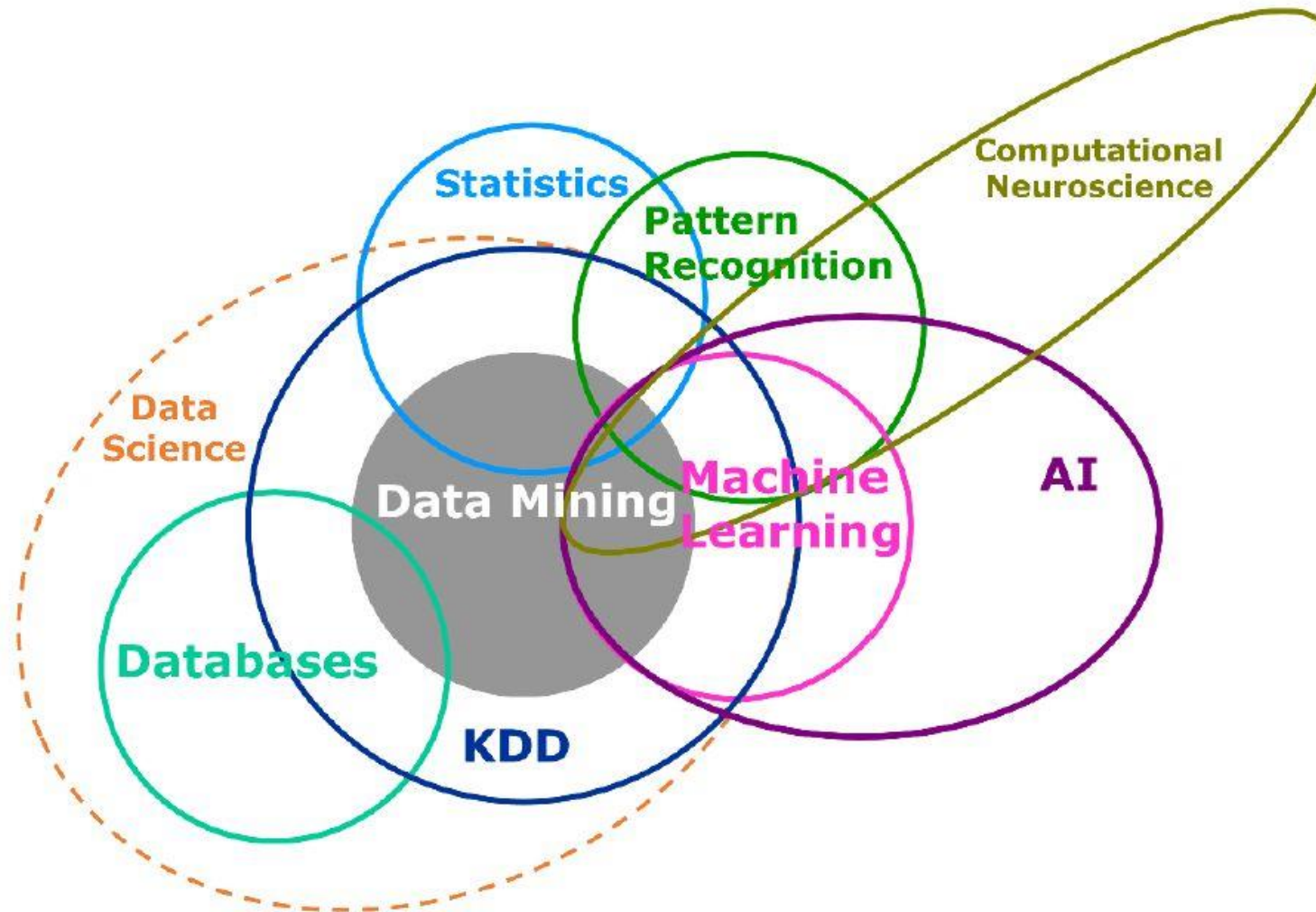


Introducción al aprendizaje automático

Jorge Bedoya

Machine Learning, Data Science, and Statistics



<https://jamesmccaffrey.wordpress.com/2016/09/29/machine-learning-data-science-and-statistics/>

<https://www.datasciencecentral.com/ms-data-science-vs-ms-machine-learning-ai-vs-ms-analytics/>

Relación entre la AI y el aprendizaje automático

- Máquina que imita las funciones “cognitivas”: percibir, razonar, aprender y resolver problemas¹
- Rama de las ciencias computacionales encargada de estudiar modelos de cómputo capaces de realizar actividades propias de los seres humanos en base a dos de sus características primordiales: el razonamiento y la conducta²
- La capacidad de un sistema para interpretar correctamente datos externos, para aprender de dichos datos y emplear esos conocimientos para lograr tareas y metas concretas a través de la adaptación flexible³

