



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

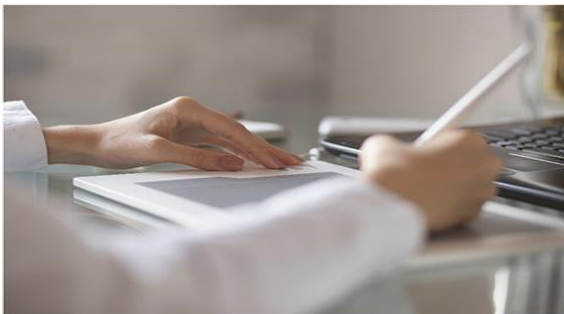
EDUCACIÓN
PROFESIONAL

Diplomado en Big Data y Analítica de Datos 2021

Curso: Fundamentos Machine & Deep Learning

Educación Profesional Escuela de Ingeniería UC

Profesor: Rodrigo Sandoval





ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

9. Proyectos Machine Learning

Fuente parcial. Ismini Lourentzou & Boris Ginzburg

Se necesita un proyecto ML/DL cuando ...

Se requiere responder ¿qué clase de ...? o ¿cuál es la predicción de ...?

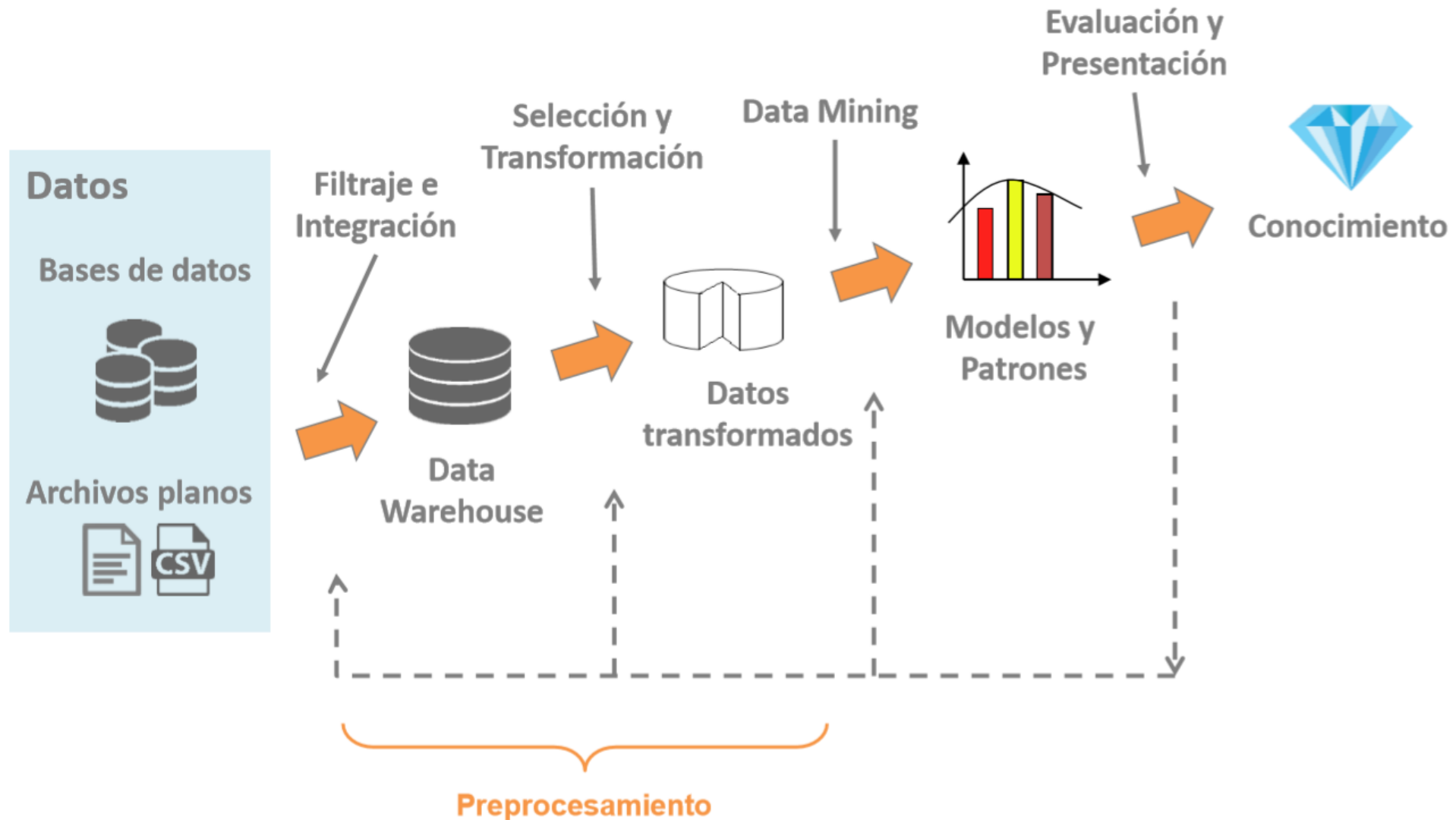
Se necesita responder ¿cuáles son los grupos que distinguen ...?

