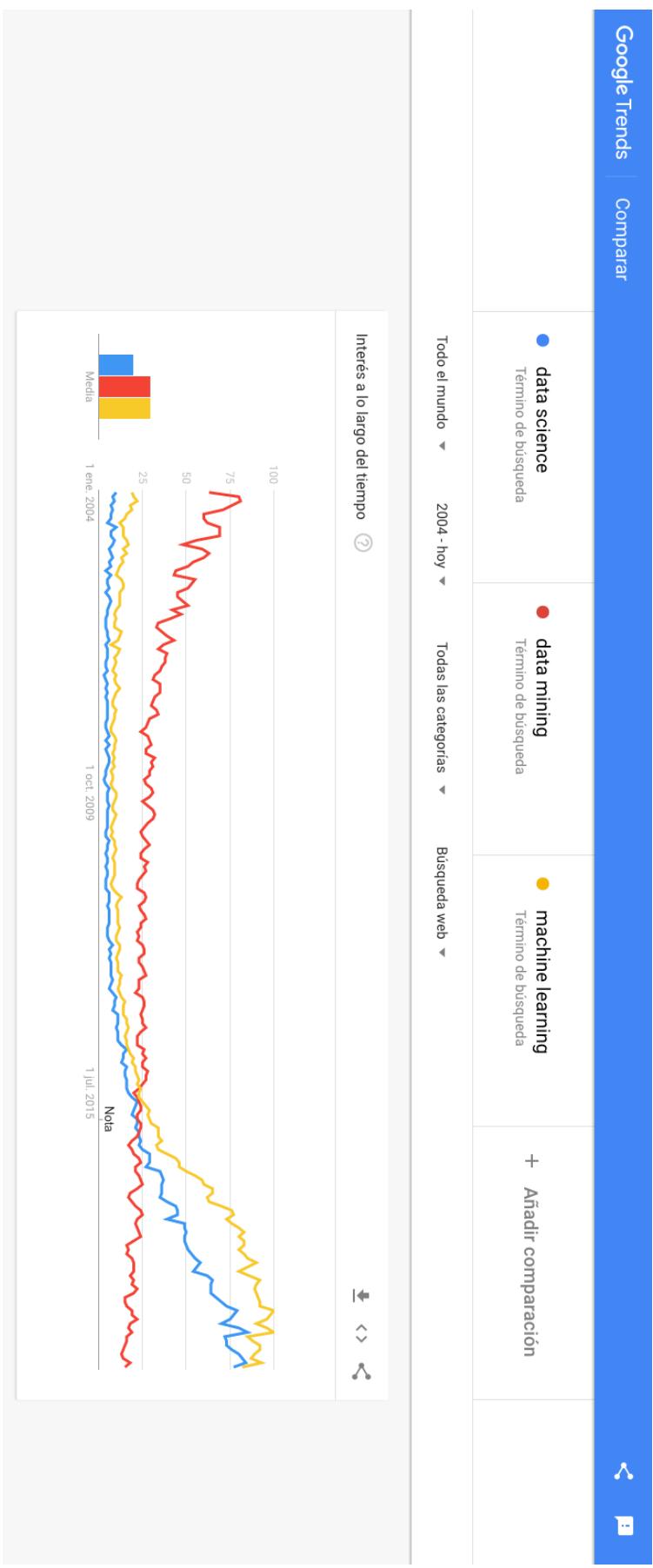
BIG DATA / VARIEDAD



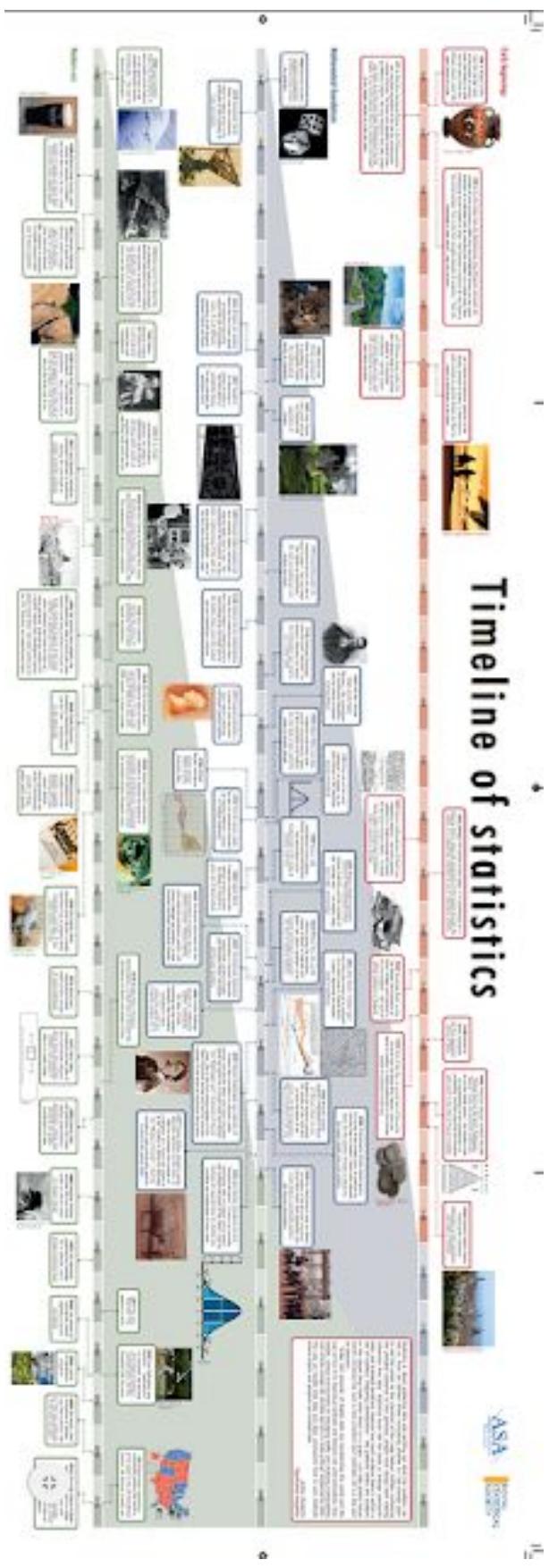
BIG DATA / MÁS V