1. Poole, David. [«Computational Intelligence: A Logical Approach»](#)

2. Takeya

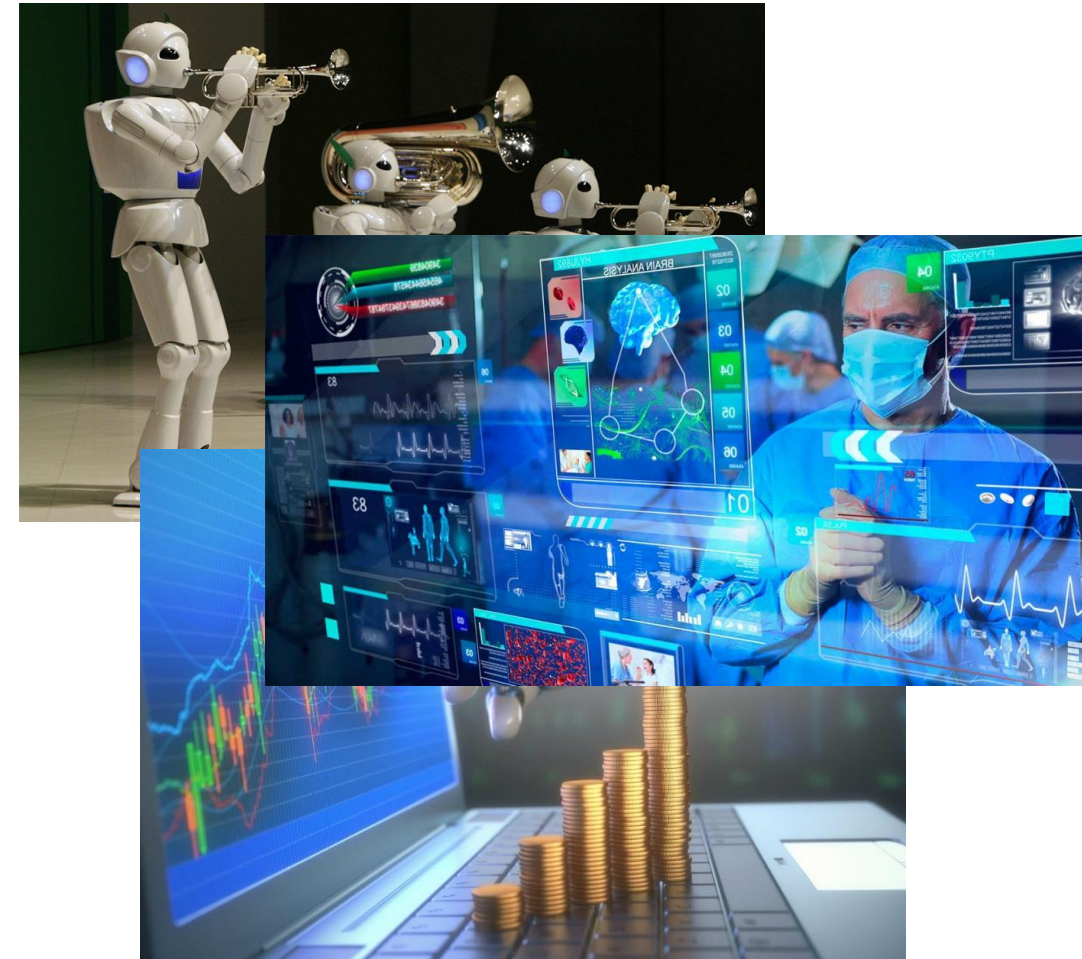
3. Andreas Kaplan y Michael Haenlein



Aplicaciones

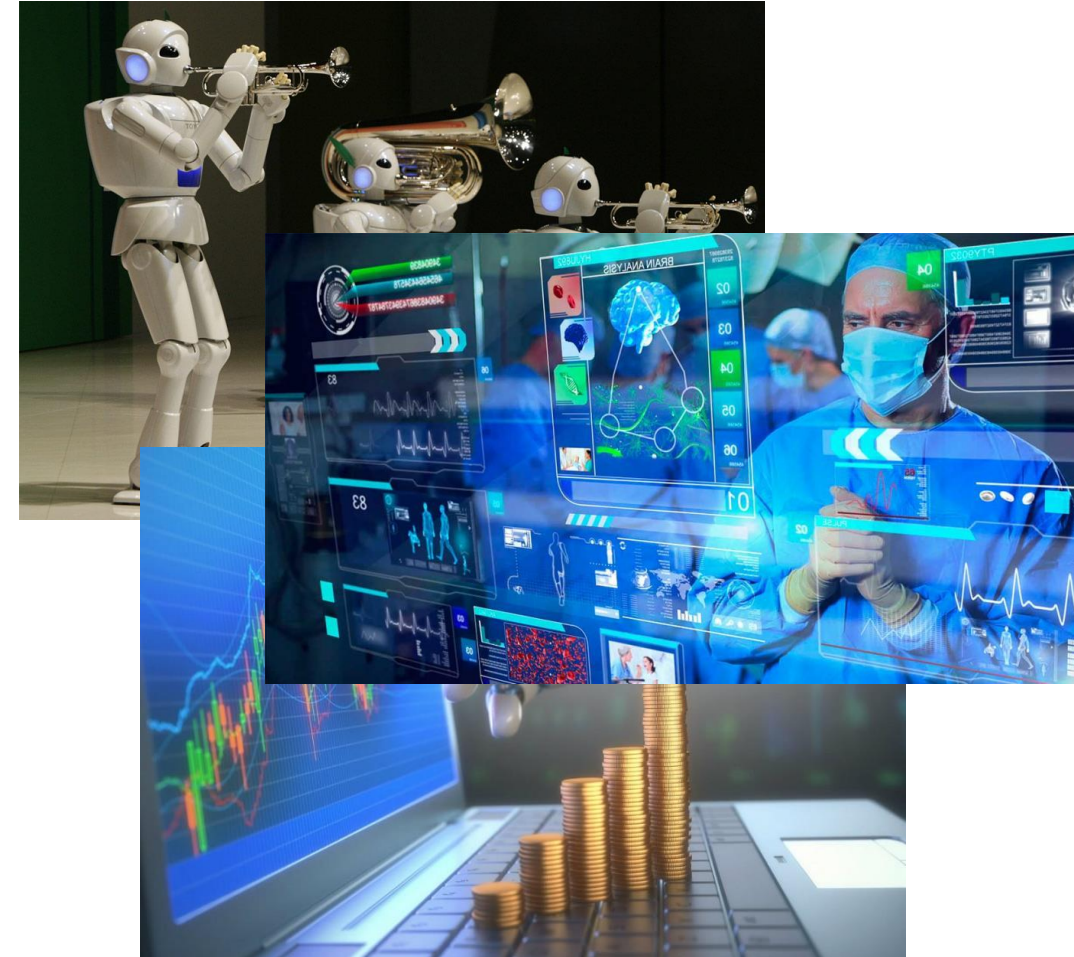
Aplicaciones

- **Financieras y de banca**
 - Análisis de riesgos de crédito
 - Obtención de patrones de fraude en tarjetas de crédito
 - Correlaciones entre indicadores financieros
- **Análisis de mercado**
 - Análisis de canasta de compra
 - Segmentación de clientes
 - Análisis de fidelidad de clientes. Reducción de fuga
- **Seguros y salud privada**
 - Predicción de clientes que contratarán nuevas pólizas.
 - Determinación de clientes que podrían ser potencialmente caros.



Aplicaciones

- **Educación**
 - Selección o captación de estudiantes
 - Detección de abandonos o fracaso
- **Medicina**
 - Identificación de patologías
 - Recomendación priorizada de fármacos
- **Biología**
 - Análisis de secuencias de genes
 - Modelos de calidad del agua
- **Telecomunicaciones**
- **Procesos industriales**
- **ETC.**

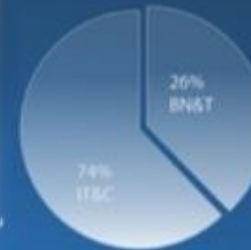


Metodologías

Projected sales of main products in 2013



Distribution of market share among the major industry players



Distribution of market share among the major industry players: IT & C and BN & T was 74% and 26% percent respectively. A further change in the economic situation in the market will be characterized by a more equal distribution of market share major players

Share of market activity



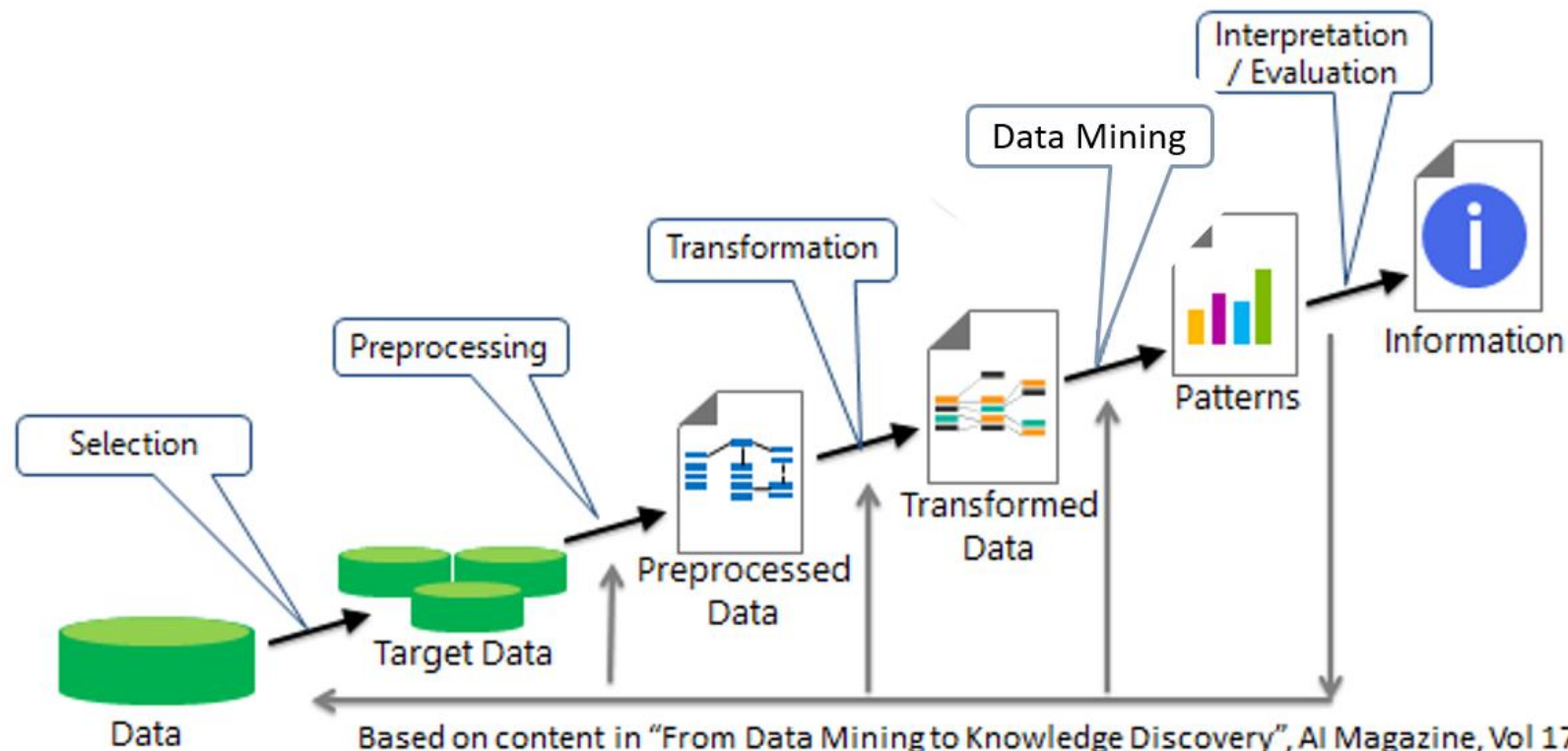
Changes in the activity of the active and passive market is uncertain. Established positive trends in various market segments.

Projected sales of main products in 2013



Perspectiva técnica

KDD: Knowledge Discovery in Databases process

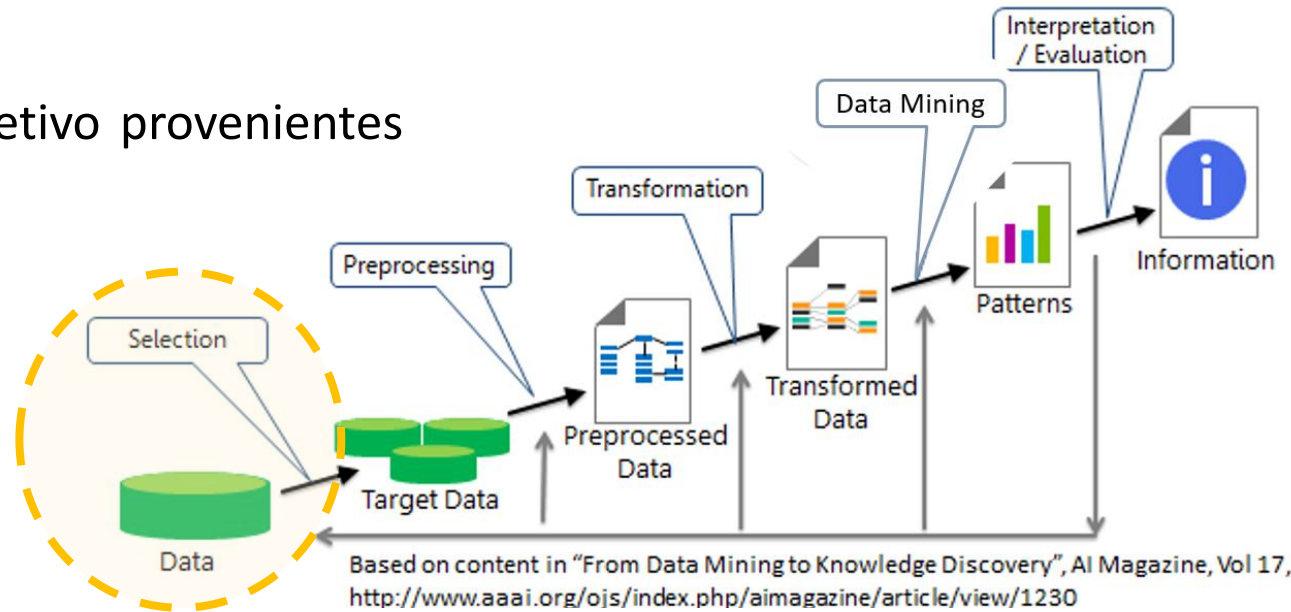


Based on content in "From Data Mining to Knowledge Discovery", AI Magazine, Vol 17, No. 3 (1996)
<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>