Se necesita interpretar texto en lenguaje natural (texto/audio) o imágenes (fotos/videos)

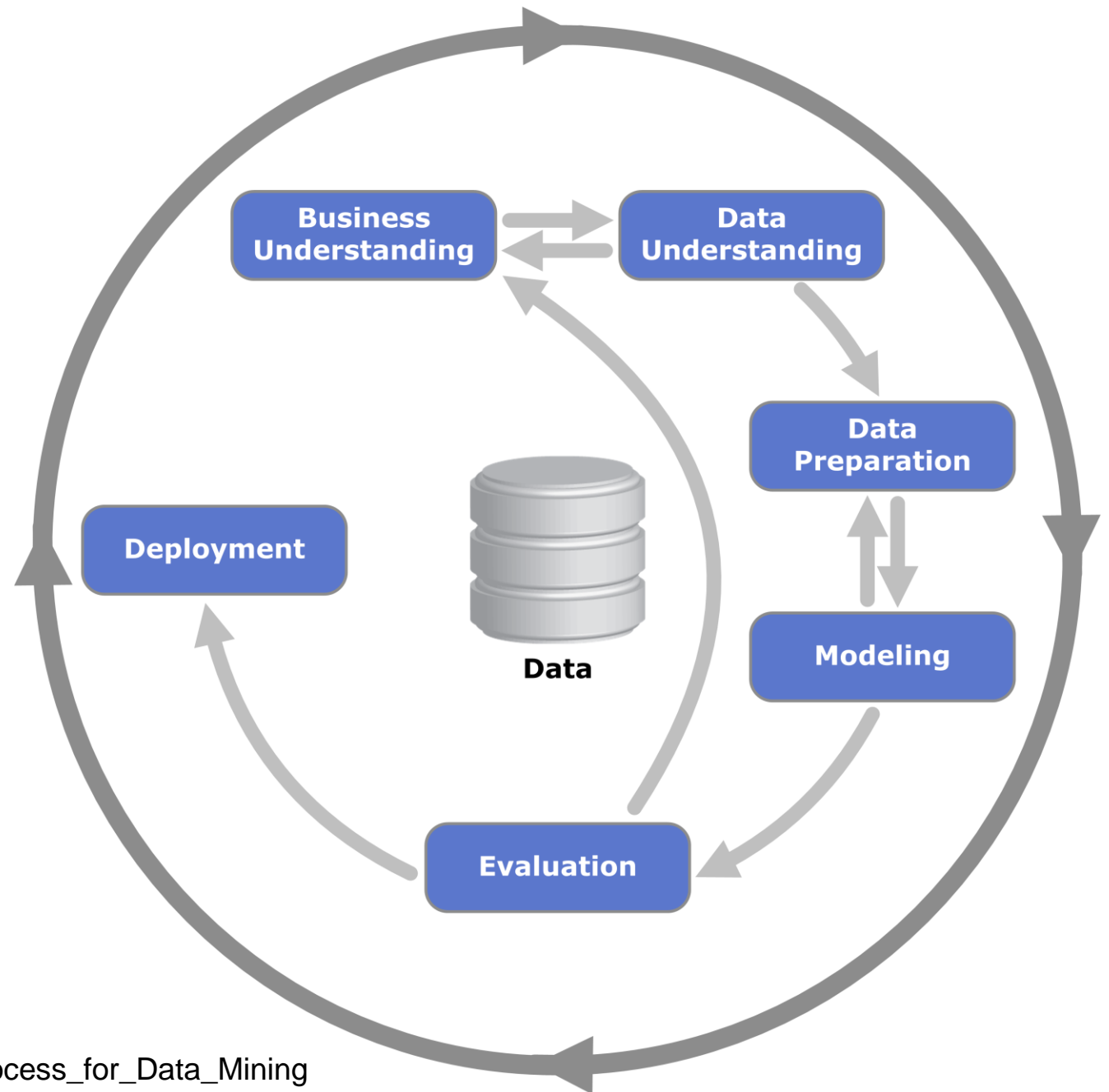
Y existe la posibilidad de contar con datos adecuados.

