

Ciencia de Datos

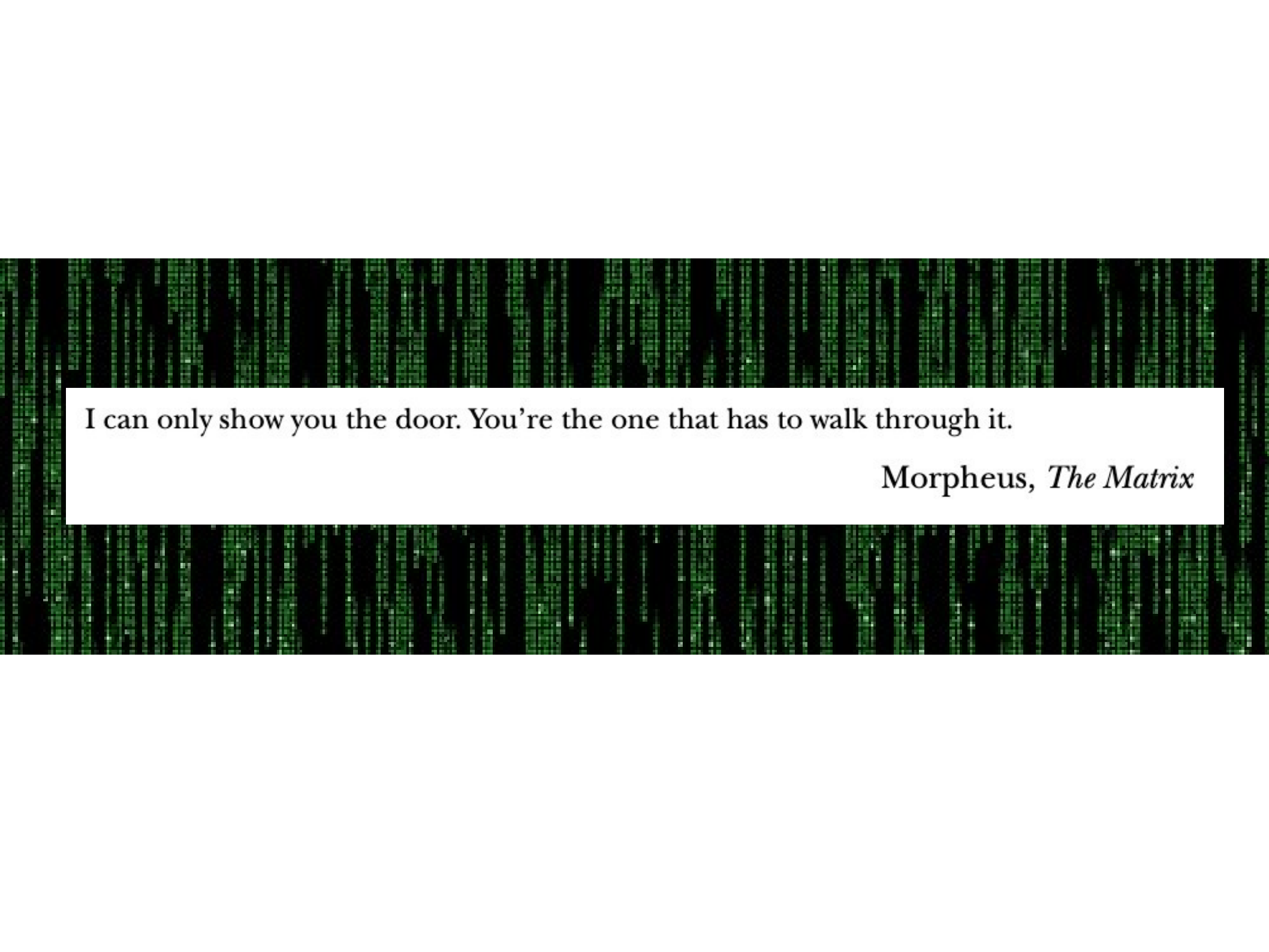
Conceptos



Rogelio Ferreira Escutia

Profesor / Investigador
Tecnológico Nacional de México
Campus Morelia





I can only show you the door. You're the one that has to walk through it.