KDD: Knowledge Discovery in Databases process

1. SELECCIÓN:

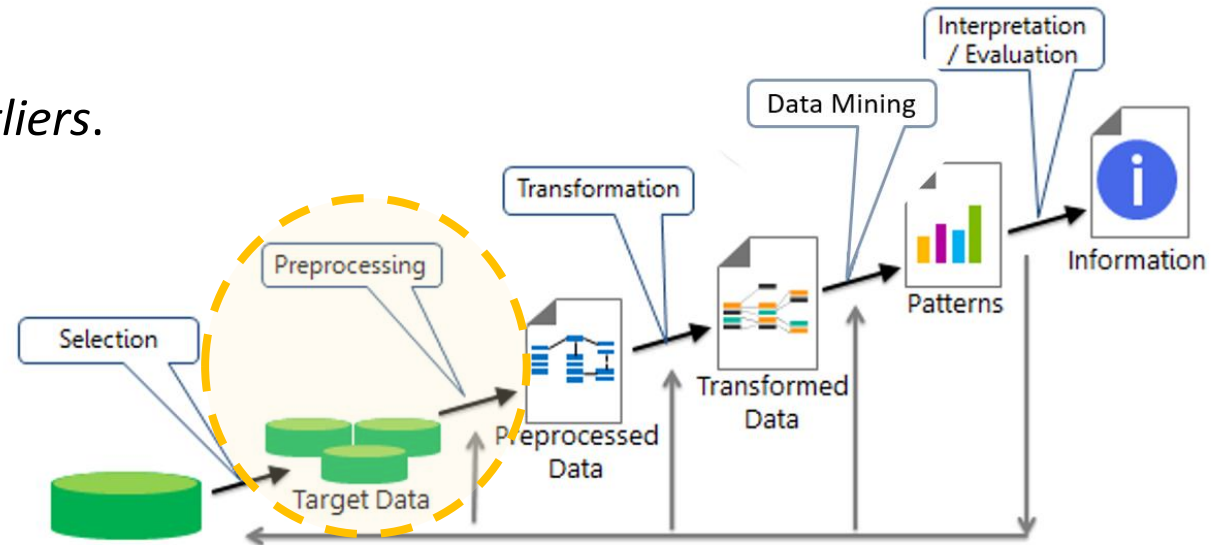
Selección e integración de los datos objetivo provenientes de fuentes múltiples y heterogéneas



KDD: Knowledge Discovery in Databases process

2. PREPROCESAMIENTO:

- Eliminación de ruido y datos aislados o *outliers*.
- Uso del conocimiento previo para Eliminar las inconsistencias y los duplicados.
- Escogencia y uso de estrategias para manejar la información faltante en los datasets.

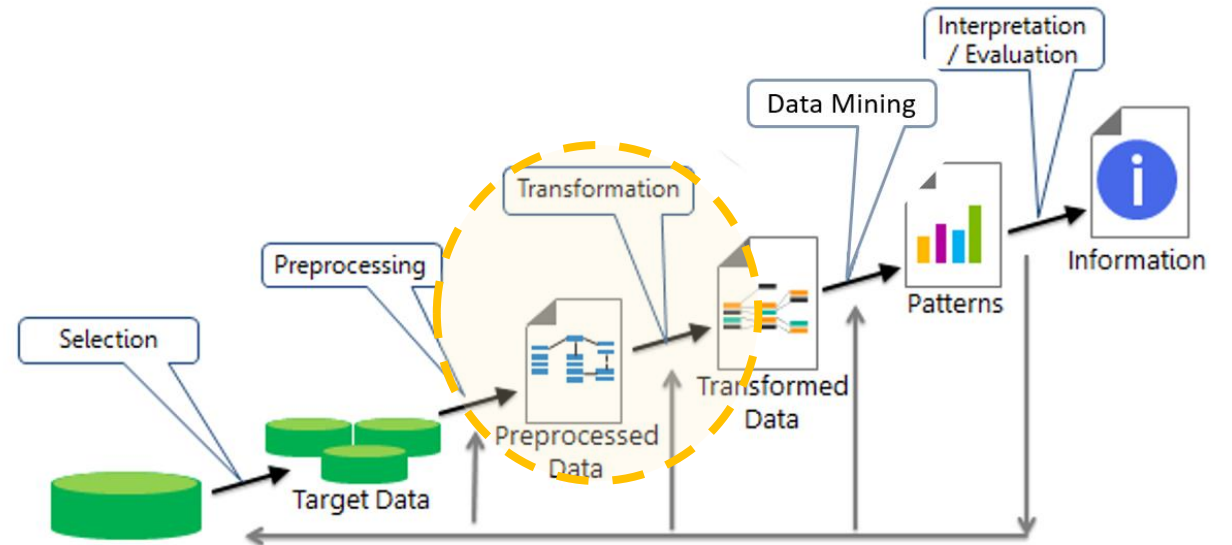


Based on content in "From Data Mining to Knowledge Discovery", AI Magazine, Vol 17, No. 3 (1996)
<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>

KDD: Knowledge Discovery in Databases process

3. TRANSFORMACIÓN:

- Preparación de los datos para el análisis
- Uso de transformaciones de atributos como: numerización, discretización, etc.
- El resultado es un conjunto de filas y columnas denominado vista minable

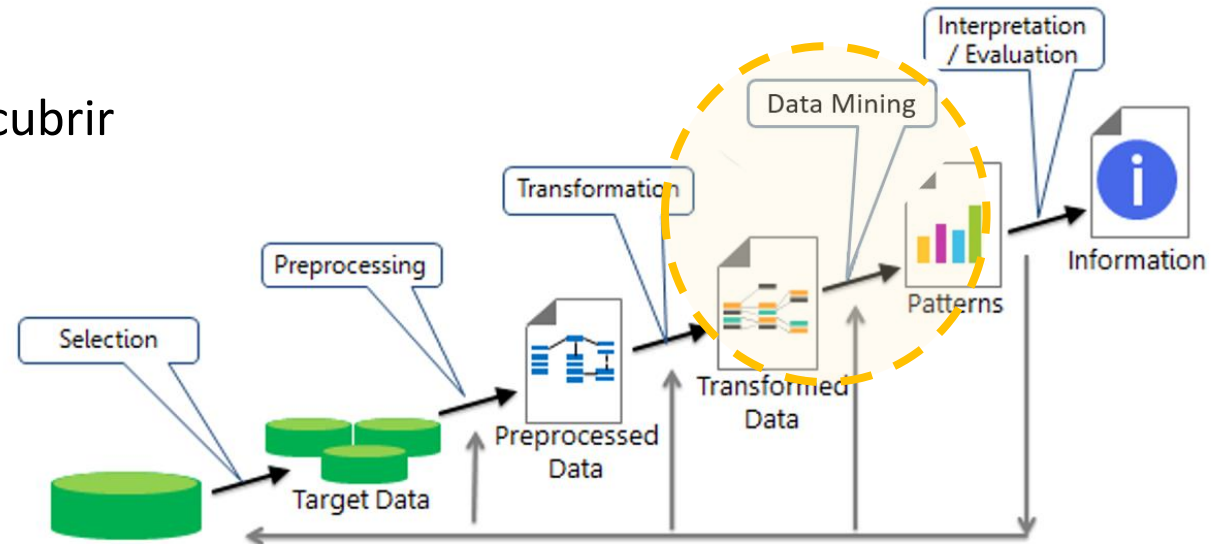


Based on content in "From Data Mining to Knowledge Discovery", AI Magazine, Vol 17, No. 3 (1996)
<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>

KDD: Knowledge Discovery in Databases process

4. MINERÍA DE DATOS:

- Análisis de los patrones o relaciones a descubrir
- Se comprende de 3 pasos:
 - Selección de la tarea
 - Selección del algoritmo(s)
 - Aplicación/Entrenamiento del algoritmo.

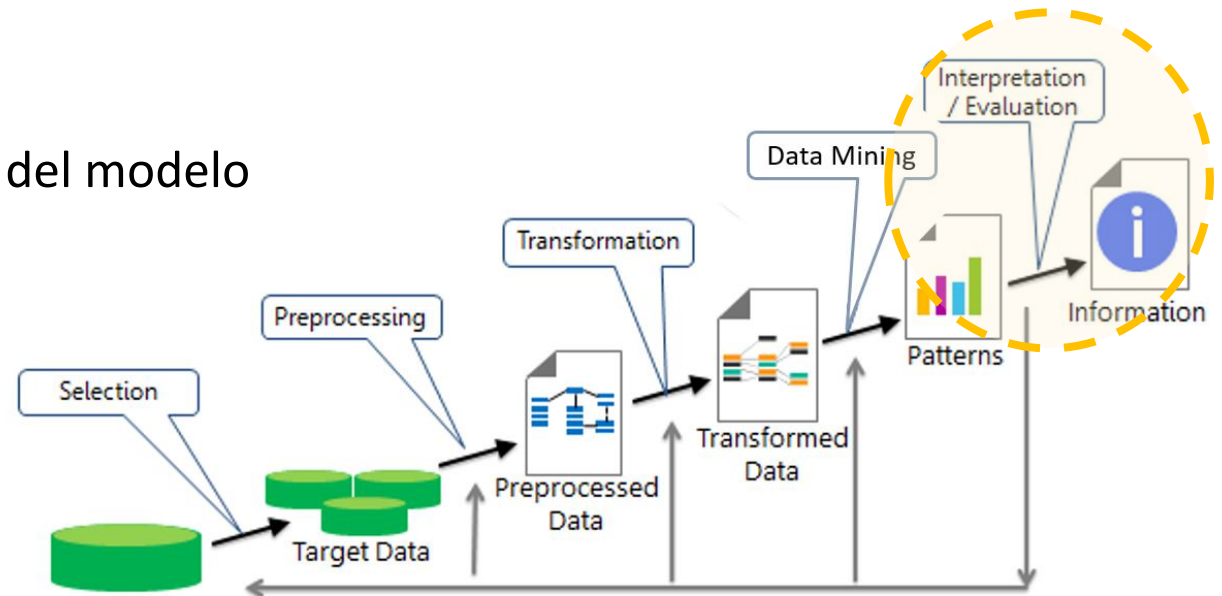


Based on content in "From Data Mining to Knowledge Discovery", AI Magazine, Vol 17, No. 3 (1996)
<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>