Enfrentando un proyecto Machine Learning

KDD – Knowledge Discovery in Databases

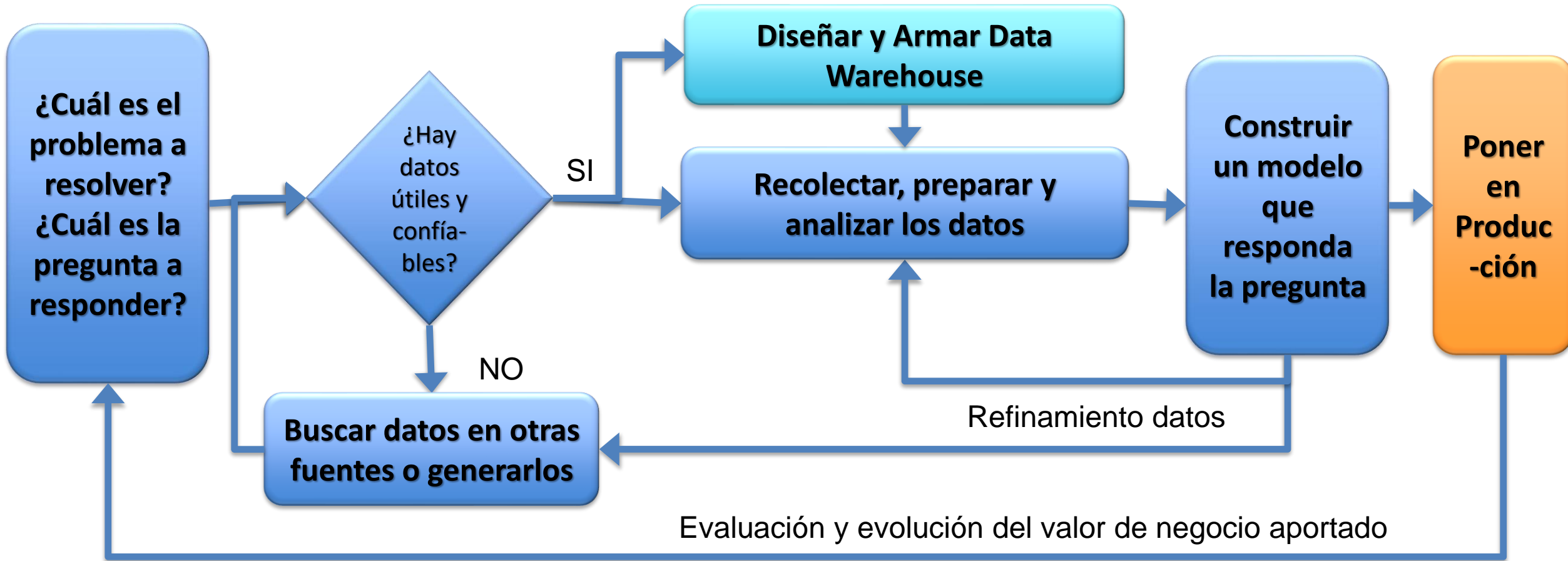


CRISP-DM: Procesos con Data Mining



https://es.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

Enfoque R:Solver para valor de negocio con Data Mining



Los puntos clave son (1) determinar la pregunta, (2) recolectar y preparar datos, (3) elegir y refinar un modelo que responda la pregunta. Siempre considerar que la situación de negocio y el modelo que la resuelve van evolucionando en el tiempo.

