

INTRODUCCIÓN A CIENCIA DE DATOS

Dr. Mauricio Rosales Rivera



¿QUÉ ES UN CIENTÍFICO DE DATOS?

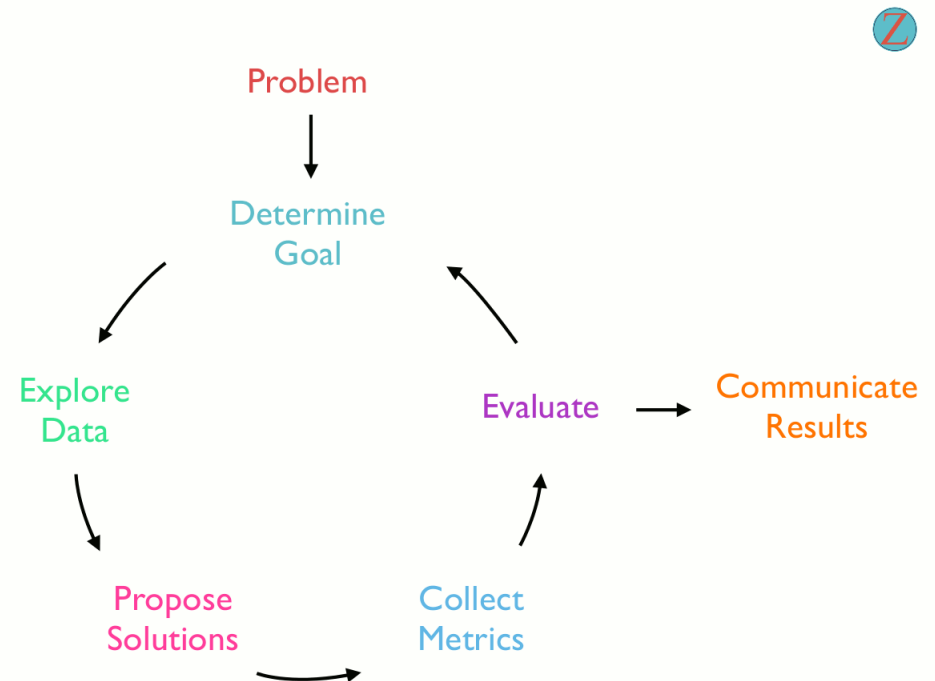
"Científico de datos (n): Persona que sabe más de estadística que cualquier programador y que a la vez sabe más de programación que cualquier estadístico"...

Un poco más en serio, un científico de datos es un profesional dedicado a **analizar e interpretar grandes bases de datos**. O lo que es lo mismo, uno de los profesionales más importantes en cualquier empresa de internet hoy en día.

¿PORQUÉ SE HA PUESTO DE MODA?

La tecnología actual no solo necesita del mejor talento sino de **datos, mucho datos**. Muchos...

Lo que vale para los entornos de *learning*, vale para casi cualquier



¿PORQUÉ SE HA PUESTO DE MODA?

Lo curioso es que este gran valor de los datos contrasta con que precisamente los datos son el recurso más abundante del planeta (se calcula que se crean 2.5 trillones de bytes de información nuevos al día).



¿CÓMO ES POSIBLE QUE ALGO TAN ABUNDANTE SEA TAN VALIOSO?

Acumular datos es algo trivial..... **Lo complejo es procesarlos.**

“La ciencia de datos es la disciplina de hacer que los datos sean útiles.”