KDD: Knowledge Discovery in Databases process

5. INTERPRETACION/EVALUACIÓN:

- Implementación, interpretación o difusión del modelo

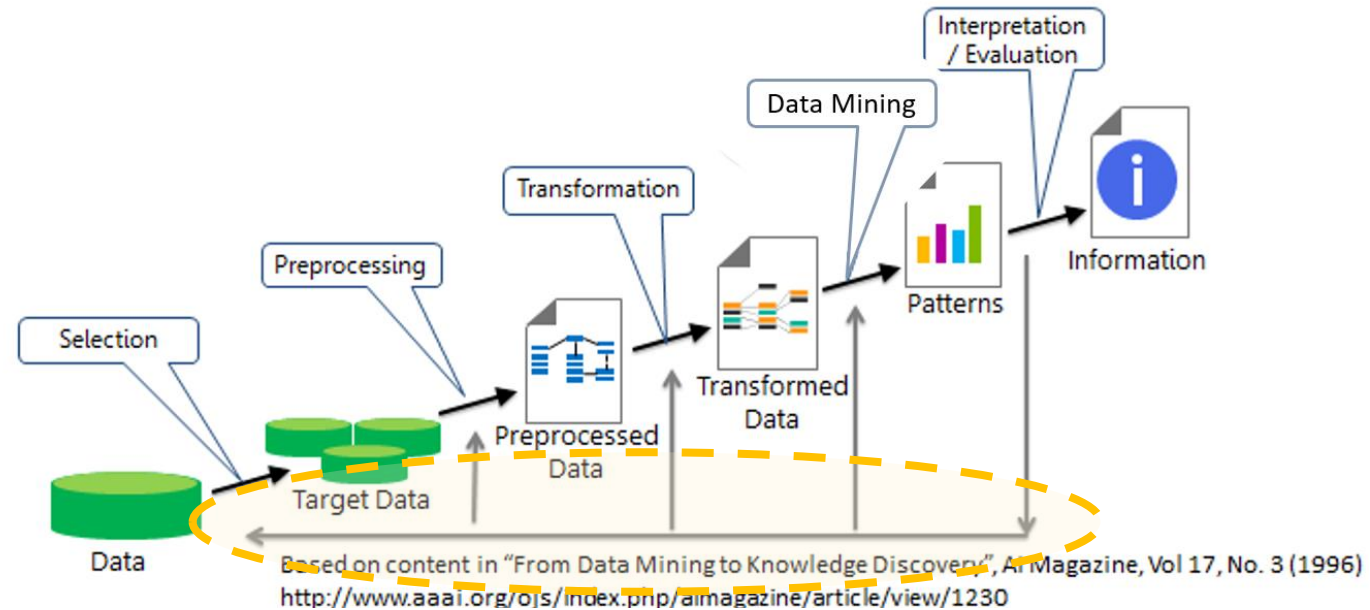


Based on content in "From Data Mining to Knowledge Discovery", AI Magazine, Vol 17, No. 3 (1996)
<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>

KDD: Knowledge Discovery in Databases process

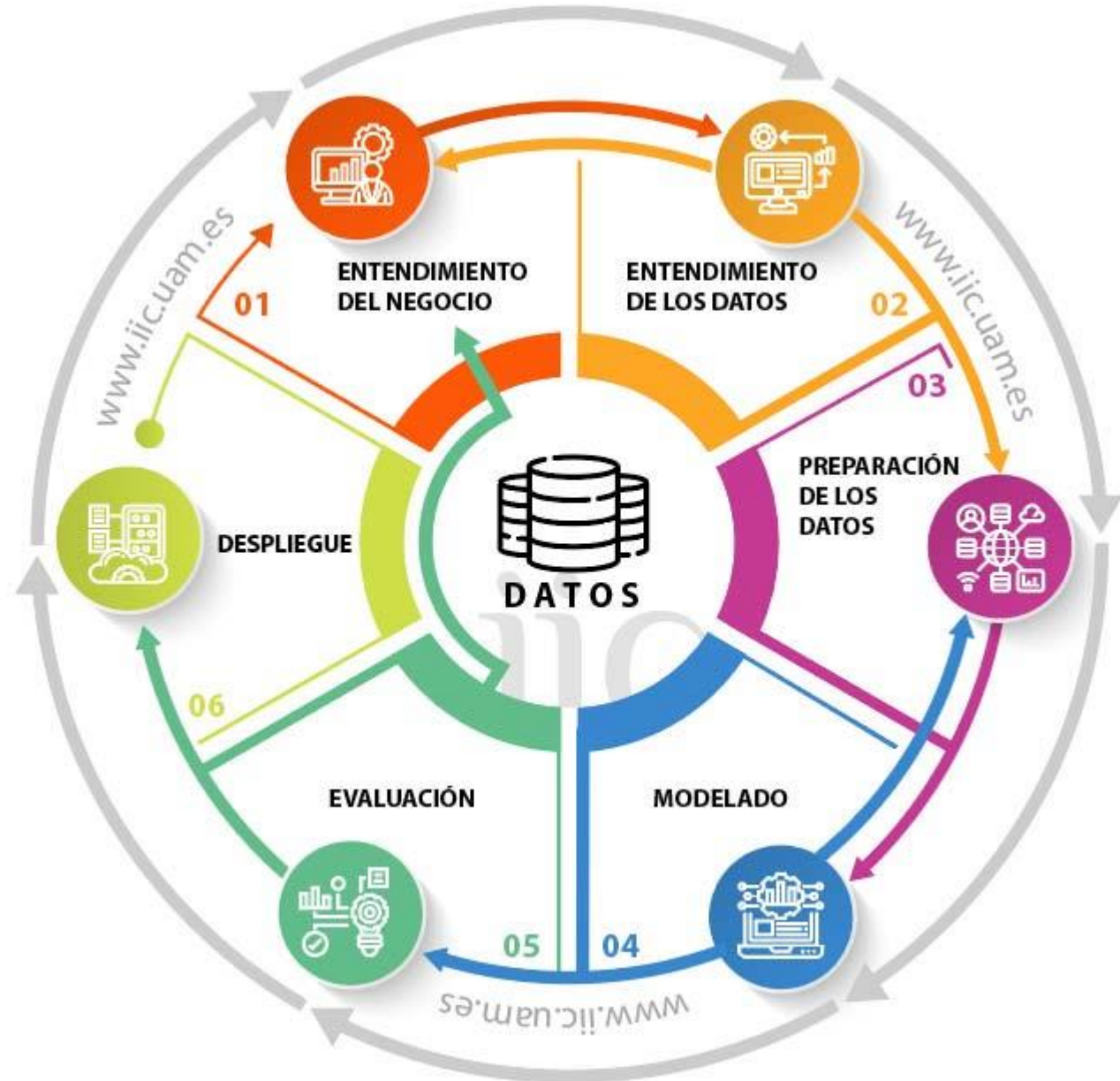
6. ACTUALIZACIÓN Y MONITORIZACIÓN

- Consiste en ir revalidando el modelo con cierta frecuencia sobre nuevos datos, con el objetivo de detectar si el modelo requiere una actualización