Dedicación de Tiempo y Esfuerzo “normal”

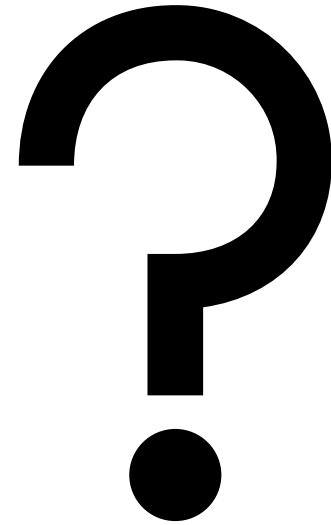


Poder armar modelos confiables depende en parte importante de contar con los datos adecuados: disponibles, confiables, limpios, manipulables, variados, balanceados, etc.

Pregunta(s) fundamental(es)

¿Cuánto tiempo tomará?

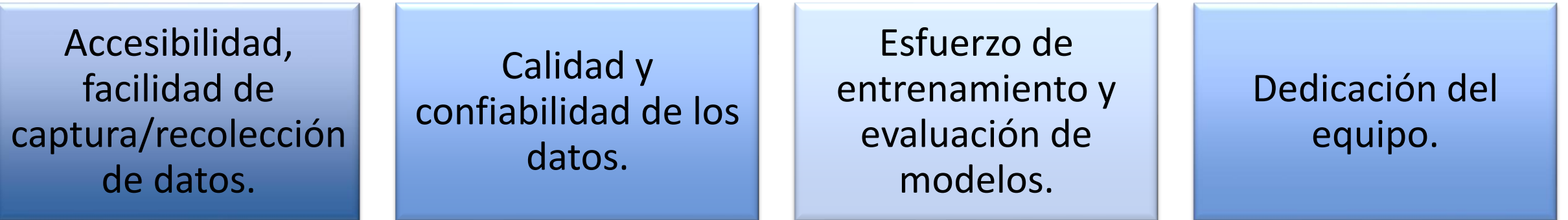
¿Cuánto va a costar?



Incerteza inherente en proyectos ML

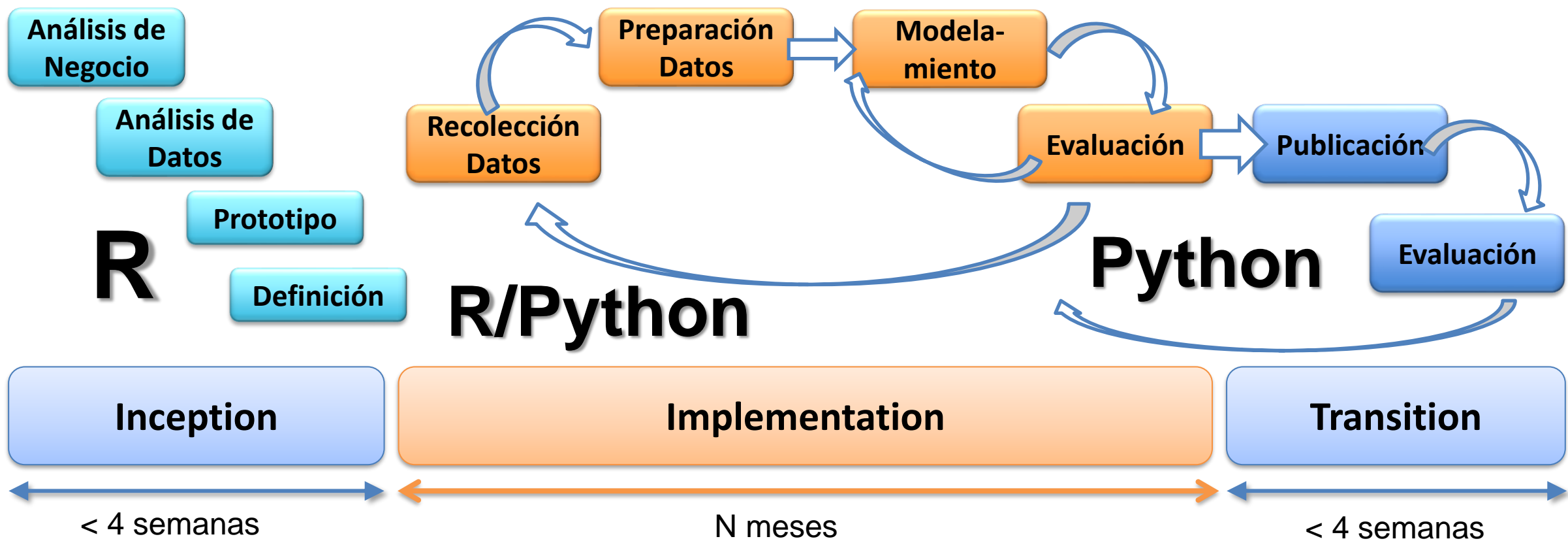
En la práctica, es casi imposible responder la clásica pregunta de “¿cuánto tiempo necesitas para tener un modelo confiable y de buen desempeño?”

