






Entendiendo la Analítica Predictiva

Esta presentación describe, bajo un marco común, los conceptos fundamentales de Analítica Predictiva.

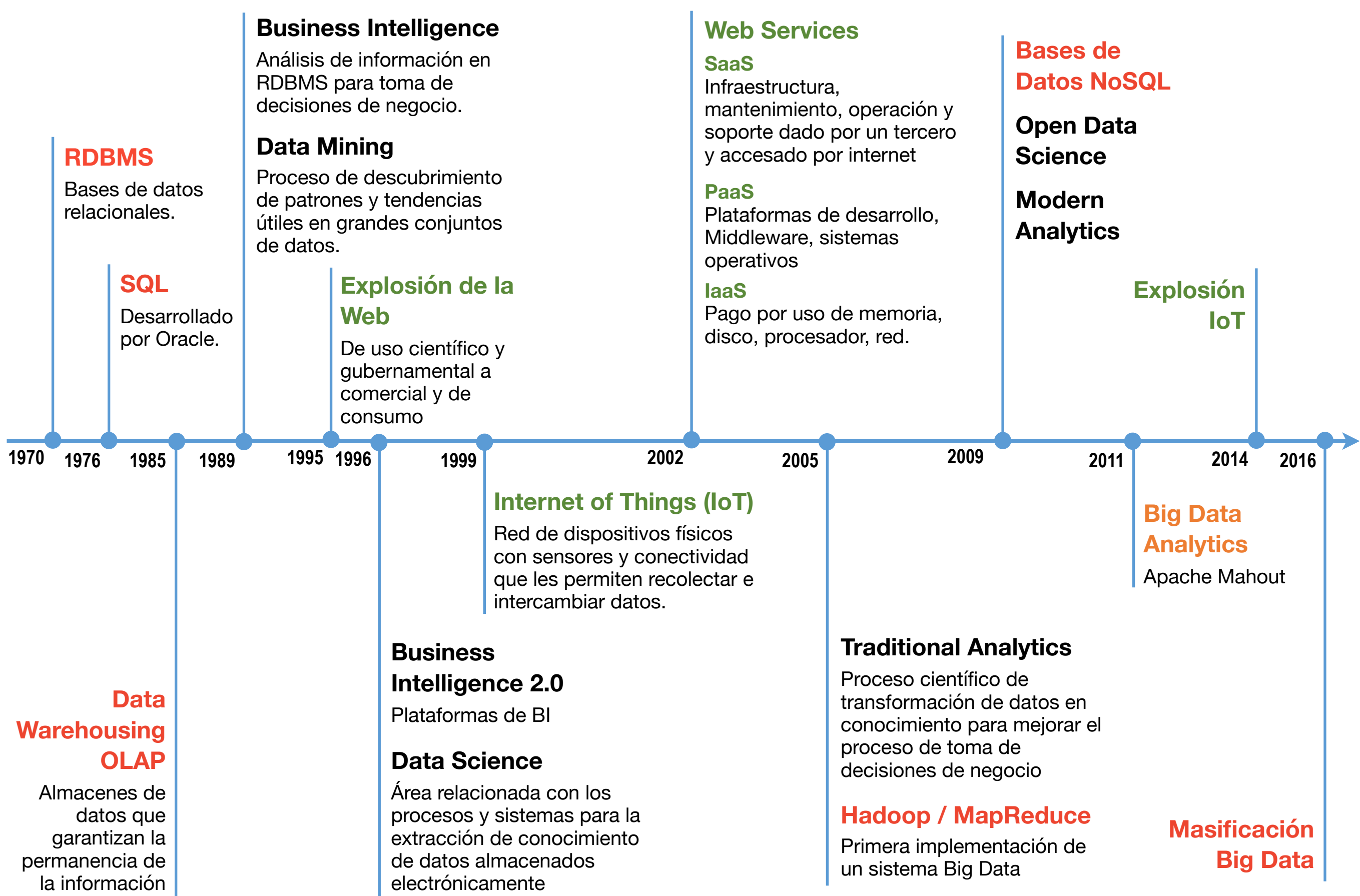
Descargue la última versión de este documento de:
<https://github.com/jdvelasq/Lecture-notes-on-analytics/blob/master/intersoftware-predictive-analytics.pdf>