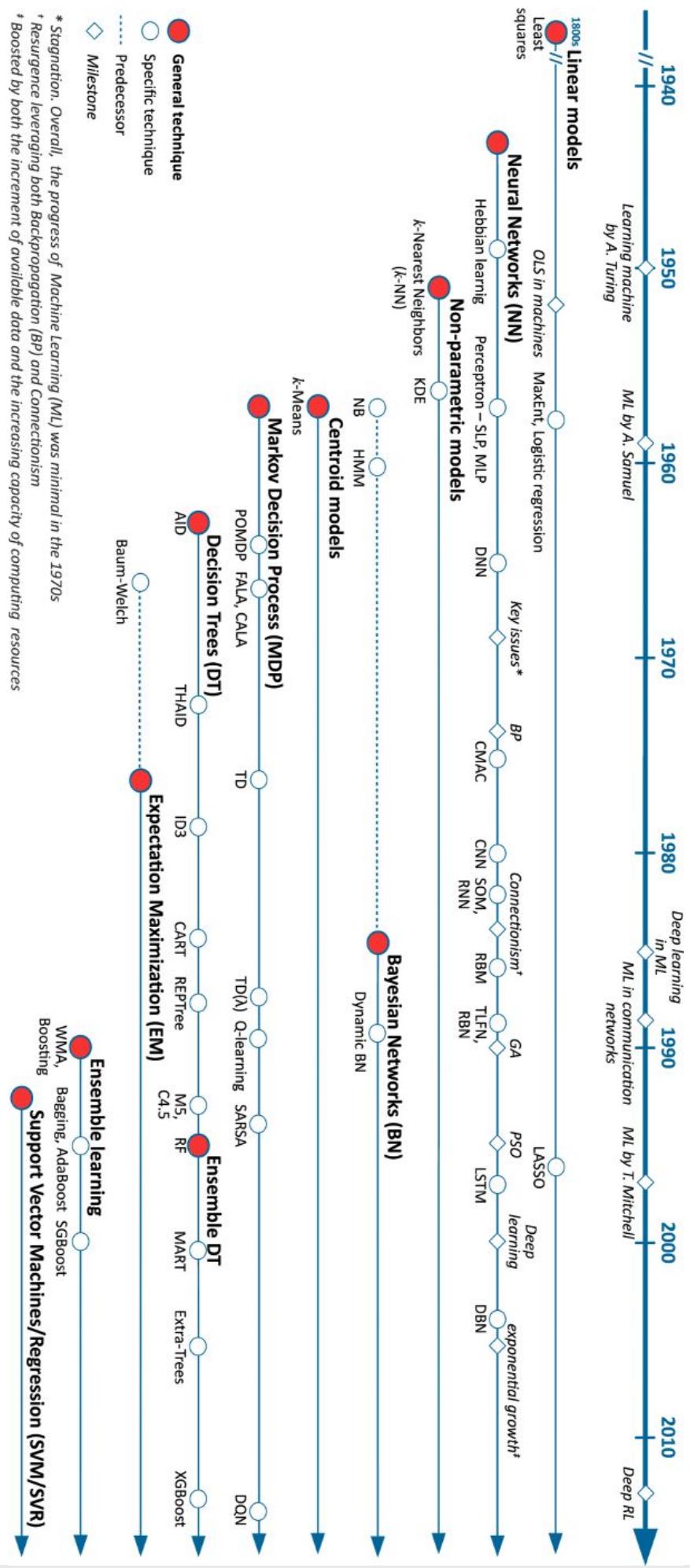
EVOLUCIÓN DEL USO DEL CONCEPTO



EVOLUCIÓN DE LAS TECNOLOGÍAS DE LA ESTADÍSTICA