Un mapa de Ciencia de Datos



Minería de datos
o Analítica

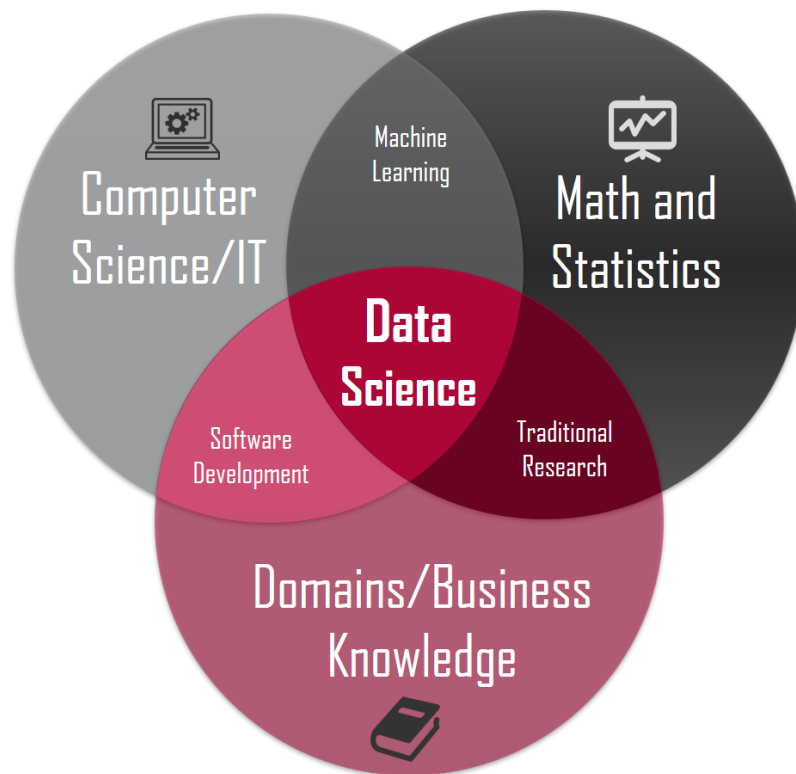


Estadística
inferencial



Machine
Learning

LA CIENCIA DE DATOS ES LA DISCIPLINA DE **HACER QUE LOS DATOS SEAN ÚTILES**



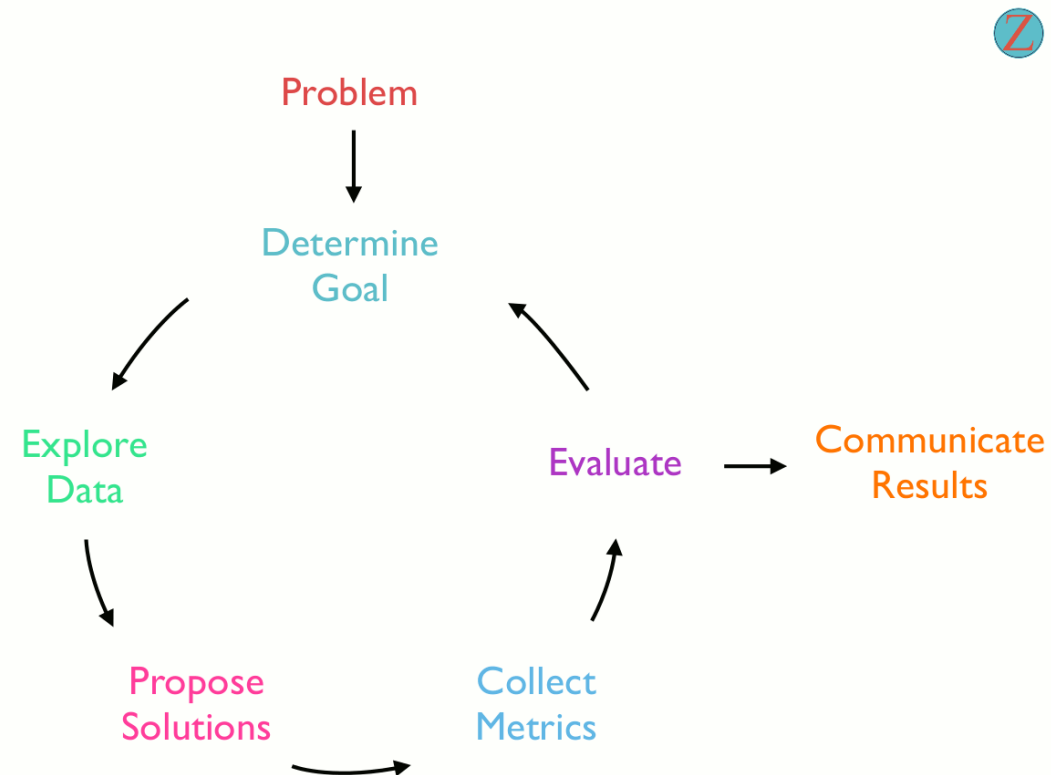
Desde hace unos años nos encontramos ante la primera gran 'democratización' de estas técnicas. Y, con ello, el boom de los científicos de datos: nadie quiere tener una mina de oro sin aprovechar.