JUAN DAVID VELÁSQUEZ HENAO, MSc, PhD
Profesor Titular
Departamento de Ciencias de la Computación y la Decisión
Facultad de Minas
Universidad Nacional de Colombia, Sede Medellín

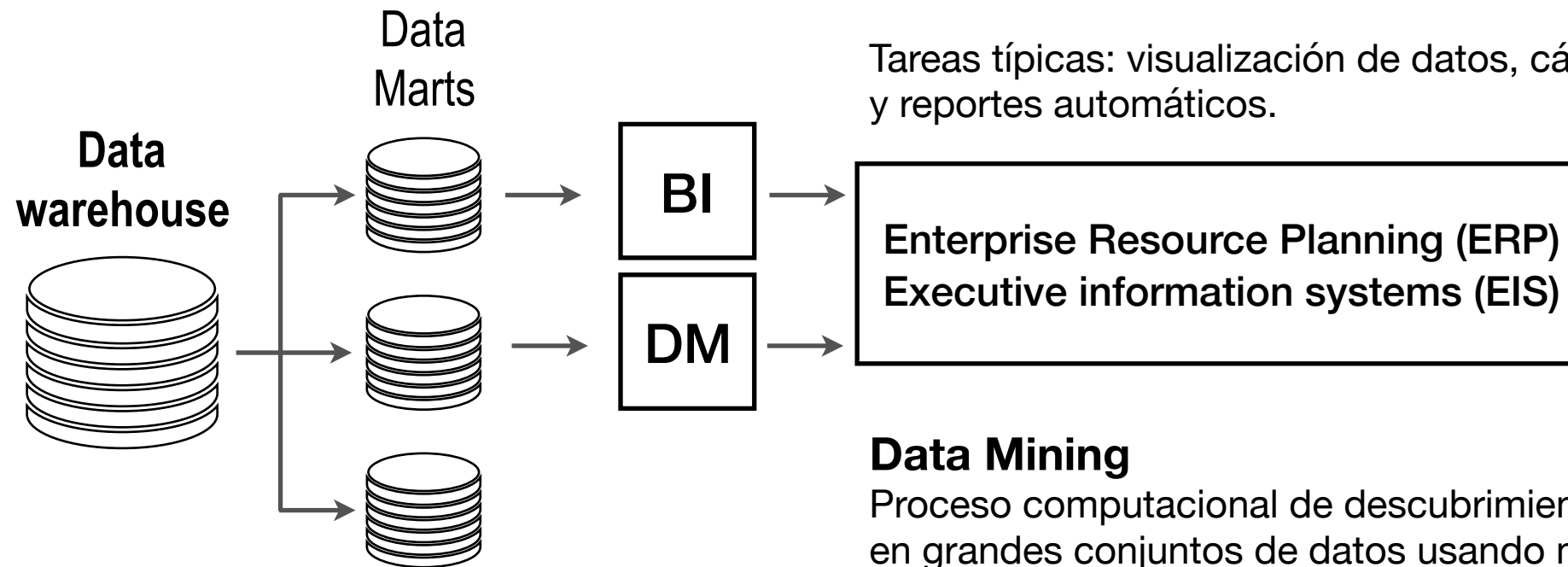
 jdvelasq@unal.edu.co
 [@jdvelasquezh](https://twitter.com/jdvelasquezh)
 <https://github.com/jdvelasq>
 <https://goo.gl/prkjAq>
 <https://goo.gl/vXH8jy>

Big Data &
Data Analytics

Categoría A,
Convocatoria 781 de 2017



Data Mining & Business Intelligence



Inteligencia de Negocios

Conjunto de herramientas, productos y tecnologías para la creación y gestión del conocimiento del medio a partir de los datos disponibles en una organización.