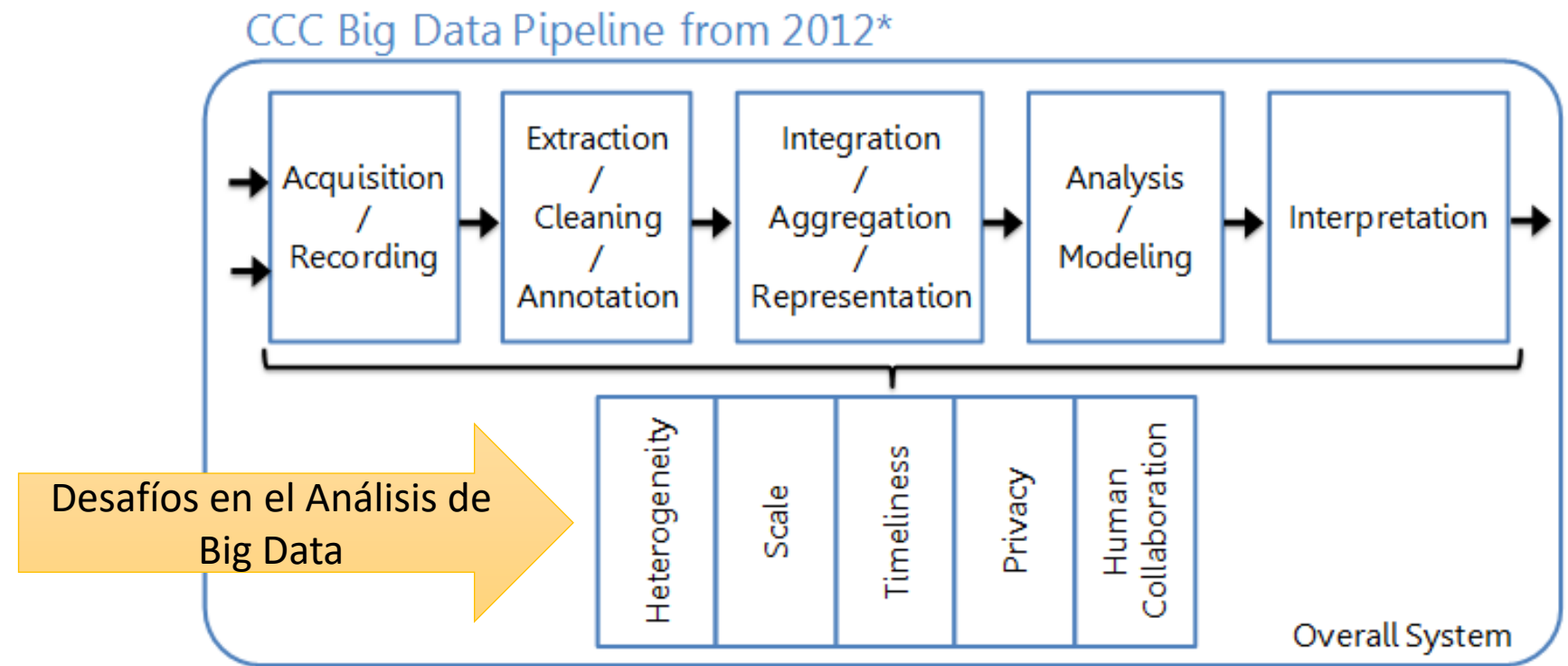


CRISP DM

Cross Industry Standard
Process for Data Mining



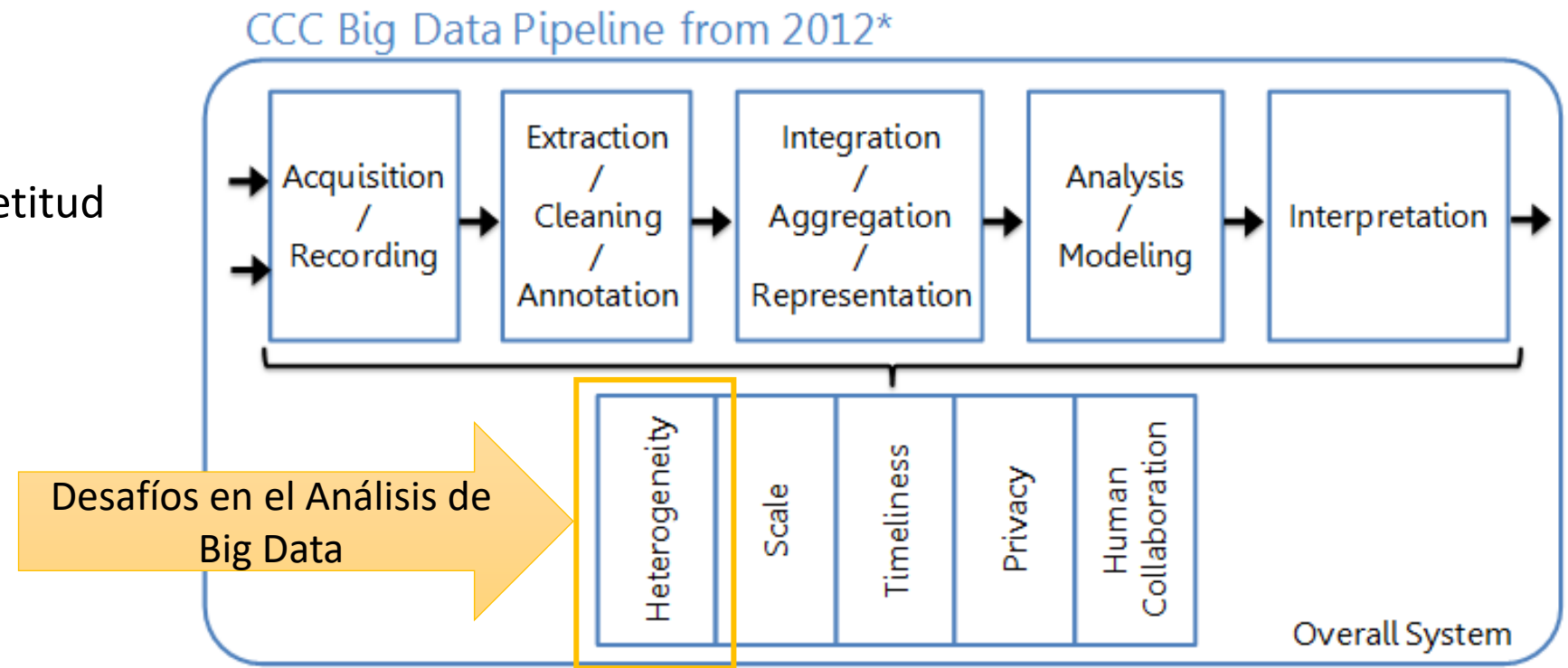
Big Data Analysis Pipeline



*From the Computing Community Consortium Big Data Whitepaper: <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>

Big Data Analysis Pipeline

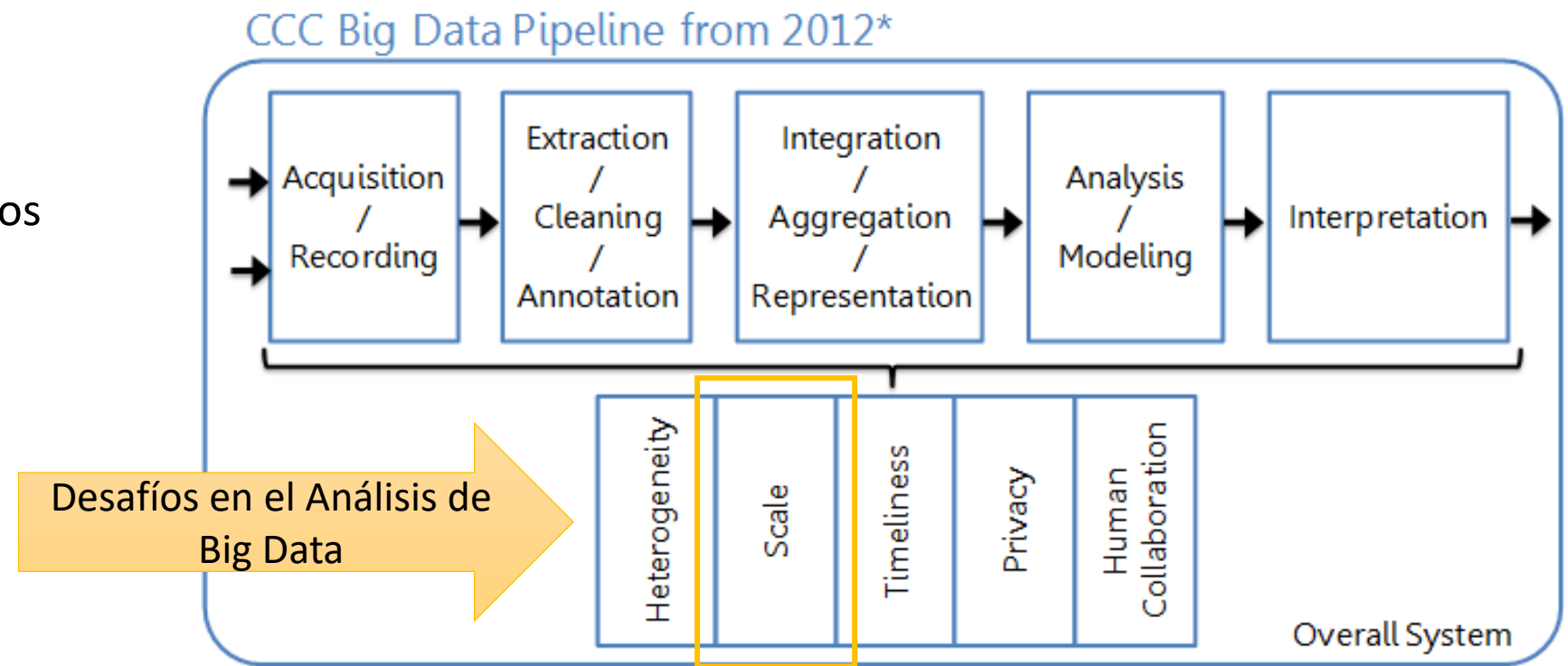
1. Heterogeneidad e incompletitud



*From the Computing Community Consortium Big Data Whitepaper: <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>

Big Data Analysis Pipeline

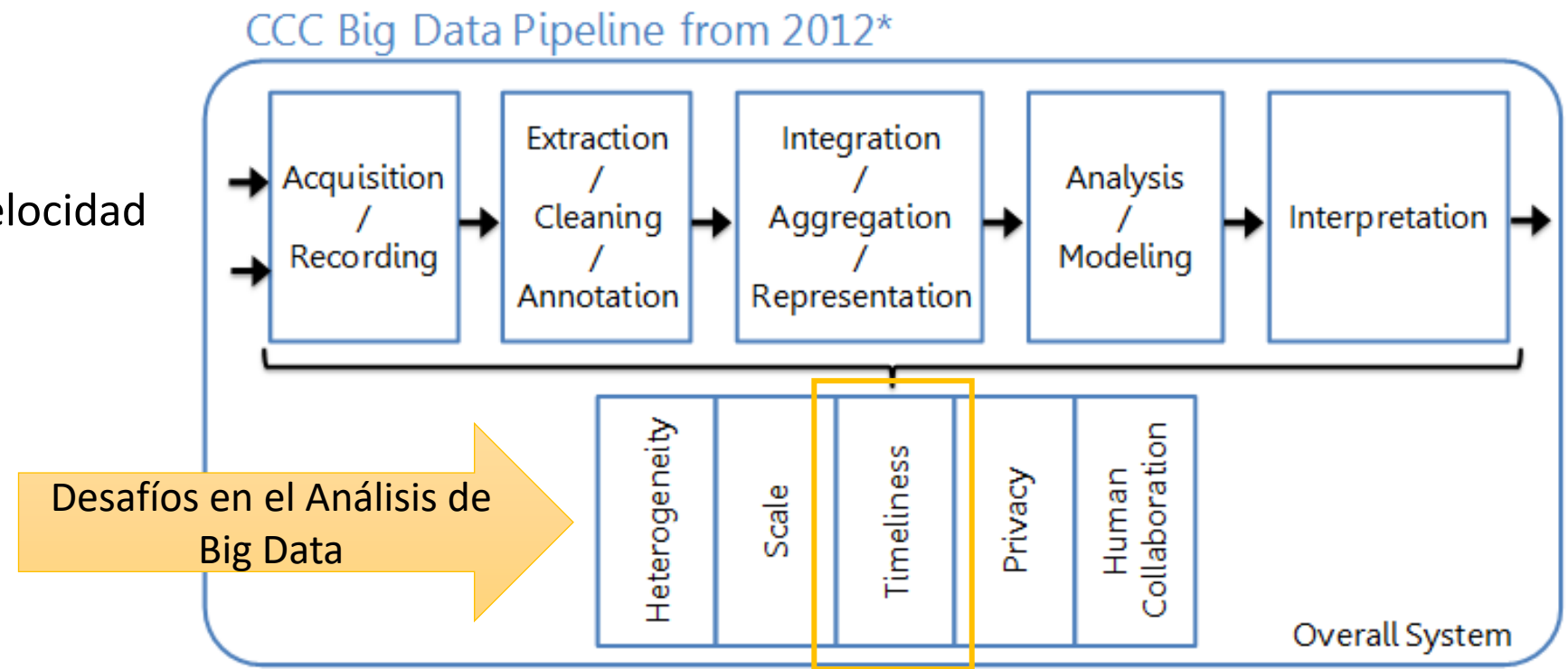
2. Escalar: Volumen de datos



*From the Computing Community Consortium Big Data Whitepaper: <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>