Es a través de nuestras acciones—nuestras decisiones—que afectamos el mundo que nos rodea.

LA CIENCIA DE DATOS ES LA DISCIPLINA DE **HACER QUE LOS DATOS SEAN ÚTILES**

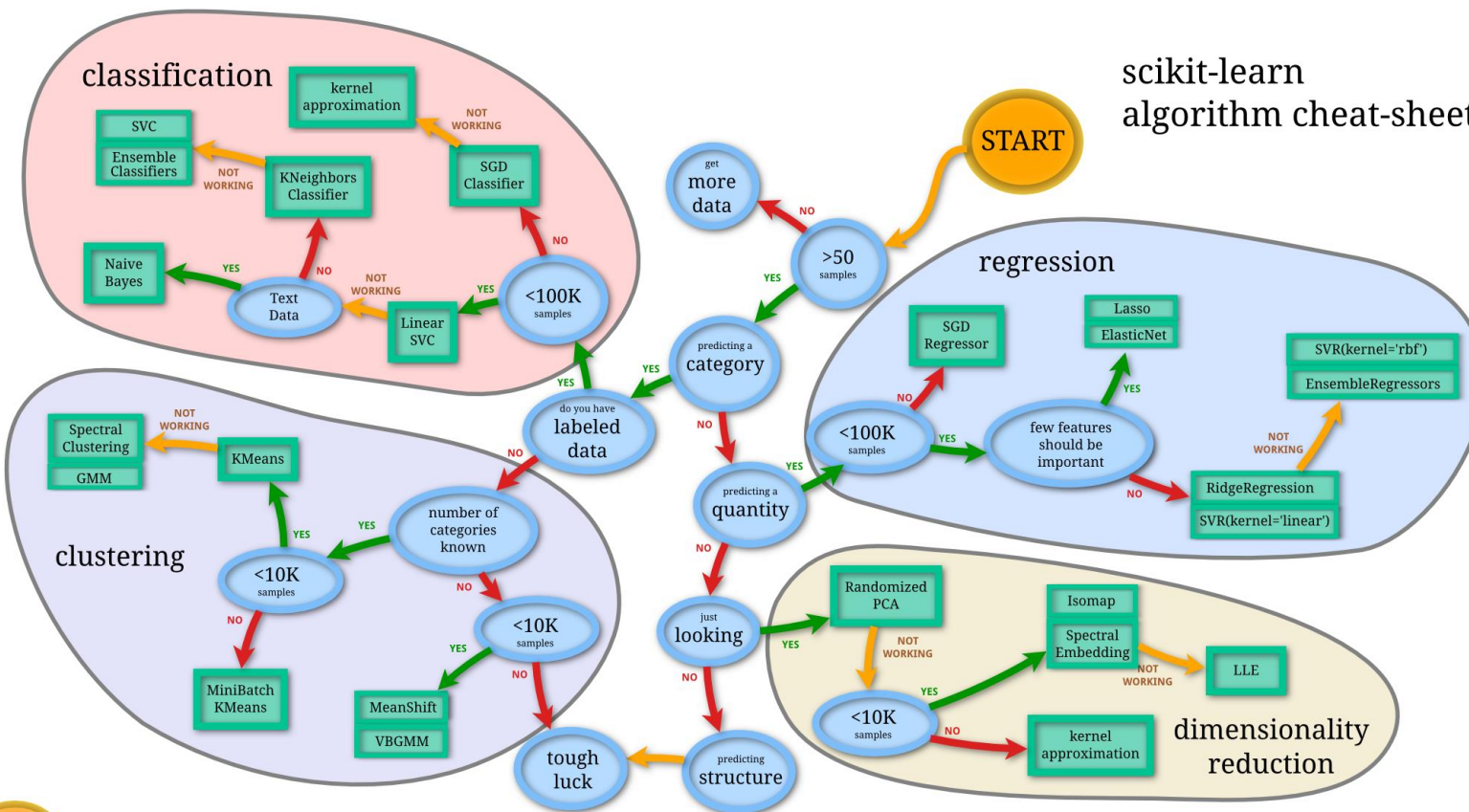


LA CIENCIA DE DATOS ES LA DISCIPLINA DE **HACER QUE LOS DATOS SEAN ÚTILES**



ALGUNAS TÉCNICAS DE *MACHINE LEARNING*





Con el tiempo, se han desarrollado muchos enfoques estadísticos de aprendizaje automático.

Puede usar este [link](#) del equipo de *scikit-learn* como guía para la mayoría métodos populares.

Diplomado en Ciencia de Datos con Python

Introducción al análisis de datos estadístico

19 de marzo de 2020

Introducción al análisis de datos estadístico

Diplomado
en Ciencia
de Datos
con
Python

¿Qué es la
Estadística?

Interpretación
de datos

Objetivos de
Interpretación

Validación
de datos

Análisis
Estadístico
- Pre procesamiento

Pasos para
ML

Objetivos
del
Modulo
Estadístico

La estadística es básicamente una ciencia que implica la recopilación de datos, la interpretación de datos y, finalmente, la validación de datos.

El análisis de datos estadístico es un procedimiento esencial para el entendimiento en la ciencia de datos.

El propósito de analizar e interpretar datos es obtener información utilizable y útil. La interpretación, independientemente de si los datos son cualitativo o cuantitativo, debe:

- 1 describir y resumir los datos
- 2 identificar relaciones entre variables