Tareas típicas: visualización de datos, cálculo de indicadores, dashboards y reportes automáticos.

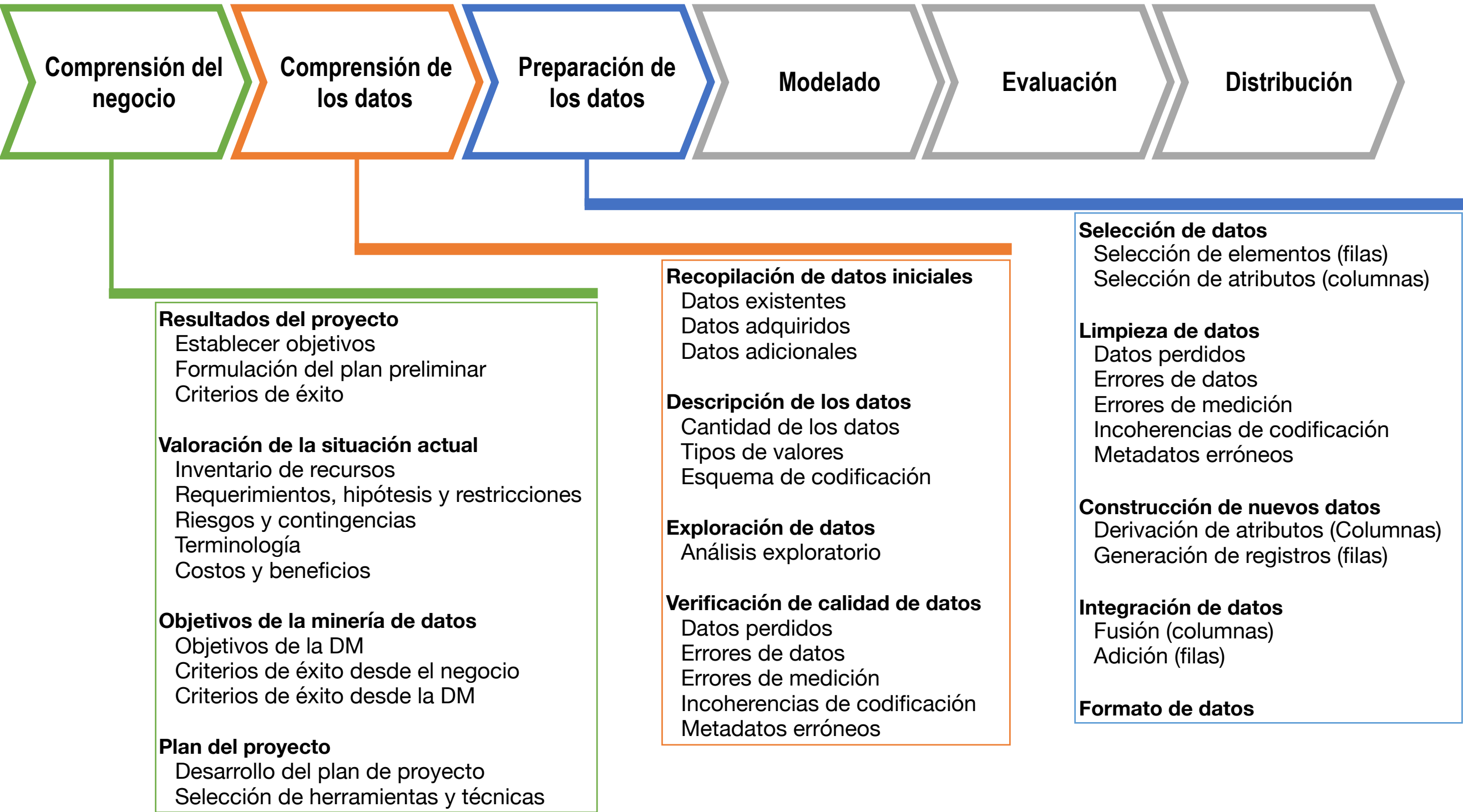
Data Mining

Proceso computacional de descubrimiento de patrones y tendencias útiles en grandes conjuntos de datos usando métodos provenientes de la Estadística, el Aprendizaje de Máquinas y los sistemas de bases de datos.