Big Data Analysis Pipeline

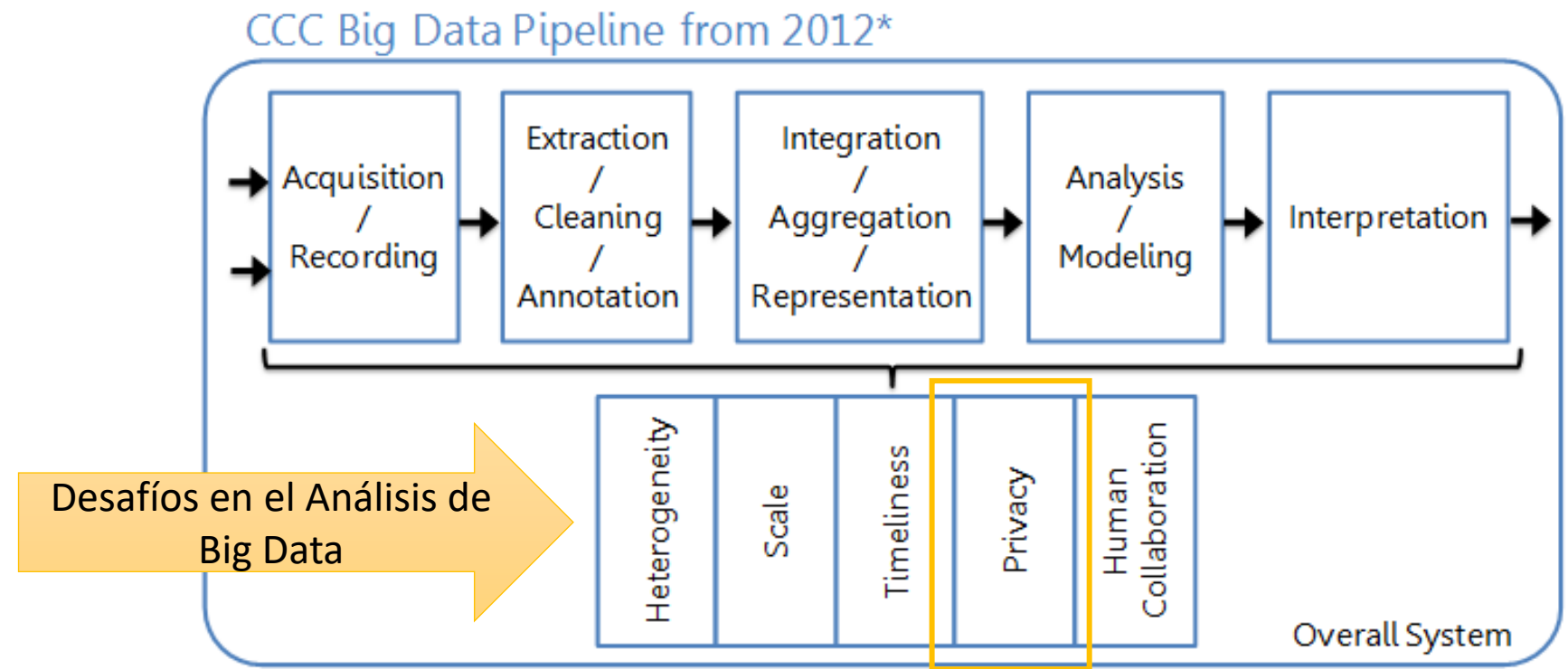
3. Puntualidad (timeliness): Velocidad



*From the Computing Community Consortium Big Data Whitepaper: <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>

Big Data Analysis Pipeline

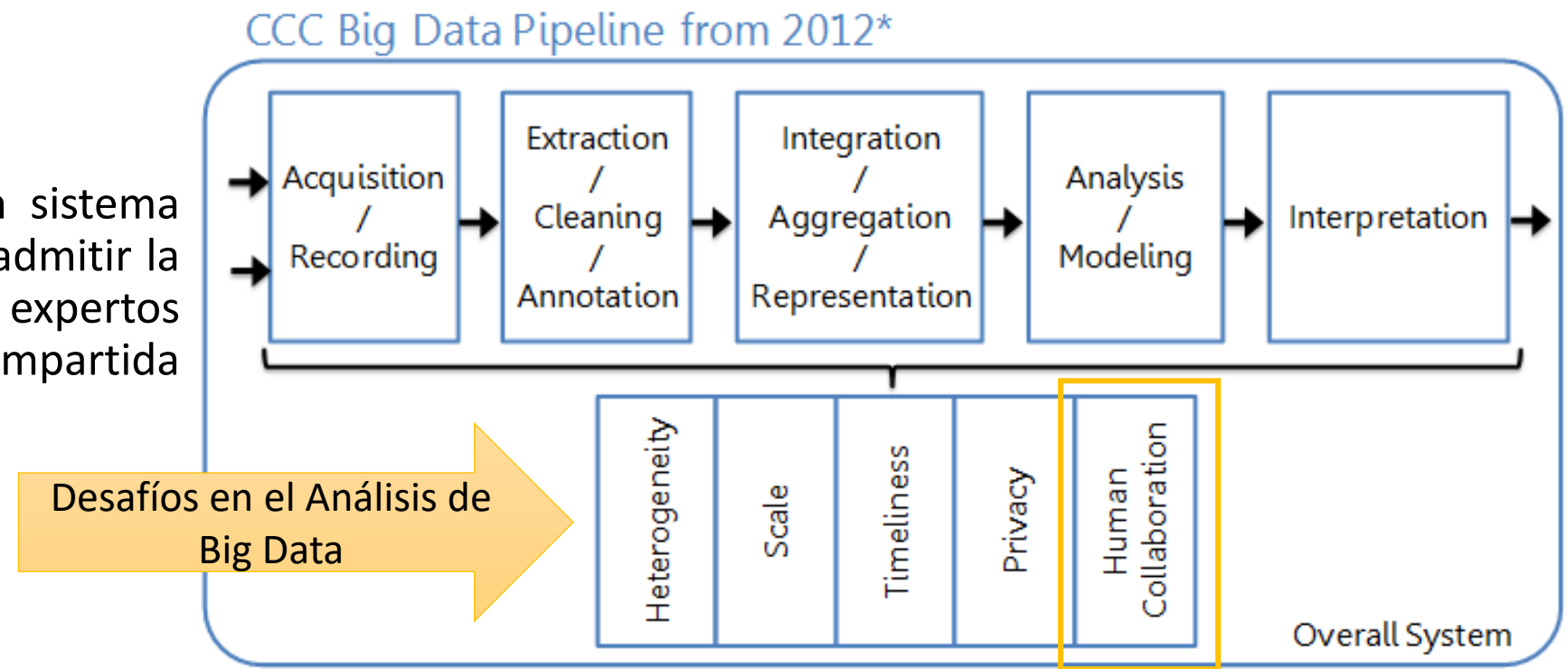
4. Privacidad



*From the Computing Community Consortium Big Data Whitepaper: <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>

Big Data Analysis Pipeline

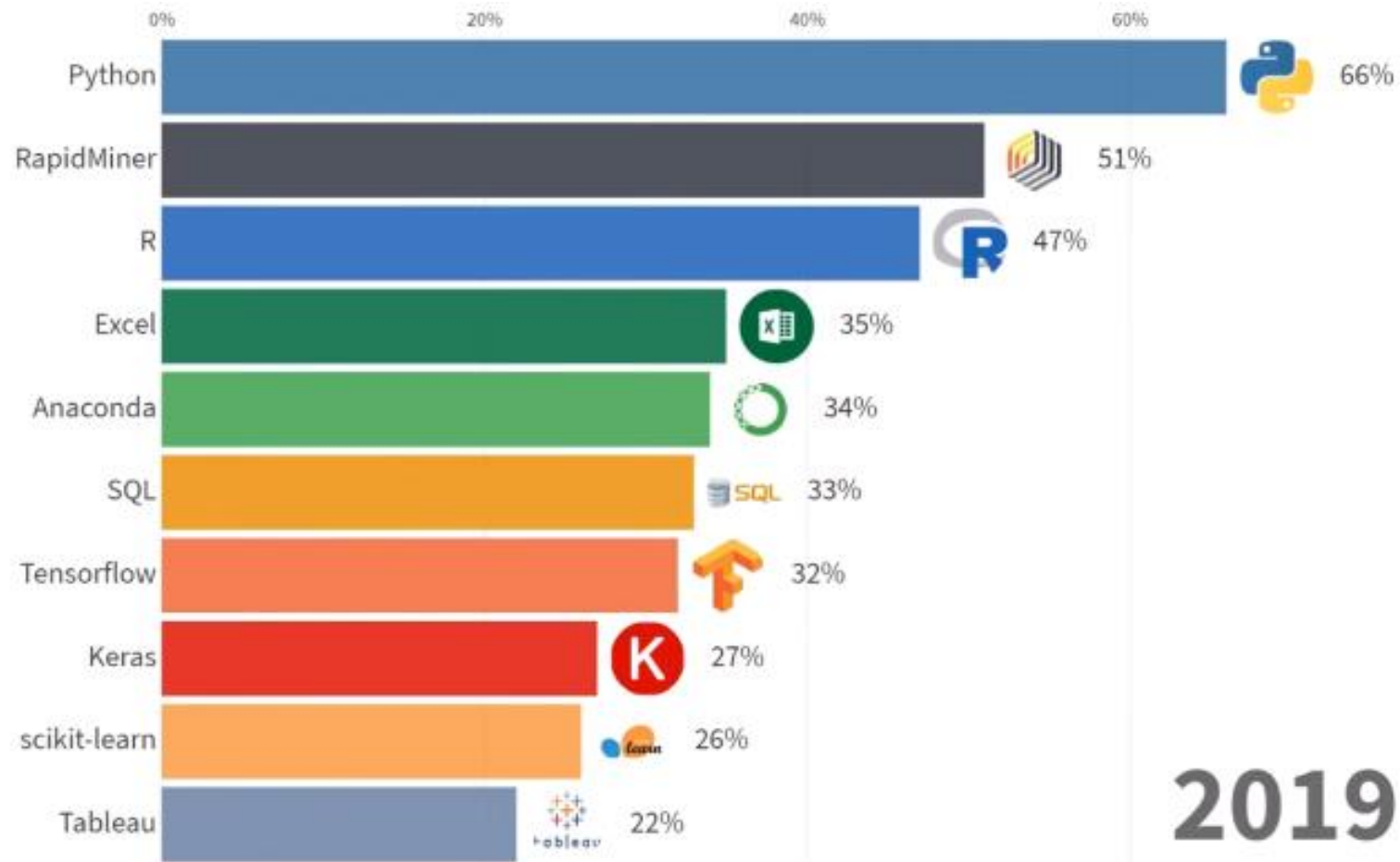
5. Colaboración humana: Un sistema de análisis de Big Data debe admitir la aportación de múltiples expertos humanos y la exploración compartida de resultados.