Morpheus, *The Matrix*

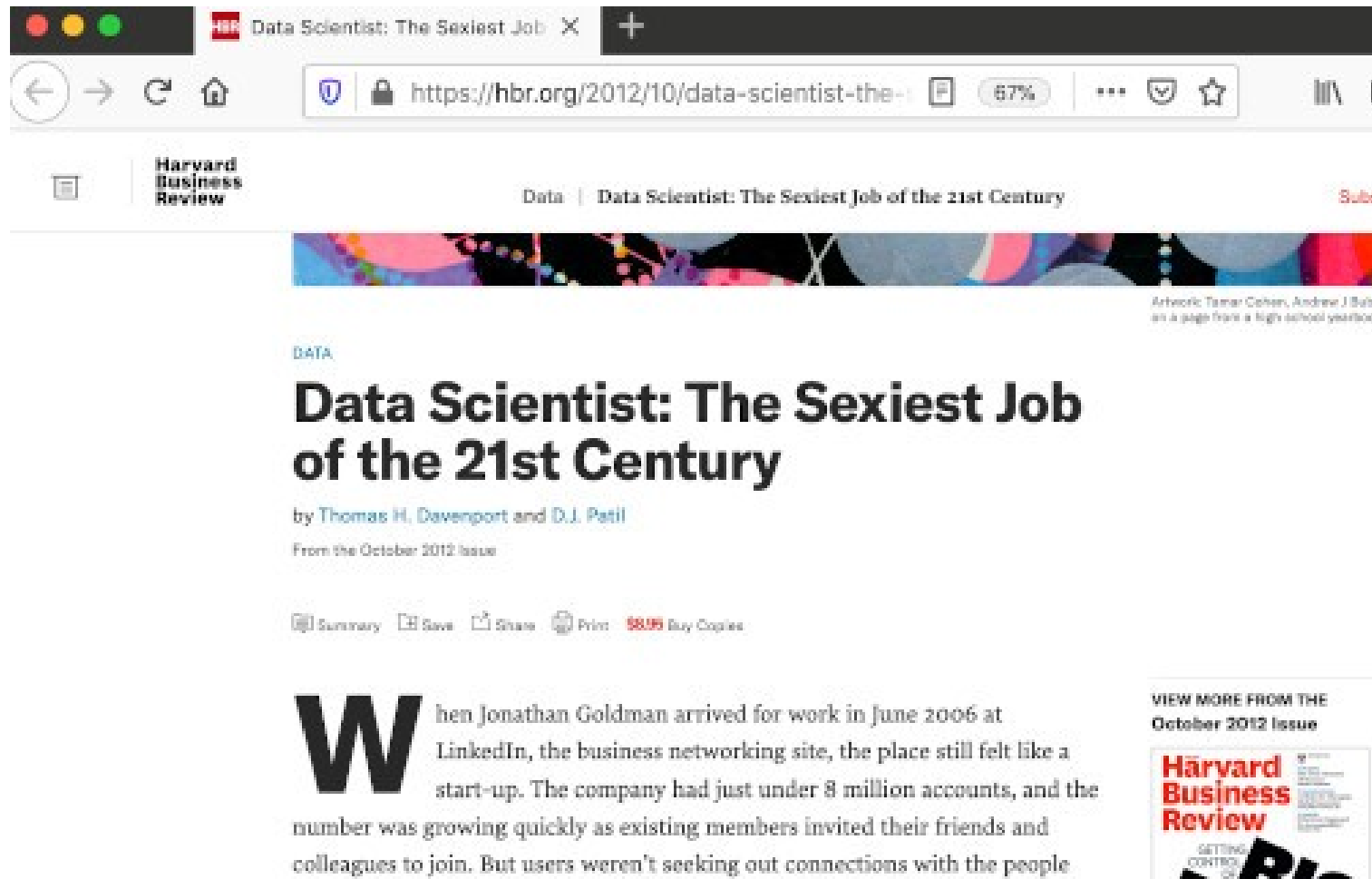
¿Qué es
Ciencia de Datos?

Definición

- **Es tan nueva que algunas personas todavía no la aceptan como una “Ciencia”, pero el crecimiento exponencial de la información ha planteado la necesidad de su análisis y por lo tanto del nacimiento y crecimiento de esta “Nueva Ciencia”.**

El trabajo mas “sexy” del siglo XXI

- Según Thomas J. Davenport:



The image is a screenshot of a web browser displaying the Harvard Business Review article "Data Scientist: The Sexiest Job of the 21st Century". The browser's address bar shows the URL "https://hbr.org/2012/10/data-scientist-the-". The page features the Harvard Business Review logo and the article title in large, bold black text. Below the title, it lists the authors as Thomas H. Davenport and D.J. Patil and notes it is from the October 2012 issue. A large, bold letter 'W' begins the first paragraph, which discusses Jonathan Goldman's experience at LinkedIn in June 2006. To the right of the main text, there is a small thumbnail image of the October 2012 issue of Harvard Business Review with the headline "GETTING CONTROL OF...".

"Data Scientist: The Sexiest Job of the 21st Century", <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>, octubre 2012



Definición

- La ciencia de datos es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados, lo cual es una continuación de algunos campos de análisis de datos como la estadística, la minería de datos, el aprendizaje automático, y la analítica predictiva.