Tareas típicas: detección de anomalías, modelado de dependencias agrupamiento, clasificación, regresión.

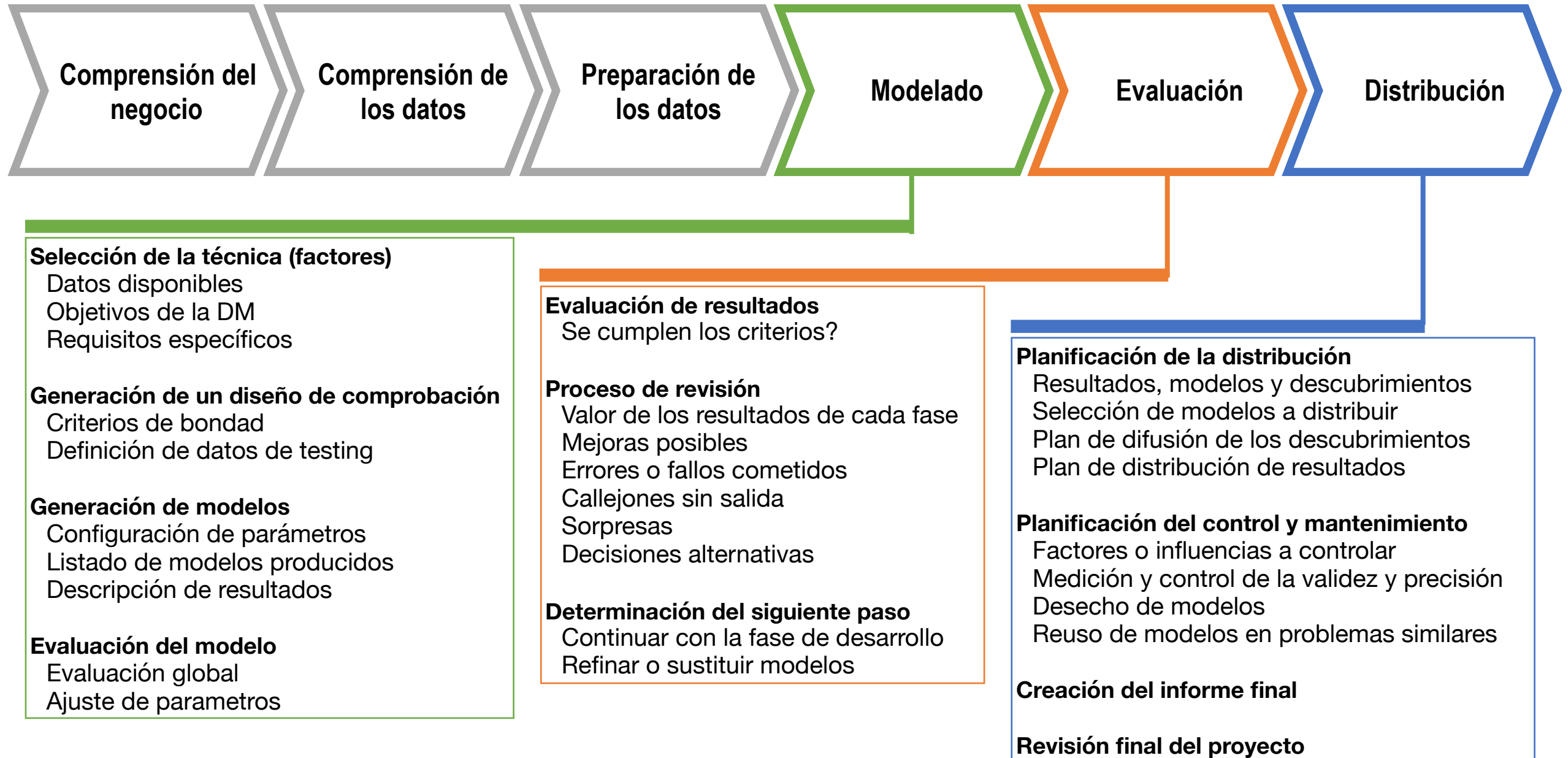
CRISP-DM (1996)