*From the Computing Community Consortium Big Data Whitepaper: <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>

Herramientas

Herramientas Data Science



Herramientas

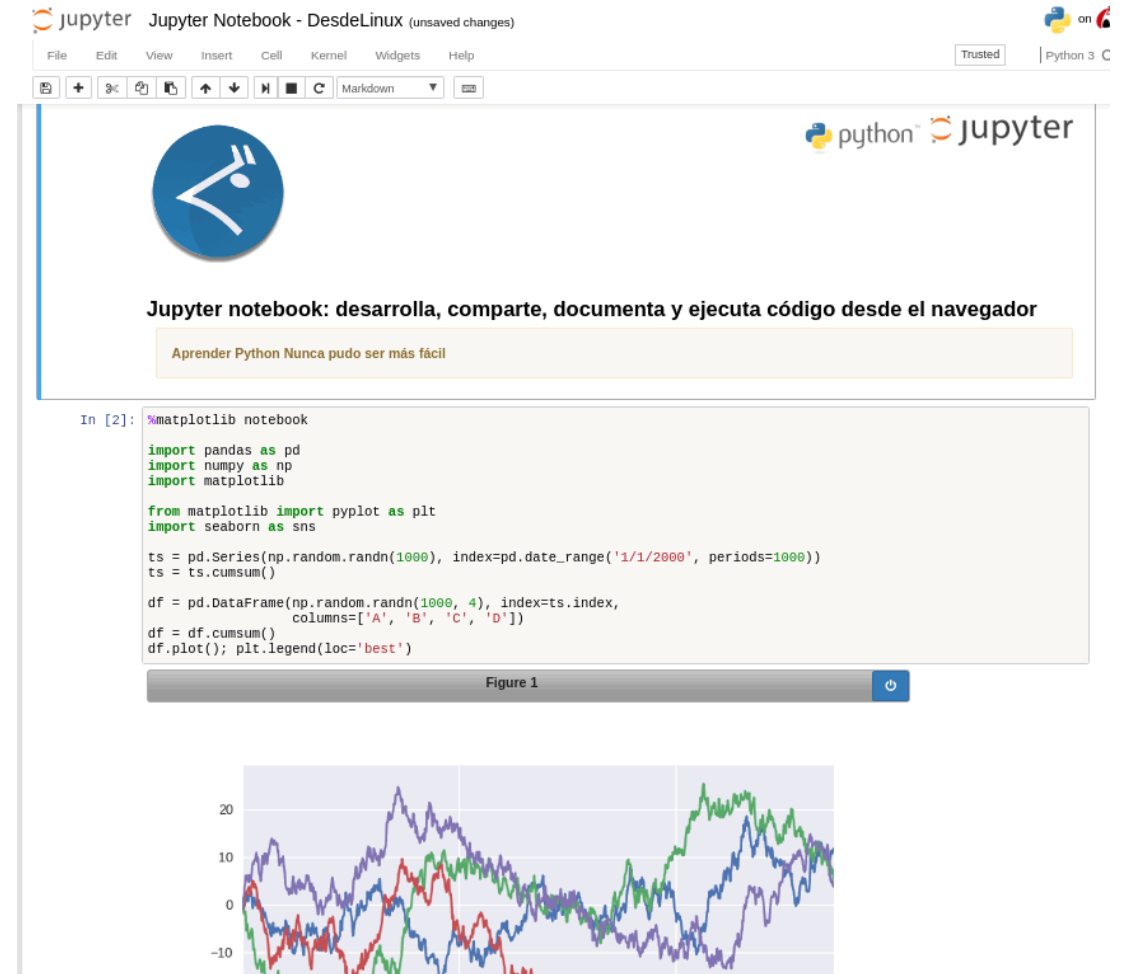


Es un **lenguaje bastante eficiente con una sintaxis bastante amigable** (No es muy complejo)

Cada vez **se está usando más este lenguaje**. Tanto por informáticos como por científicos e ingenieros.

Las numerosas librerías creadas para el análisis de datos son:

- Numpy y Pandas que implementan funciones para cálculos matemáticos y estadísticos
- Mlpy con algoritmos de aprendizaje máquina
- Matplotlib que permite la visualización y representación gráfica de los datos.



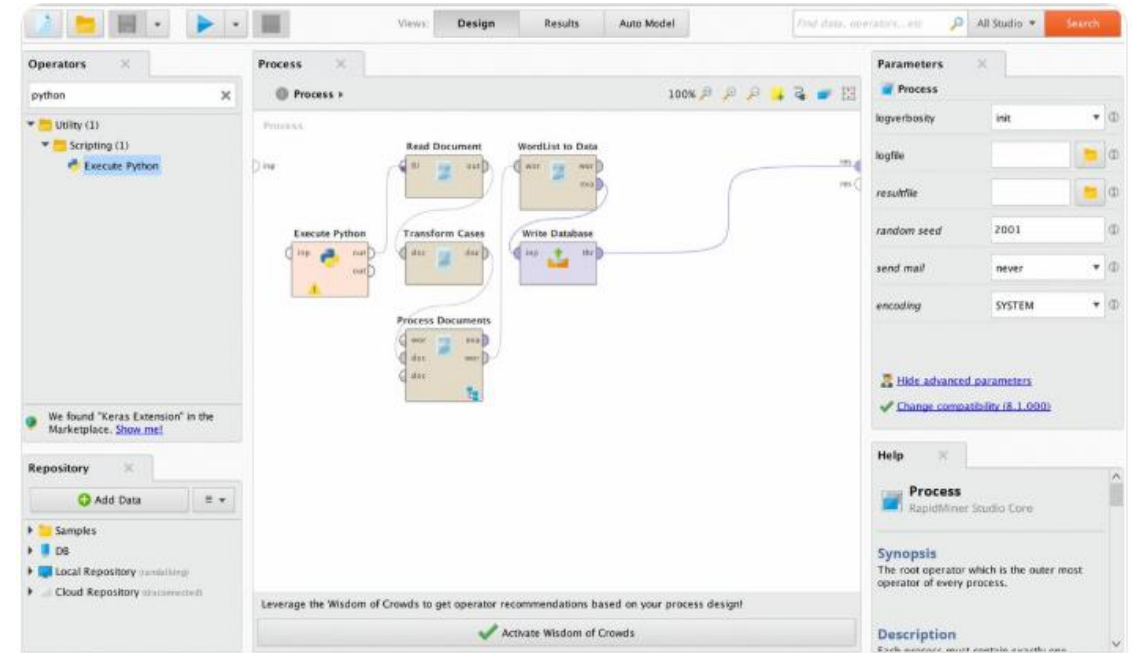
Su integración con aplicaciones como MongoDB, Hadoop o Pentaho.

Herramientas



La herramienta está formada por diferentes módulos entre ellos: RapidMiner Studio, RapidMinder Server, RapidMiner Radoop, RapidMiner Real-Time Scoring, cada uno encargado de una técnica diferente de minería de datos

El punto fuerte de RapidMiner, si se compara con el resto de software de data mining, reside en los **análisis predictivos**



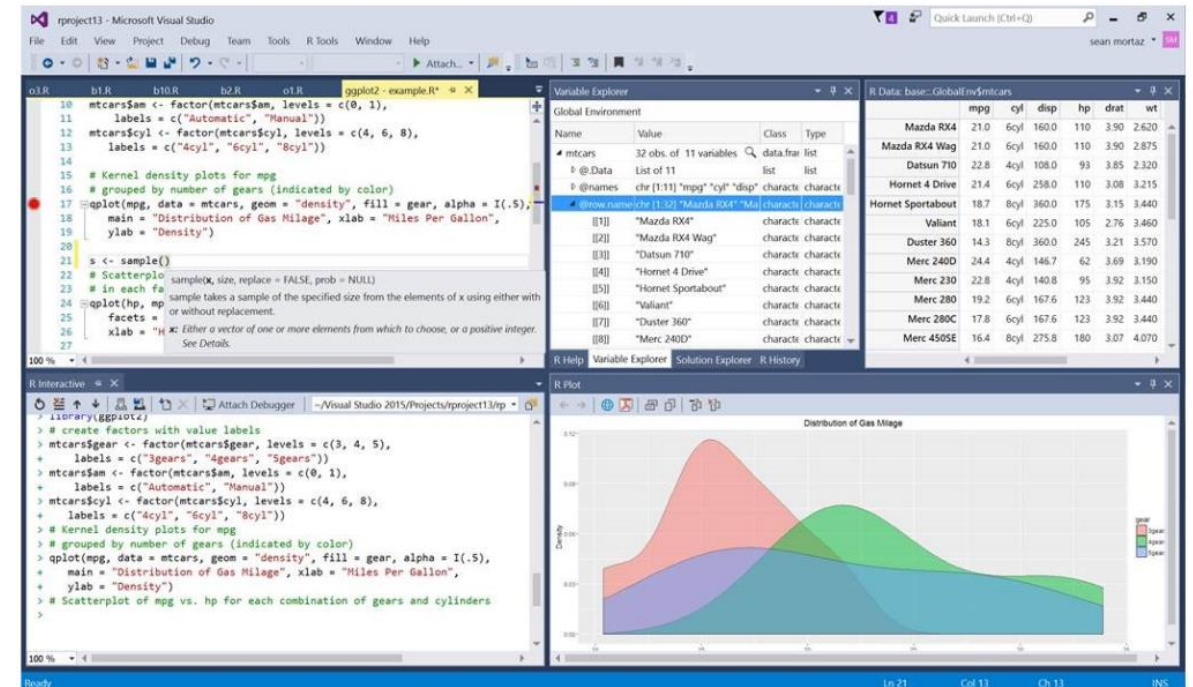
Herramientas



R es un lenguaje ampliamente usado para el análisis de datos y computación estadística

La inclusión de paquetes poderosos en R lo ha hecho más y más poderoso con el tiempo

R tiene suficientes disposiciones para implementar algoritmos de aprendizaje automático de una manera rápida y sencilla.