Definición

- **Data Science es la resolución de los problemas de negocios/organizaciones a través de las matemáticas, la programación y el método científico que implica la creación de hipótesis, experimentos y pruebas a través del análisis de datos y la generación de modelos predictivos.**



Historia de la Ciencia de Datos

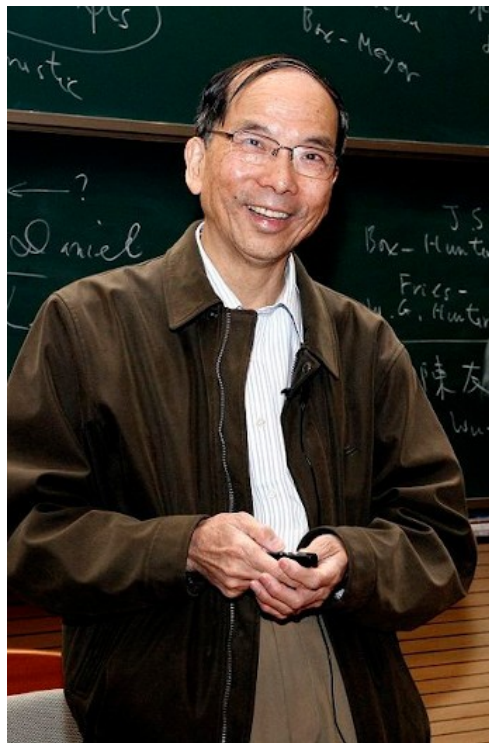
John Tukey (1962)

- En 1962, John W. Tukey precedió al término “Ciencia de Datos” en su artículo “The Future of Data Analysis” al explicar una evolución de la estadística matemática.
- Tukey introdujo la palabra "bit" como contracción de «Dígito binario» (por sus siglas en inglés Binary Digit).
- También usó el término "Software de Computación" (Computer Software) en un contexto computacional en un artículo escrito en 1958.



Jeff Wu (1997)

- En 1997 un profesor de la Universidad de Michigan Jeff Wu, durante una conferencia expuso la pregunta "Statistics = Data Science?", y a partir e ahí popularizó el término "Data Science".



Leo Breiman (2001)

- En el 2001 un profesor de la Universidad de California en Berkeley Leo Breiman publicó el artículo "Statistical Modeling: The two cultures", donde comenta acerca de la estadística antigua y el surgimiento de una tipo de nueva estadística, donde se fusiona la estadística con las Ciencias Computacionales.



¿Qué hace un
Científico de Datos?

¿Qué es un científico de datos?

- Un Científico de Datos es una persona (¿o sistema?) Encargada de analizar problemas de negocios/organizaciones ofreciendo una solución estructurada que comienza convirtiendo este problema en una pregunta válida y completa, luego usa programación y herramientas computacionales para desarrollar códigos que preparan , limpian y analizan los datos para crear modelos y responder la pregunta inicial.



Habilidades de un Científico de Datos

Principales Habilidades

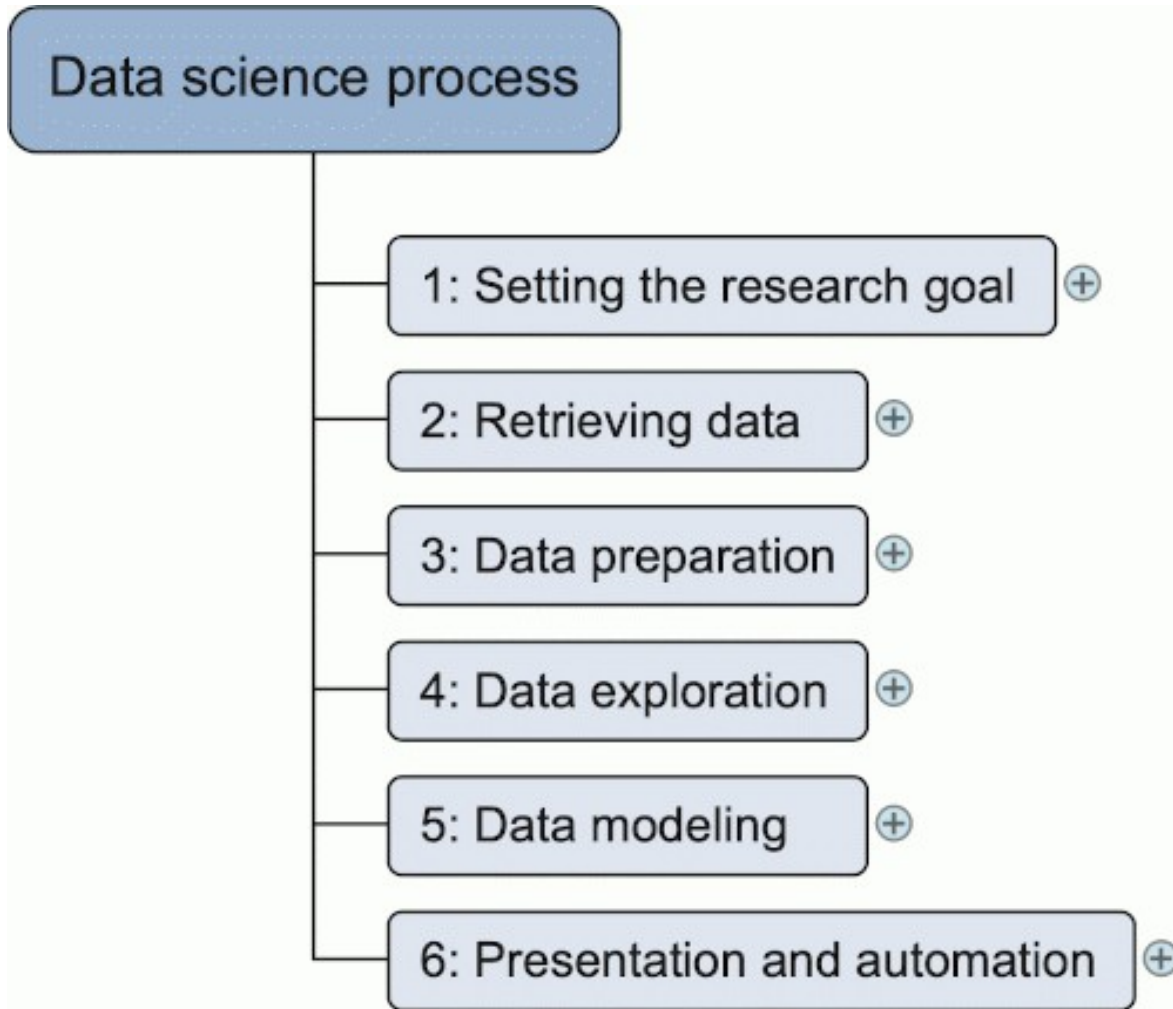
- **1) Habilidades Matemáticas y Estadísticas**
- **2) Habilidades del área de Computación**
- **3) Conocimiento del negocio**