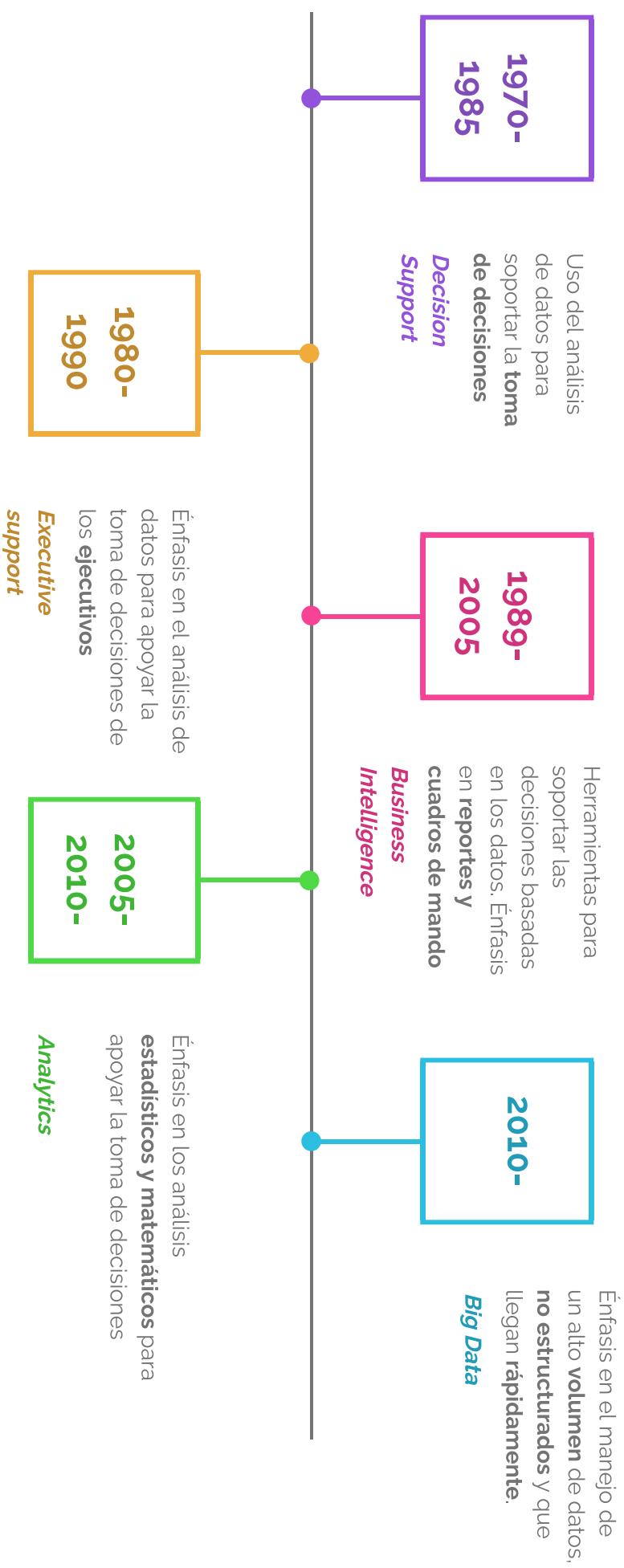
EVOLUCIÓN DE LAS TECNOLOGÍAS DE ML



Boutaba, R., Salahuddin, M.A., Limam, N. et al. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. *J Internet Serv Appl* 9, 16 (2018)