Normalmente, esto depende de:



Disciplined Agile aplicada a proyectos Machine Learning

Propuesta de R. Sandoval / R:Solver



¿Cuánto tiempo total proyecto?

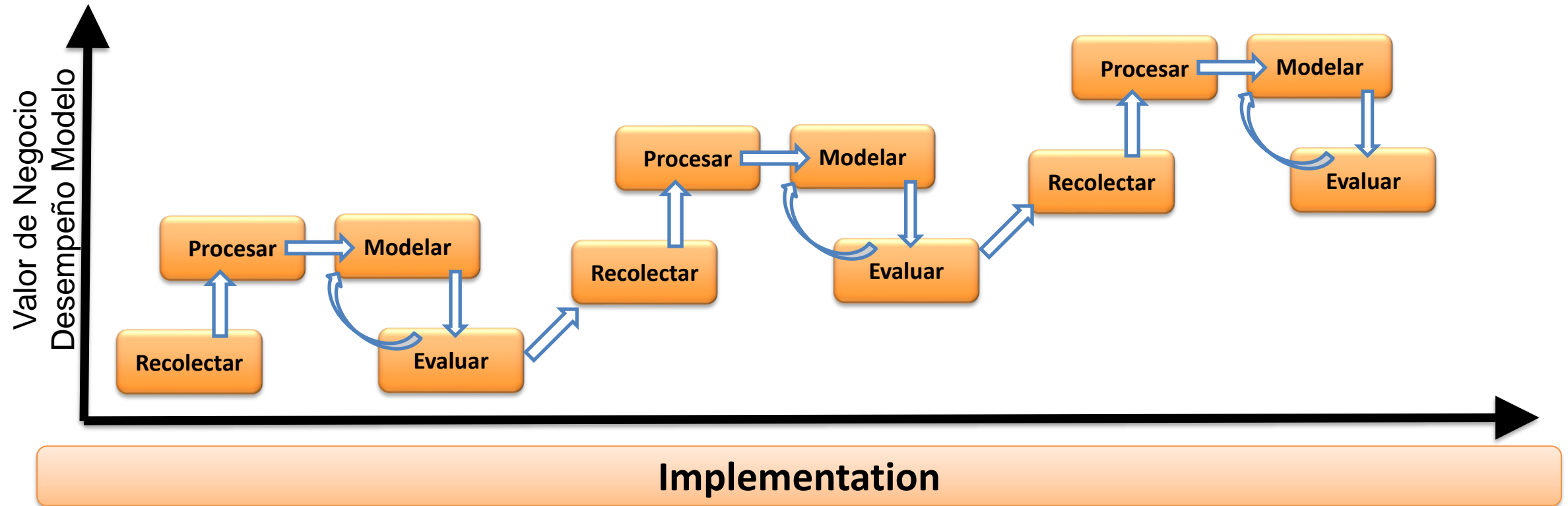
Mínimo 2 meses (depende de esfuerzo de recolección de datos y preparación)

¿Cuánto esfuerzo (costo) en el proyecto?