Retos a Resolver

- Entender el contexto del problema.
- Tener datos de calidad.
- Limpieza de datos.
- Preparación de la información.
- Seguridad de la información.
- Pruebas de calidad, unidad, integración y estrés.
- Integración continua.
- Despliegue de modelos
- Monitores de proyectos, métricas y modelos

Etapas de un Proyecto de Ciencia de Datos

Pasos que sigue un Científico de Datos



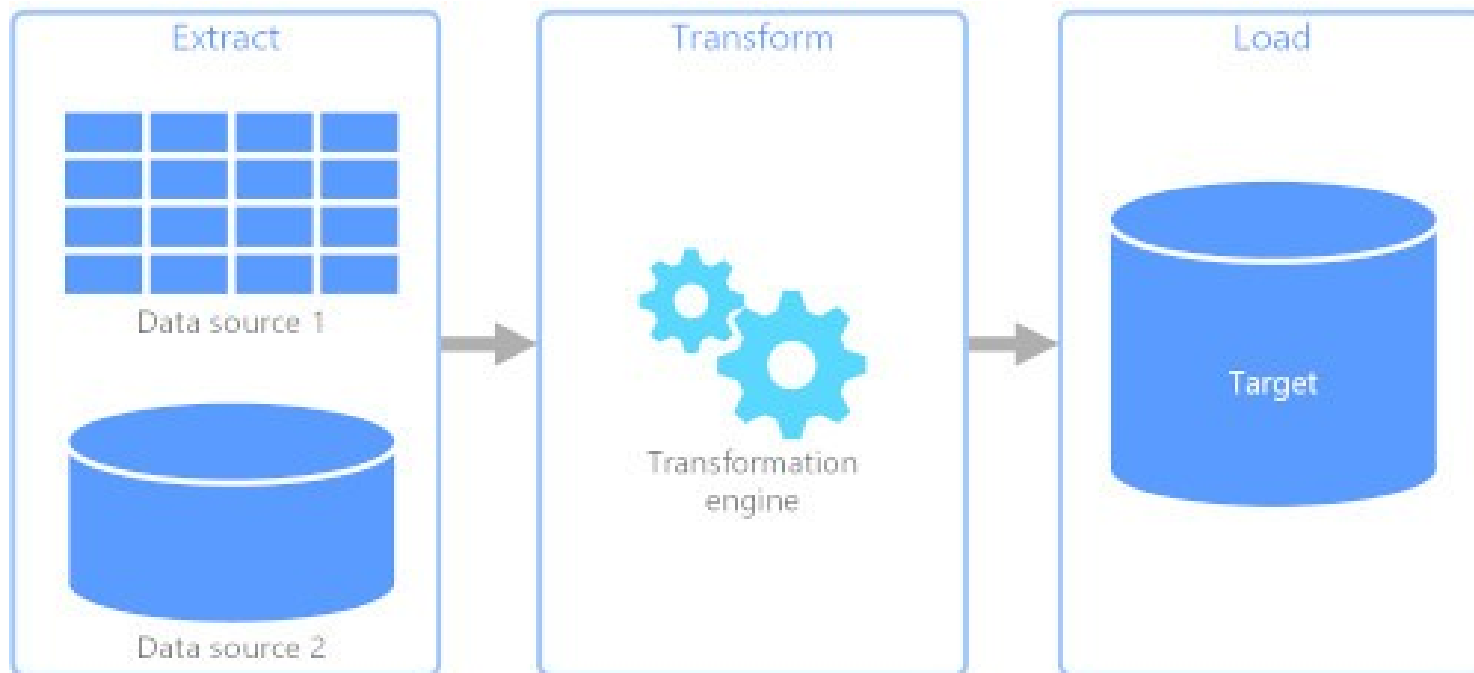
Paso 1) Definir objetivo

- **Especificar a donde se quiere llegar:**
 - **Predecir resultados.**
 - **Descubrir tendencias.**
 - **Encontrar patrones.**