Cross-industry standard process for data mining

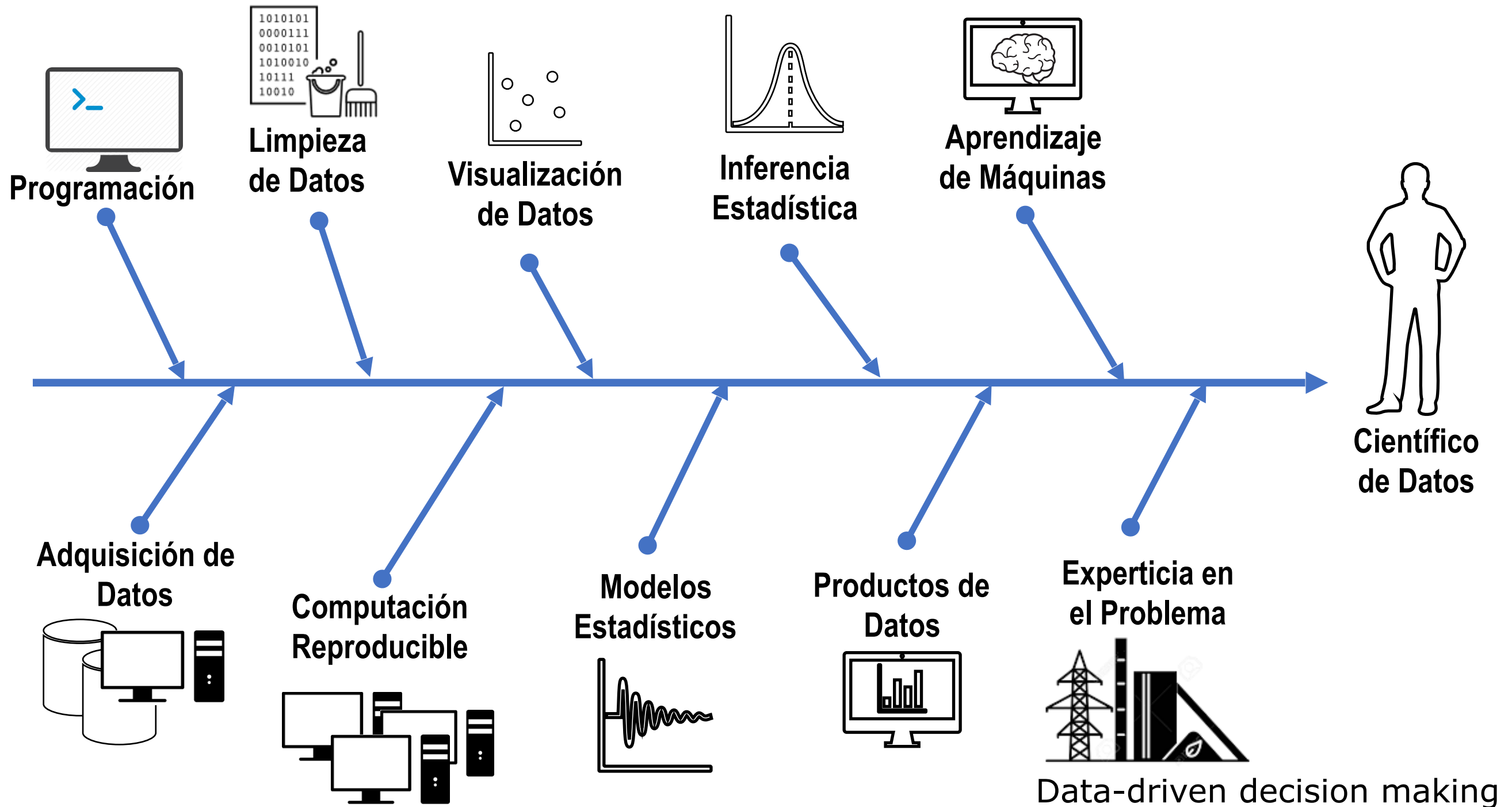


CRISP-DM (1996)

Cross-industry standard process for data mining



Data Science (1996)



Traditional Analytics (2005)



Traditional Analytics (2005)

- **Compresión de la historia.**
- **Pronóstico del futuro.**
- **Los datos están listos.**
- **Sólida fundamentación matemática.**

Aprendizaje Estadístico

Conjunto de herramientas fundamentadas en conceptos estadísticos para la comprensión de conjuntos de datos con el fin de modelar y predecir.

Aprendizaje de Máquinas

Área de las ciencias de la computación orientada al desarrollo de sistemas inteligentes (solución de problemas como lo haría un experto humano).

- **No se quiere comprender que pasó.**
- **Se desea mimificar la inteligencia.**
- **Pronóstico del futuro.**
- **Los datos están listos.**
- **Fundamentación matemática, pero sin el rigor de la estadística.**

Modelado Predictivo

Área enfocada al uso de técnicas estadísticas utilizadas para pronosticar resultados (de un proceso).

- **No se quiere comprender que pasó.**
- **Pronóstico del futuro.**
- **Los datos están listos.**
- **Sólida fundamentación matemática.**

Minería de Datos

Descubrimiento de patrones y conocimiento de grandes conjuntos de datos usando técnicas estadísticas, aprendizaje de máquinas y herramientas de bases de datos.