Considerar ~4 meses y un equipo de 1 a 4 Data Scientists.
Sumar costos de entrenamiento *cloud* para contextos de datos no-estructurados (texto, imágenes), si no, podría lograrse con computadores de trabajo. Costos *cloud* pueden ser entre US\$100 y US\$2000 por mes.

Disciplined Agile aplicada a proyectos Machine Learning

Propuesta de R. Sandoval / R:Solver



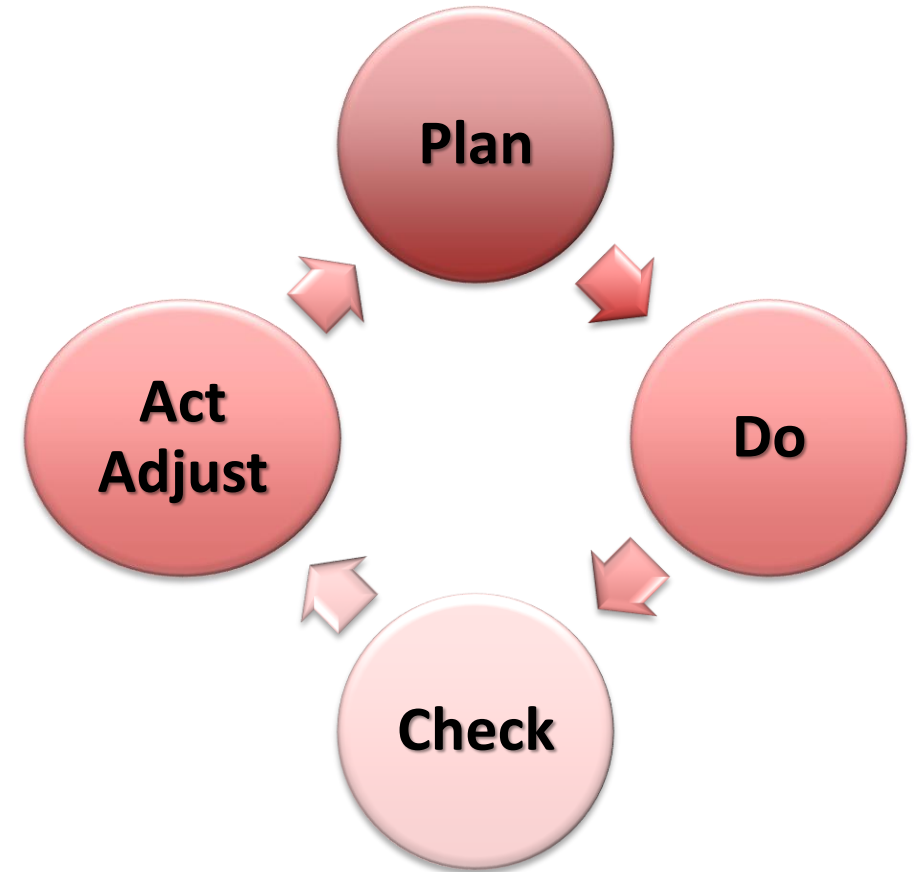
¿Cuánto tiempo?

Si se designa tiempo (presupuesto) de N meses, entonces se itera y mejora hasta que se acaba el tiempo, entregando cada pocas semanas una nueva versión de mejor desempeño incrementalmente.

Las Ventajas de un esquema Ágil

- Utilizando enfoque ágil, se producen versiones visibles, de desempeño cuantificable, que entregan claridad de puntos de desafío, complejidad y capacidad de ser usable en “producción”.
- Se puede establecer un proyecto “Etapa 1” cuyo objetivo es producir una primera versión que tenga el mejor desempeño posible.
- Evaluar si es necesario continuar con “Etapa 2”, pero siempre teniendo una versión usable de la Etapa 1.

Ciclo Deming Mejora Continua



Costos infraestructura Cloud

Se requiere cuando hay contextos de datos no-estructurados y de gran volumen (~ Big Data)

The screenshot displays the Google Cloud Pricing Calculator interface. At the top, a navigation bar includes icons for various services: COMPUTE ENGINE, APP ENGINE, KUBERNETES ENGINE, CLOUD RUN, VMWARE ENGINE, CLOUD STORAGE, NETWORKING EGRESS, CLOUD LOAD BALANCING, and INTE & CL. Below this is a search bar with the placeholder text "Search for a product you are interested in." and a magnifying glass icon.