Paso 2) Extracción de datos

- **Localizar la ubicación de los datos para poder leerlos y en su caso almacenarlos en nuestro sistema de base de datos.**



Paso 3) Preparar los datos

- **Antes de hacer algún proceso, se hace un preprocesamiento o limpiado de los datos, con el objetivo de poder tener solamente los datos requeridos y en el formato deseado.**



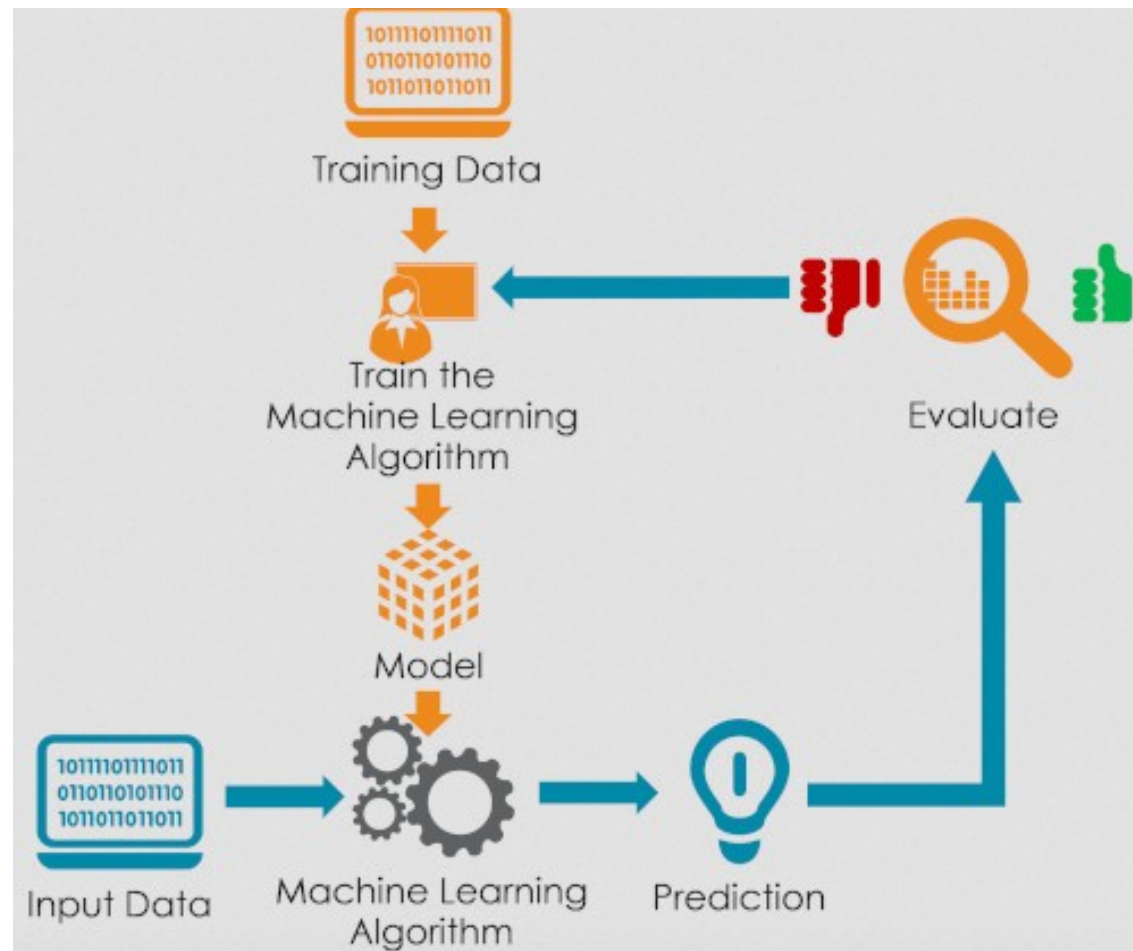
Paso 4) Exploración de los datos

- En este paso se analizan y procesan los datos



Paso 5) Modelado

- A partir de los datos se crea un modelo para lograr llegar a nuestro objetivo y encontrar resultados.