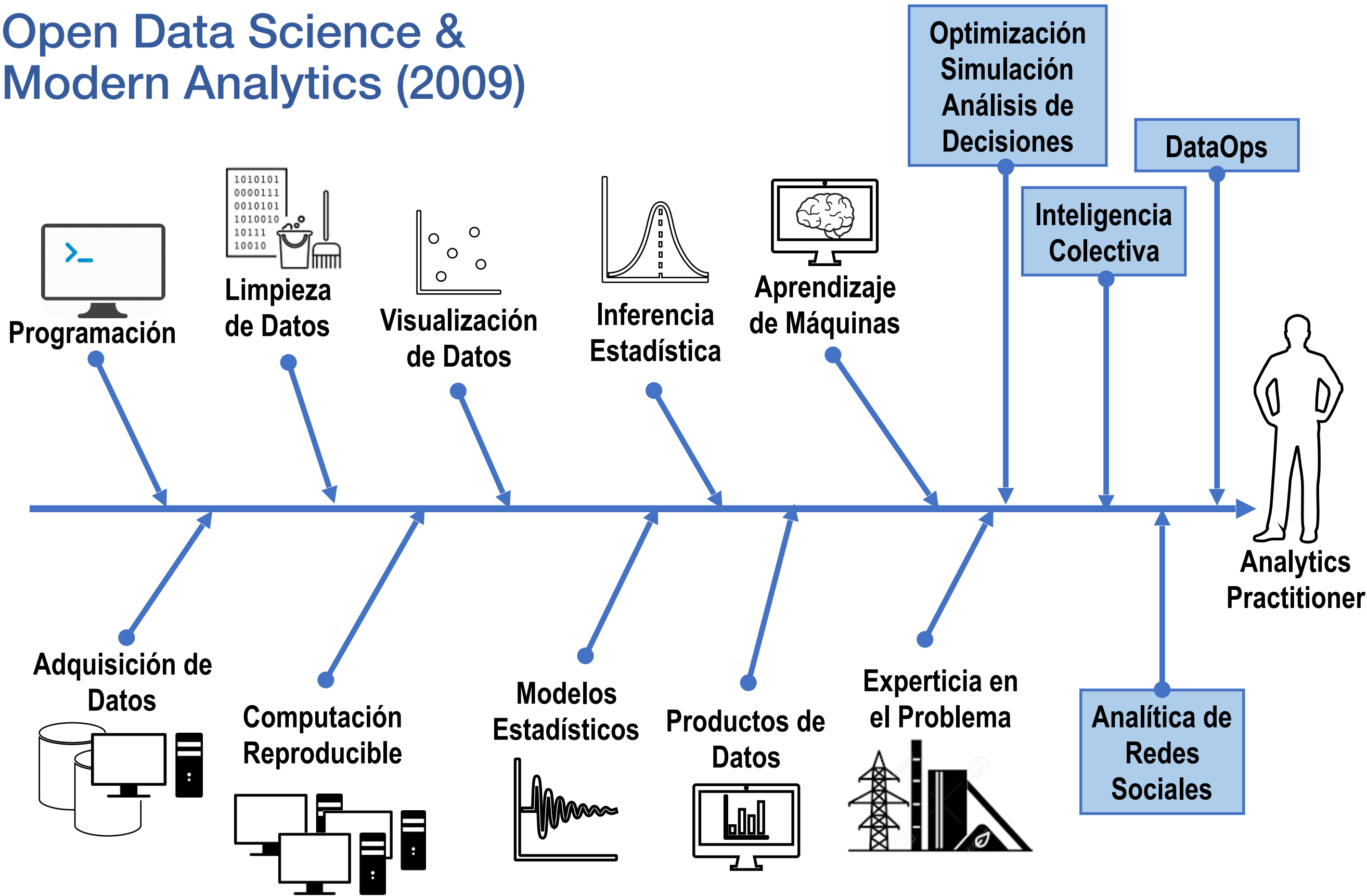
- **No se quiere comprender que pasó.**
- **Pronóstico del futuro.**
- **Los datos están NO están listos.**
- **Sólida fundamentación matemática.**

Analítica Predictiva

Área enfocada al uso de técnicas de modelado predictivo, aprendizaje de máquinas y minería de datos para pronosticar los resultados de un proceso en un contexto organizacional.

- **Pronóstico del futuro.**
- **Los datos están NO están listos.**
- **Agrupar todas las anteriores.**

Open Data Science & Modern Analytics (2009)



Modern Analytics (2009)

FASES / DIMENSIONES

Definición del problema de negocio.	Definición del problema de analítica.	Datos
<p>La habilidad de entender un problema de negocios y determinar cuando el problema es solucionable mediante la analítica.</p> <ul style="list-style-type: none">•Obtener o recibir la definición del problema y los requisitos de usabilidad.•Identificar stakeholders.•Determinar si el problema es solucionable mediante analítica.•Refinar la definición del problema y definir restricciones.•Definir el conjunto inicial de beneficios y costos para el negocio.•Obtener consenso entre los stakeholders sobre la definición del problema.	<p>Habilidad para reformular un problema de negocio en un problema de datos con una solución analítica potencial.</p> <ul style="list-style-type: none">•Reformular el problema en términos de los datos.•Desarrollar un conjunto propuesto de variables explicativas y relaciones con las salidas.•Establecer el conjunto de supuestos relacionados con el problema.•Definir los criterios de éxito.•Realizar el inventario de recursos para la ejecución del proyecto.•Obtener consenso entre los stakeholders.	<p>Habilidad de trabajar efectivamente con datos para identificar relaciones potenciales entre variables que ayudarán a refinar la formulación del problema en términos del negocio y de la analítica.</p> <ul style="list-style-type: none">•Identificar, priorizar los datos necesarios y sus fuentes.•Adquirir los datos.•Armonizar, escalar, limpiar y compartir datos.•Construir, integrar y formatear los datos.•Análizar los datos e identificar relaciones entre los datos.•Verificar la calidad de los datos.•Documentar y reportar hallazgos.•Refinar la formulación del problema en términos del negocio y de la analítica.