EVOLUCIÓN DE LOS SISTEMAS DE SOPORTE A LA TOMA DE DECISIONES

1970-Actualmente

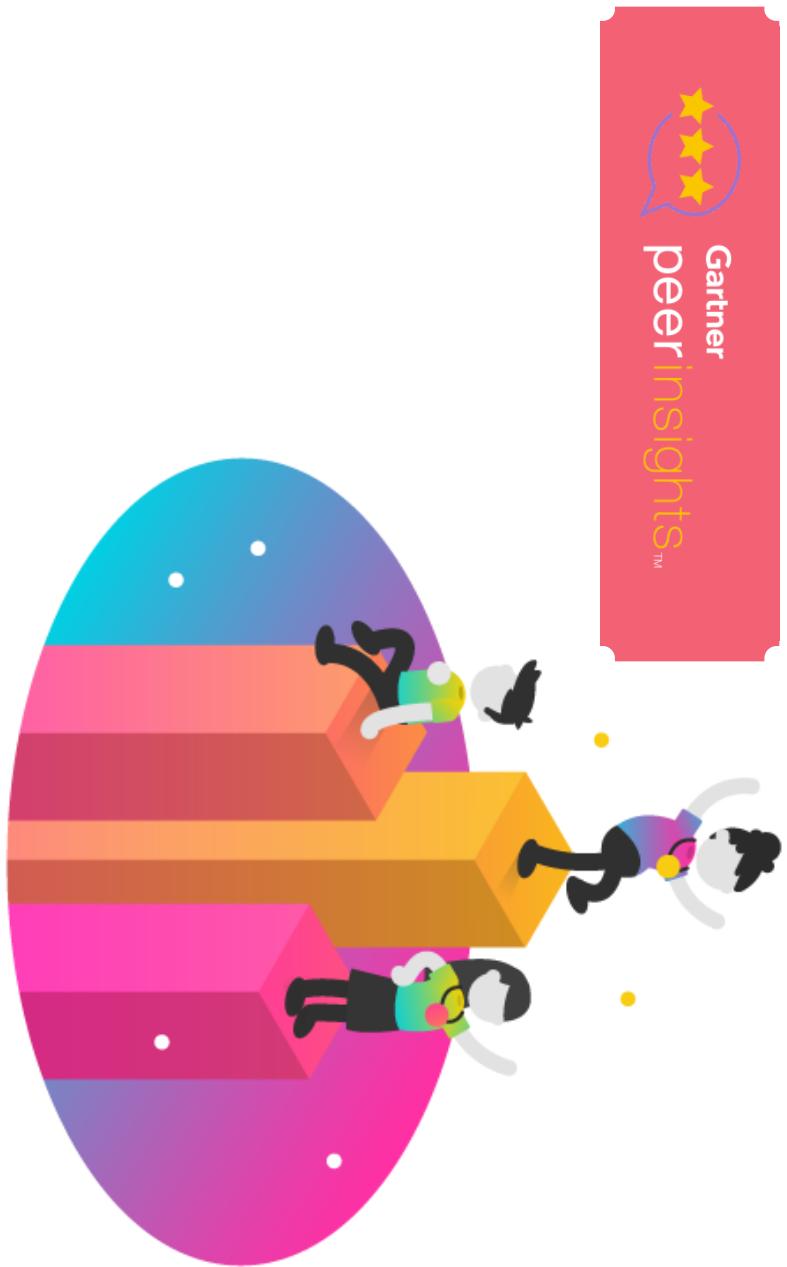


Definición y conceptos básicos de ciencia de datos

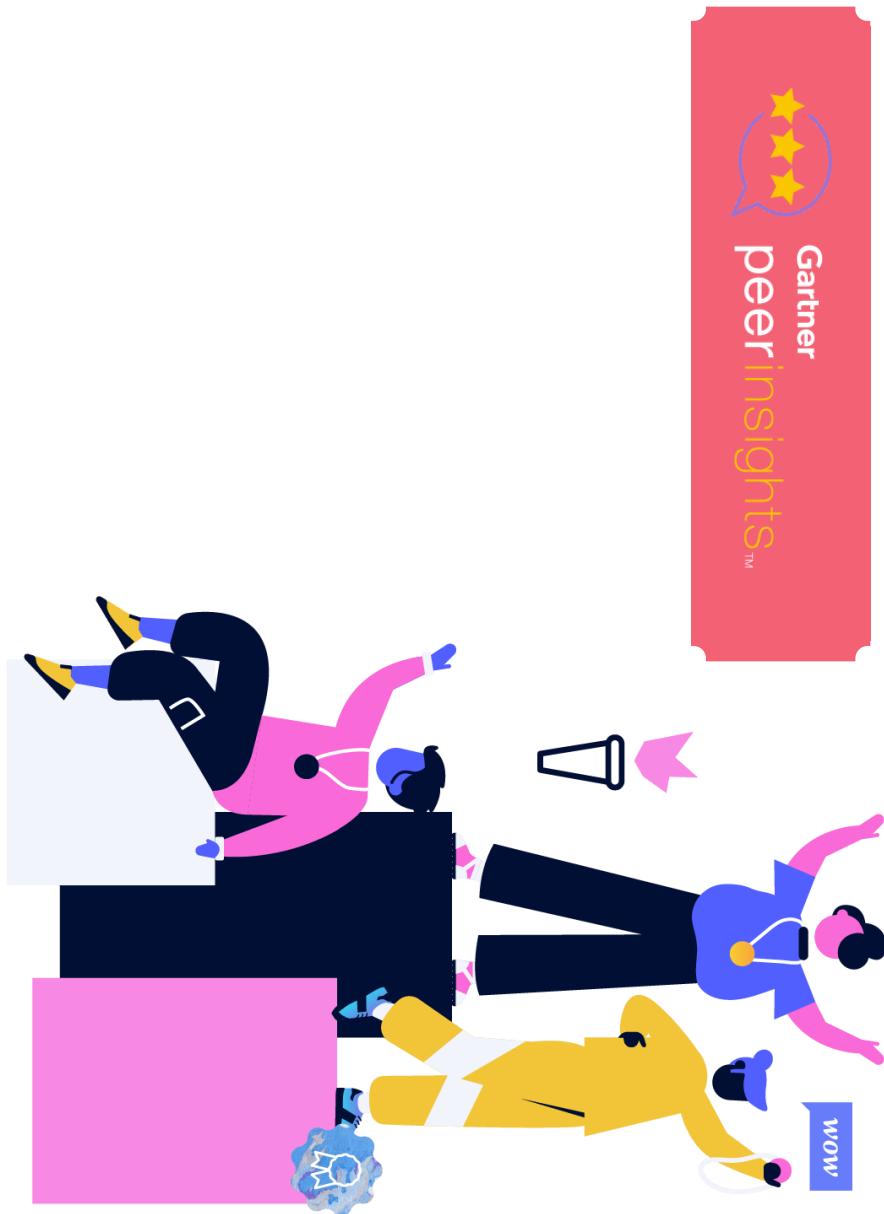
QUE HE APRENDIDO



HERRAMIENTAS DE CIENCIA DE DATOS - MACHINE LEARNING



HERRAMIENTAS DE ANALÍTICA E INTELIGENCIA DE NEGOCIOS



HOME



Contexto de la empresa

Decisiones

- ¿Qué series/películas comprar o producir, qué series mantener?
- ¿Qué recomendar?
- ¿Qué trailers utilizar?
- ¿Cómo lograr que usemos más el servicio?

Datos

- Más de 165 millones de usuarios
- ¿Qué ve?
- ¿Cuándo hace pausa, adelanta, pasa rápido?
- ¿Qué días y a qué hora ve películas?
- ¿Dónde ve, qué dispositivo usa, cómo busca qué ver?
- Información de las películas: actores, directores, género, etc.
- Información de series más pirateadas

Nuevo conocimiento y acciones

General

- Las personas que ven más de 15 horas al mes tienen menos probabilidad de cancelar la suscripción.
- Los que ven menos de 5 horas al mes tienen más probabilidad de cancelar (95%)

Acciones

- Recomendación adaptada
- Usuarios independientes
- Postplay

YES, I'M STILL WATCHING
STOP JUDGING ME, NETFLIX

► Continue Playing

✗ Back to Browse

Nuevo conocimiento y acciones

Series y películas

- Metadatos de películas vs. Metadatos de usuarios
- Análisis por director, por actor, por género vs. Usuarios

¿Compro House of Cards?

Muchos usuarios ven películas dirigidas por David Fincher (The Social Network).

A la versión británica de "House of Cards" le fue bien en los 90s.

Aquellos que vieron la versión británica les gusta ver películas de Kevin Spacey y/o dirigidas por David Fincher.

Acciones

- Principio de la mayor felicidad por dólar
- ¿Qué películas entregan una felicidad similar a un menor precio?

¿Compro Caballero de la Noche?

Memento dirigida por Christopher Nolan

El Secreto de la Montaña protagonizada por Heath Ledger

Gracias por fumar protagonizada por Aaron Eckhart

El maquinista protagonizada por Christian Bale



Contexto de la empresa - Mayor empresa de alimentos

Decisiones

- Qué nuevos alimentos generar para responder a las necesidades de los consumidores?
- Cómo llegar a los nuevos consumidores a través de las marcas?
- Qué perfiles de usuarios tienen?