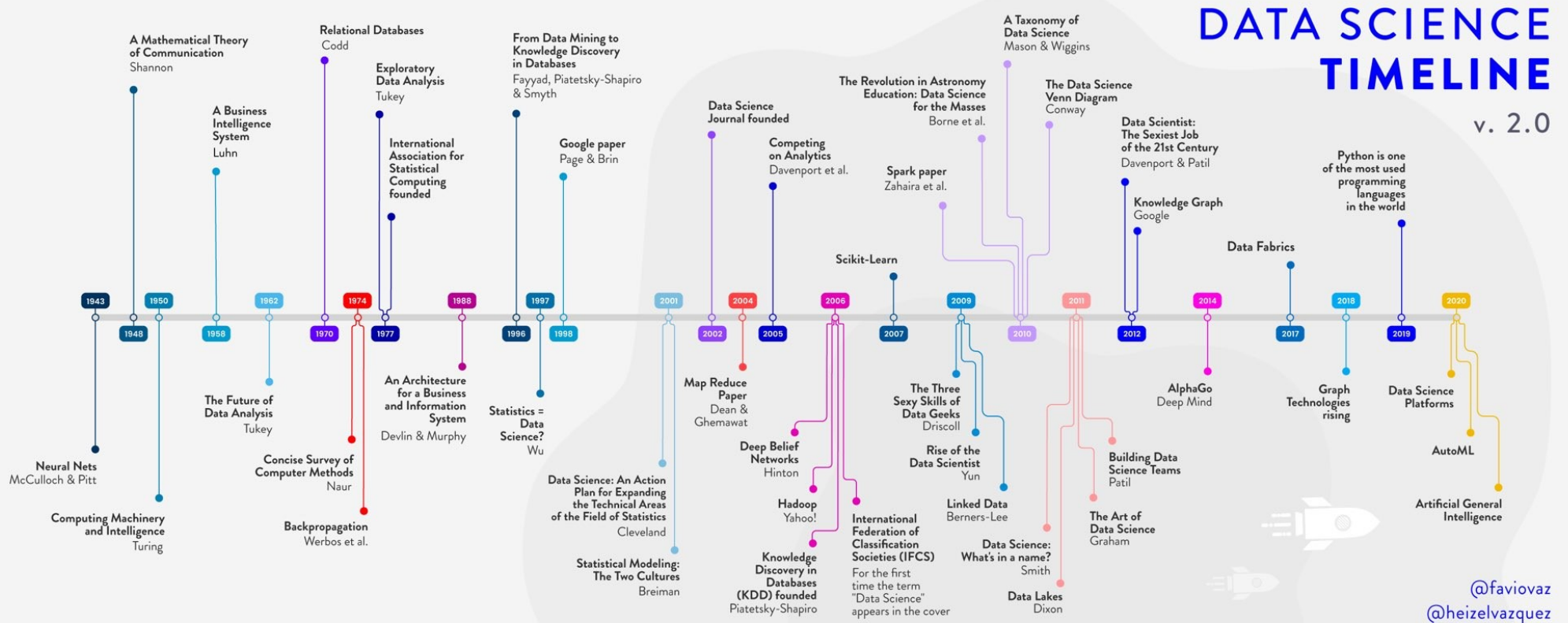
Paso 6) Visualización

- Finalmente hay que visualizar (graficar) los datos obtenidos para poder entender más fácilmente los resultados.