Modern Analytics (2009)

FASES / DIMENSIONES

Selección de la metodología.	Desarrollo del modelo	Despliegue del modelo	Gestión del ciclo de vida del modelo
<p>Habilidad para identificar y seleccionar aproximaciones potenciales para resolver el problema de negocio</p> <ul style="list-style-type: none">•Identificar las aproximaciones disponibles para la solución del problema.•Seleccionar las herramientas de software.•Definir los métodos para probar los modelos.•Refinar el análisis de costos y beneficios•Desarrollar un plan inicial del proyecto.•Seleccionar las aproximaciones a utilizar.	<p>Habilidad para identificar y construir modelos efectivos para ayudar a resolver el problema de negocio.</p> <ul style="list-style-type: none">•Identificar las estructuras del o los modelos.•Correr y evaluar modelos.•Calibrar modelos y datos.•Integrar modelos.•Documentar y reportar hallazgos.	<p>Habilidad para realizar el despliegue del modelo que ayuda a solucionar el problema de negocio.</p> <ul style="list-style-type: none">•Evaluación del modelo en términos del negocio.•Publicar reportes con hallazgos.•Desarrollar los requerimientos del modelo, del sistema y de usabilidad para producción.•Despliegue del modelo/ sistema en producción.•Soportar el desarrollo	<p>Habilidad para gestionar el ciclo de vida con el fin de evaluar los beneficios del modelo para el negocio sobre el tiempo.</p> <ul style="list-style-type: none">•Documentar la estructura del modelo.•Evaluar permanentemente la calidad del modelo•Recalibrar y mantener el modelo.•Desarrollar actividades de entrenamiento.•Evaluar periódicamente los beneficios del modelo.

Problemas abordados en la analítica predictiva

- Clasificación de patrones
- Agrupamiento
- Asociación
- Regresión
- Series de tiempo.

Problemas detectados en el medio






- Hay fuertes debilidades en los curriculums de pregrado y postgrado.
- No hay pregrados que den formación integral.
- Fuertes barreras de entrada a nivel conceptual y de herramientas.
- Falta de reconocimiento de la profesión.
- Las organizaciones son usuarios de herramientas y no desarrolladores de algoritmos.
- Se terminan adaptando productos genéricos.
- Falta de reconocimiento del producto de datos.
- Las empresas deben formar personal sobre la marcha.
- No es claro como medir el ROI en productos de analítica.
- Vendedores y Clientes sin sociedad estratégica.

Entendiendo la Analítica Predictiva

Esta presentación describe, bajo un marco común, los conceptos fundamentales de Analítica Predictiva.

Descargue la última versión de este documento de:
<https://github.com/jdvelasq/Lecture-notes-on-analytics/blob/master/intersoftware-predictive-analytics.pdf>

JUAN DAVID VELÁSQUEZ HENAO, MSc, PhD
Profesor Titular
Departamento de Ciencias de la Computación y la Decisión
Facultad de Minas
Universidad Nacional de Colombia, Sede Medellín

 jdvelasq@unal.edu.co
 [@jdvelasquezh](https://twitter.com/jdvelasquezh)
 <https://github.com/jdvelasq>
 <https://goo.gl/prkjAq>
 <https://goo.gl/vXH8jy>

Big Data &
Data Analytics

Categoría A,
Convocatoria 781 de 2017