Datos

- Más de 6 millones de usuarios de twitter
- Descripción del usuario
- Mensajes de los tweets
- Metadatos de los tweets (hora, día, dispositivo, lugar, idioma)

Solución



Perfilamiento de consumidores

Asignación de Segmentos de Consumidor

Segmento: A

Descripción:
"Ingeniera devoradora de kilómetros, libros y poesías. Training for @maratonmedellin"

Analisis de Tendencias

Interés	Porcentaje
Deportes	38%
Música	20%
Amigos	19%
Familia	12%
Salud	5%
Moda	3%
Tecnología	2%
Política	1%
Entretenimiento	1%
Religión	1%

Analisis de Clustering

A: Clúster 1

Características del clúster:
Mujeres, viven en Región Andina, considerables presencia de profesionales, hablan positivamente de deportes y entretenimiento.

Analisis de Comunidades

Comunidad 1:
Hashtags de la comunidad: tbt, happy, Cartagena, Friends, love, music, family, travel, selfie, Instagram y 90 más.

Maria Victoria Hulk
@cardonamariav
Género: Femenino.
Antigüedad: 6 años.
Tipo persona: Persona.
Rango edad: 18 – 24.
Nivel educativo: Profesionales de las Ciencias Y de la Ingeniería.
Estrato: Universitario o Superior.
Idioma: Español.
Ubicación: Medellín, Colombia.
Nivel de influencia: Común.

Nota: Aplica para usuarios con más de 20 tuits.

OPEN DATA



MOVIMIENTO INFORMACIÓN BIEN COMÚN

Dato de origen público o privado difundido de manera estructurada según un método establecido o una licencia que garantiza su acceso libre sin restricciones de tipo técnico, jurídico o financiero.

3

2

1

Index





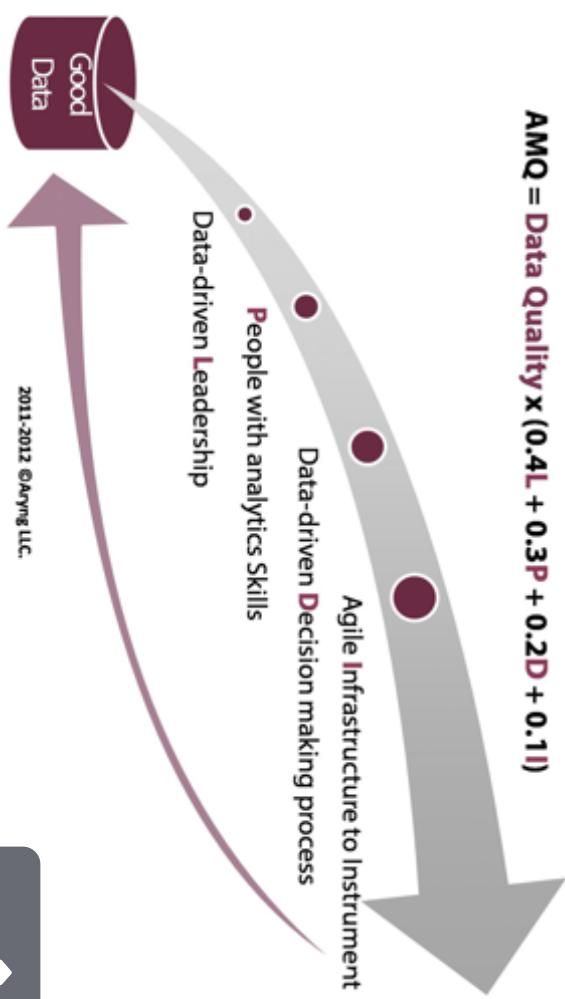
5 niveles para la

MADUREZ EN LA ARQUITECTURA DE INFORMACIÓN

MODELO DE MADUREZ ANALITICO

Analytics Maturity QuotientTM (AMQTM)

$$\text{AMQ} = \text{Data Quality} \times (0.4L + 0.3P + 0.2D + 0.1I)$$



▲ HOME

PROCESO

CRISP-DM

Cross Industry Standard Process for Data Mining. Creada por:

- DaimlerChrysler
- SPSS- IBM
- NCR