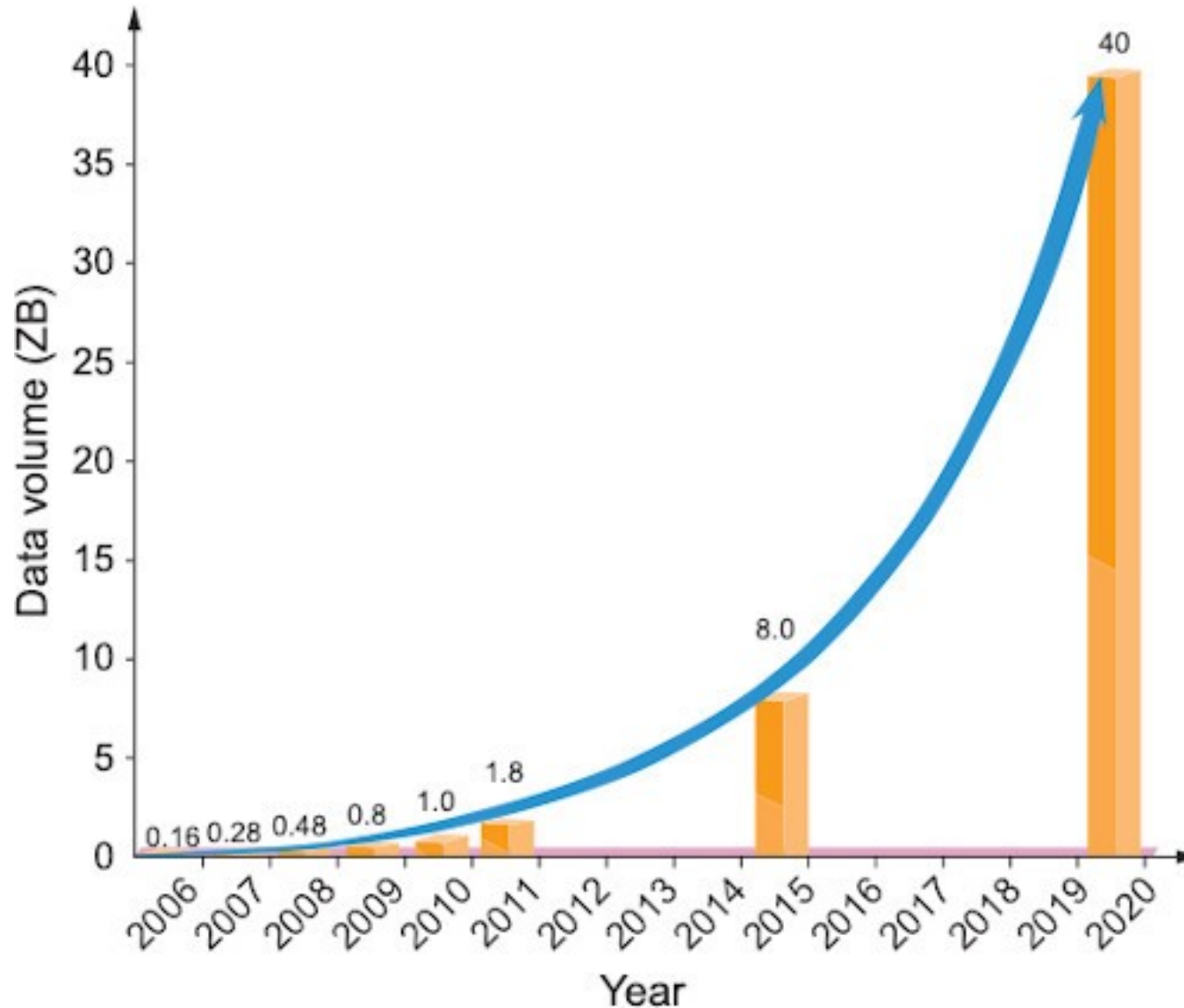


Surgimiento y Evolución de la Ciencia de Datos

Data Science Timeline



Crecimiento de datos



Ciencia de Datos en México

Infotec (2016)

- Lanza Infotec la Maestría en Ciencia de Datos



Infotec (2016)

- **Lanza Infotec el Doctorado en Ciencia de Datos**



UNAM (2019)

- **Lanza la UNAM la Licenciatura en Ciencia de Datos**

La Jornada: Ante mar de inform... x +

jornada.com.mx/2019/05/25/sociedad/030n1soc

SOCIEDAD

Sábado 25 de mayo de 2019

INICIO EDITORIAL CORREO ILUSTRADO OPINIÓN POLÍTICA ECONOMÍA MUNDO ESTADOS

CAPITAL SOCIEDAD CULTURA ESPECTÁCULOS DEPORTES CARTONES

NOTICIAS DE HOY ESPECIALES MULTIMEDIA SERVICIOS OTROS SITIOS CONTACTO

USTED ESTÁ AQUÍ: INICIO / SOCIEDAD Y JUSTICIA / ANTE MAR DE INFORMACIÓN, CREA UNAM LA CARRERA CIENCIA DE DATOS /

◀ A nterior

Siguiente ▶

Facebook Twitter Correo Más... 22

Conocimiento digital

Ante mar de información, crea UNAM la carrera ciencia de datos

Alumnos aprenderán a interpretar a partir de algoritmos

di Para entender mejor Haz doble clic sobre cualquier palabra de nuestra web para ver su significado