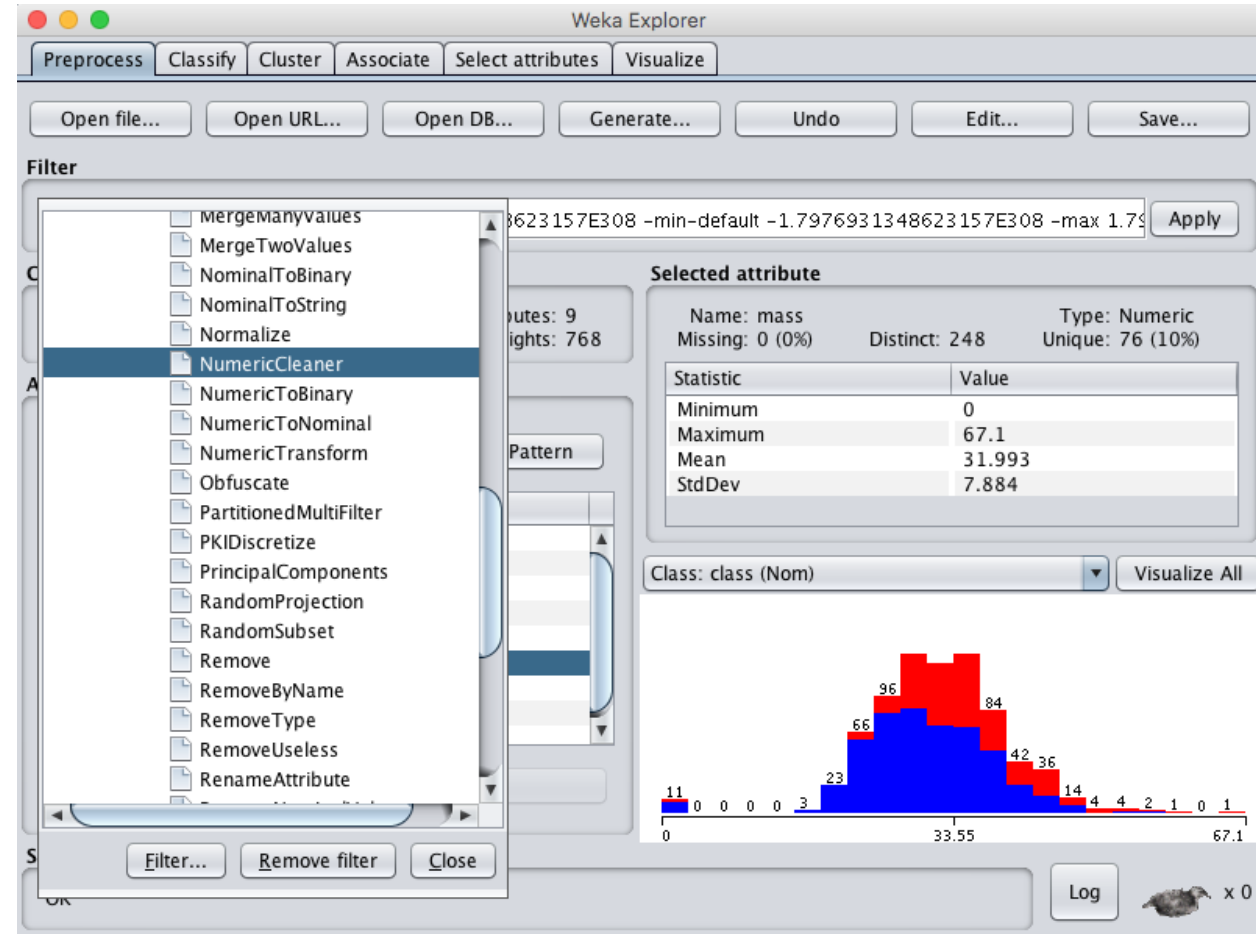
Herramientas



Weka

Es un software de código abierto desarrollado por la Universidad de Waikato.

Presenta un sinnúmero de **funciones** de aprendizaje automático y secunda tareas tan relevantes del data mining como el análisis de clústeres, de correlación o de regresión, así como la clasificación de datos, punto fuerte este último del data mining software al usar redes de neuronas artificiales, árboles de decisión y algoritmos ID3 o C4.5.



Herramientas



Azure Machine Learning

La propuesta de Azure corresponde con un servicio de aprendizaje automático a nivel empresarial para crear e implementar modelos



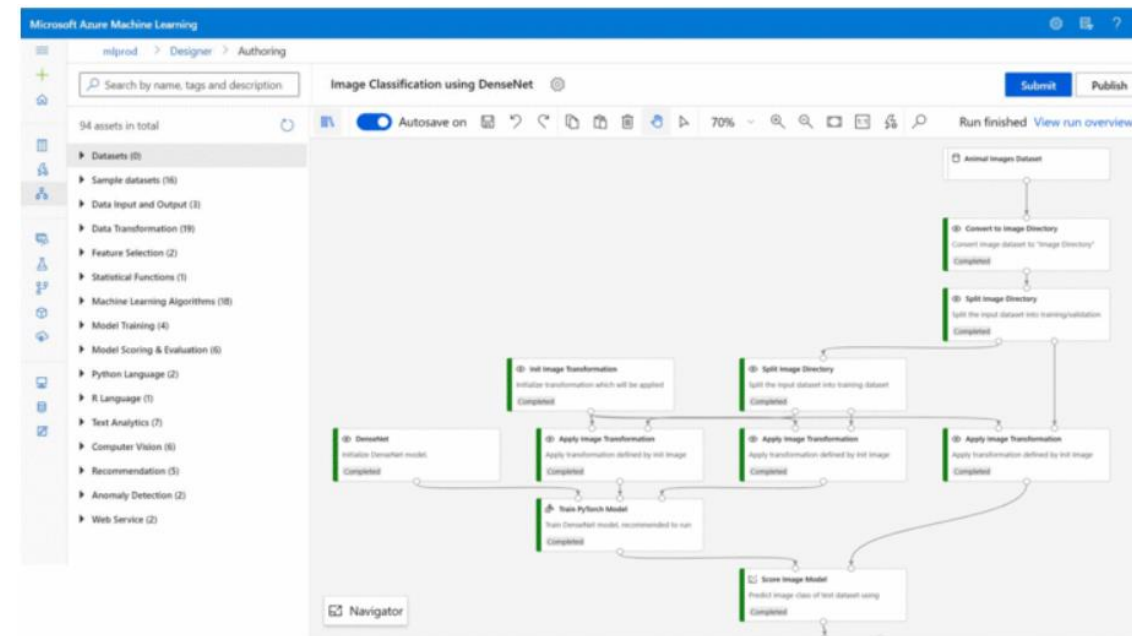
Notebooks



Aprendizaje automático



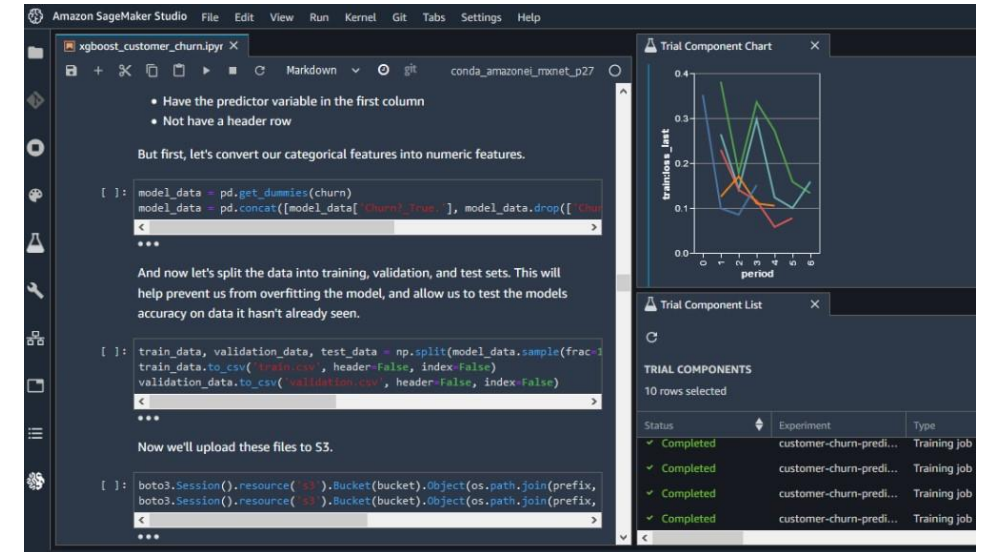
Aprendizaje automático con funcionalidad Drag and Drop



Herramientas



AWS tiene una amplia gama de servicios de Machine Learning, que pueden ser usados tanto por expertos como por gente que apenas sabe programar.



Servicios sin código:

- Amazon Comprehend
- Amazon Translate
- Amazon Transcribe
- Amazon Polly
- Amazon Rekognition
- Amazon Forecast
- Amazon Lex
- Amazon Personalize
- Amazon Textract
- AWS DeepRacer
- AWS DeepLens
- AWS DeepComposer

<https://themachinelearners.com/servicios-machine-learning-aws/>

Relación entre las tareas y las técnicas de aprendizaje

Tareas de aprendizaje

Supervisado

Predictivo

- Regresión
- Clasificación

No supervisado

Descriptivo

- Análisis exploratorio
- Agrupamiento (o Clustering)
- Reducción de la dimensionalidad

- Correlaciones (y dependencias)
- Asociaciones

Técnicas

TÉCNICA	PREDICTIVO / SUPERVISADO		DESCRIPTIVO / NO SUPERVISADO		
	Clasificación	Regresión	Clustering (agrup.)	Reglas asociación	Otros (factoriales, correl, dispersión)
Redes Neuronales	✓	✓	✓ *		
Árboles de Decisión	✓ (c4.5)	✓ (CART)	✓		
Kohonen			✓		
Regresión lineal (local, global), exp..		✓			
Reg. Logística	✓				
Kmeans	✓ *		✓		
A Priori (asociaciones)				✓	
Estudios Factoriales, análisis multivariante			.		✓
CN2 (Covering Algorithm - Basado en reglas)	✓				
K-NN	✓		✓		
RBF Radial-Basis Function	✓				
Bayes Classifiers	✓	✓			