- 3 comparar variables
- 4 identificar la diferencia entre las variables
- 5 pronósticar resultados

Objetivos de Interpretación

Diplomado
en Ciencia
de Datos
con
Python

¿Qué es la
Estadística?

Interpretación
de datos

Objetivos de
Interpretación

Validación
de datos

Análisis
Estadístico
- Pre procesamiento

Pasos para
ML

Objetivos
del
Módulo
Estadístico

Al analizar los datos, comience siempre desde sus objetivos de investigación, es decir, la razón por la que realizó el análisis en primer lugar.

El análisis e interpretación de los datos, independientemente del método y del estado cualitativo / cuantitativo, pueden incluir las siguientes características:

- 1 Identificación de datos y explicación
- 2 Comparación y contraste de datos
- 3 Identificación de datos atípicos
- 4 Predicciones futuras
- 5 Toma de decisiones
- 6 Identificación de tendencias

¿Importancia?

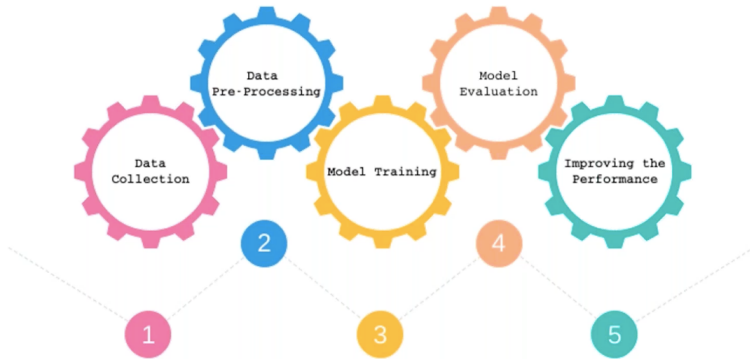
Los datos reales pueden ser impuros, ya que podrían conducir a la extracción de patrones o reglas poco útiles.

Esto se puede deber a:

- 1 Datos incompletos
- 2 Datos con ruido
- 3 Datos inconsistentes

La validación de datos también se conoce como validación de entrada y es parte esencial del procesamiento de datos. La verificación de datos también es un proceso relacionado pero distinto.

General Machine learning algorithm flow



Pasos para desarrollar el Aprendizaje Máquina (ML)