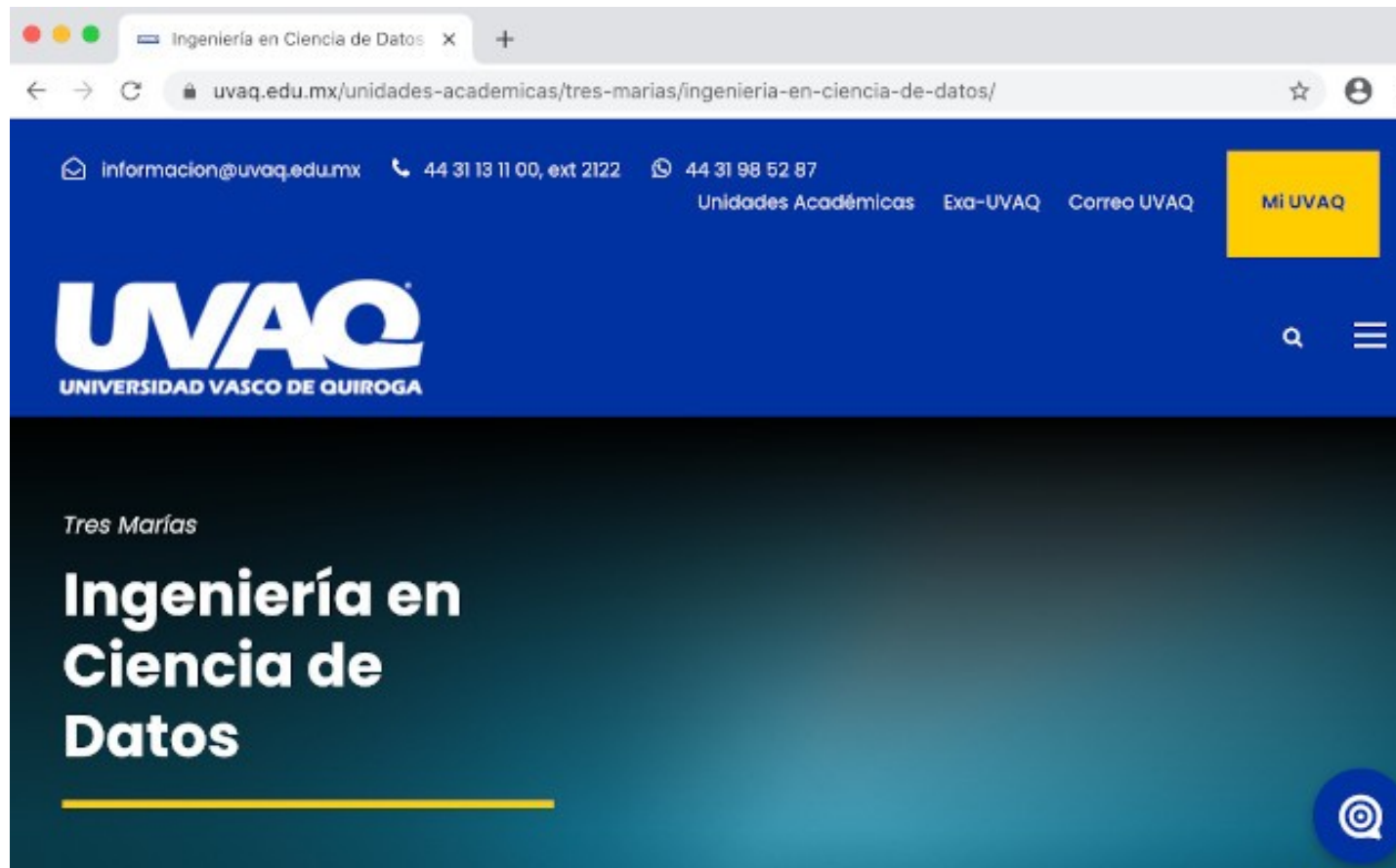
Conalep (2020)

- **Lanza Conalep la carrera de Ciencia de Datos e IA**



UVAQ (2020)

- **Lanza la UVAQ la Ingeniería en Ciencia de Datos**



IPN (2021)

- El IPN lanza la “Licenciatura en Ciencia de Datos”:



The screenshot shows the IPN website header with the logo and the motto "Instituto Politécnico Nacional 'La Técnica al Servicio de la Patria'". Navigation icons for language, search, and font size are visible. The breadcrumb trail reads "Inicio / Oferta Educativa / Educación Superior /". Below this is a large image of a digital data network with glowing blue and orange lines. The main heading is "Programa académico / Licenciatura en Ciencia de Datos". A dark blue box at the bottom contains the text "Se imparte en: ESCOM | UPIIC Campus Coahuila | UPIIT Campus Tlaxcala".

Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"

Inicio / Oferta Educativa / Educación Superior /

Programa académico / Licenciatura en Ciencia de Datos

Se imparte en: ESCOM | UPIIC Campus Coahuila | UPIIT Campus Tlaxcala

ITESM (2021)

- El ITESM lanza el primer “Hub de Ciencia de Datos”:



UDLAP (2023)

- La UDLAP arranca la “Licenciatura en Ciencia de Datos”:

The screenshot shows the UDLAP website interface for the 'Licenciatura en Ciencia de Datos' program. At the top, there is a navigation bar with 'INICIO', 'SITIOS A-Z', 'ENGLISH', and social media icons (Facebook, YouTube, LinkedIn, Twitter). A search bar labeled 'Búsqueda' is also present. Below the navigation, the main header features the UDLAP logo and 'CIENCIA DE DATOS'. The central content area is divided into three main sections: 1. A video player showing an aerial view of a modern building with the text 'LICENCIATURA EN CIENCIA DE DATOS' overlaid. 2. A registration form titled 'Sé un Estudiante UDLAP' with fields for 'Nombre', 'Email', 'Teléfono', 'Institución de procedencia', and a math question '¿Cuánto es 1 + 11?'. It includes a 'Cel' radio button, a 'Teléfono' radio button, a 'Aviso de privacidad' link, and an 'ENVIAR' button. 3. A teal banner at the bottom right with the text 'PROCESO DE ADMISIÓN'.



Rogelio Ferreira Escutia

Profesor / Investigador
Tecnológico Nacional de México
Campus Morelia



rogelio.fe@morelia.tecnm.mx



rogeplus@gmail.com



xumarhu.net



[@rogeplus](https://twitter.com/rogeplus)



[https://www.youtube.com/
channel/UC0on88n3LwTKxJb8T09sGjg](https://www.youtube.com/channel/UC0on88n3LwTKxJb8T09sGjg)



[rogelioferreiraescutia](https://www.linkedin.com/in/rogelioferreiraescutia)