The main content area is divided into two columns. The left column, titled "Instances", contains several input fields with question mark icons to the right of each field:

- Number of instances *
- What are these instances for?
- Operating System / Software: Free: Debian, CentOS, CoreOS, Ubuntu, or other User Provided OS
- Machine Class: Regular
- Machine Family: General purpose
- Series: N1
- Machine type: f1-micro (vCPUs: shared, RAM: 0.60 GB)

The right column, titled "Estimate", shows the calculated values for the selected configuration:

- Compute Engine
- 2 x
- 208.571 total hours per month
- VM class: regular
- Instance type: n1-standard-96
- Region: Iowa
- GPU dies: 8 NVIDIA TESLA K80
- GPU's Cost: USD 750.86
- GCE Instance Cost: USD 951.08
- Total available local SSD space 2x375 GiB
- Estimated Component Cost: USD 1,719.08 per 1 month
- Total Estimated Cost: USD 1,719.08 per 1 month
- Estimate Currency: USD - US Dollar

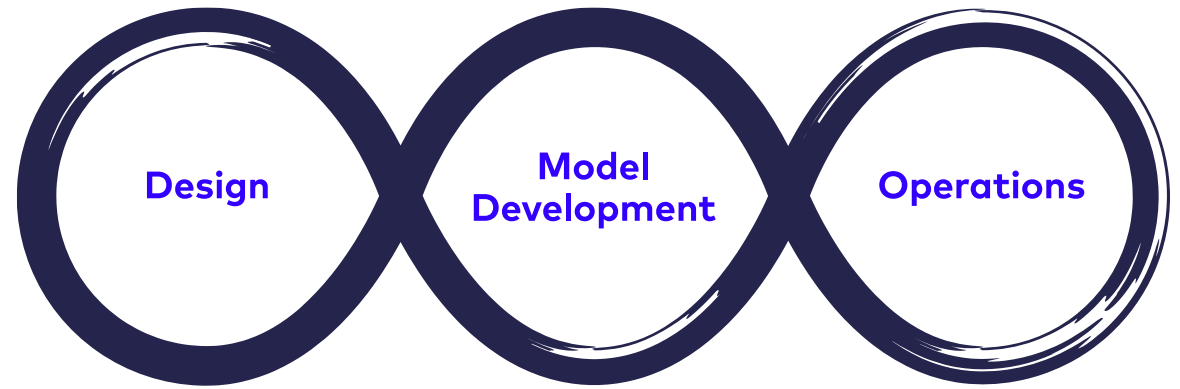
At the bottom of the right column, there are two buttons: "EMAIL ESTIMATE" and "SAVE ESTIMATE".

Rango:
US\$200 a
US\$2000
por mes

<https://cloud.google.com/ai-platform?hl=es>

MLOps – Machine Learning Operations

- MLOps es un campo de disciplina técnica, que apunta a la entrega y operación continua .
- Busca unificar el ciclo de liberación de aplicaciones con ML + Software.
- Facilita la automatización del testing y puesta en producción integrada.
- Reduce la deuda técnica entre modelos ML.
- Combina prácticas independientes de lenguajes, plataformas, frameworks e infraestructura.



MLOps – Pilares Fundamentales.

Evaluación de Riesgos

- El riesgo depende de la operación del modelo.
- Foco en disponibilidad, calidad predicción, exactitud en el tiempo, las habilidades necesarias para su mantención.

Responsabilidad e Intencionalidad

- Tener un control central de los esfuerzos con IA, con equipos y datos utilizados.
- Aseguramiento de calidad de los datos, según su origen. “Explicabilidad”

Escalamiento y Personas

- Versionamiento y trazabilidad de desempeño en el tiempo.
- Equipos multi-disciplinarios: expertos de negocio, Data Scientists, Engineers, etc.

MLOps – Metas

Despliegue y automatización

Reproductibilidad de modelos y predicciones

Diagnósticos y evaluación

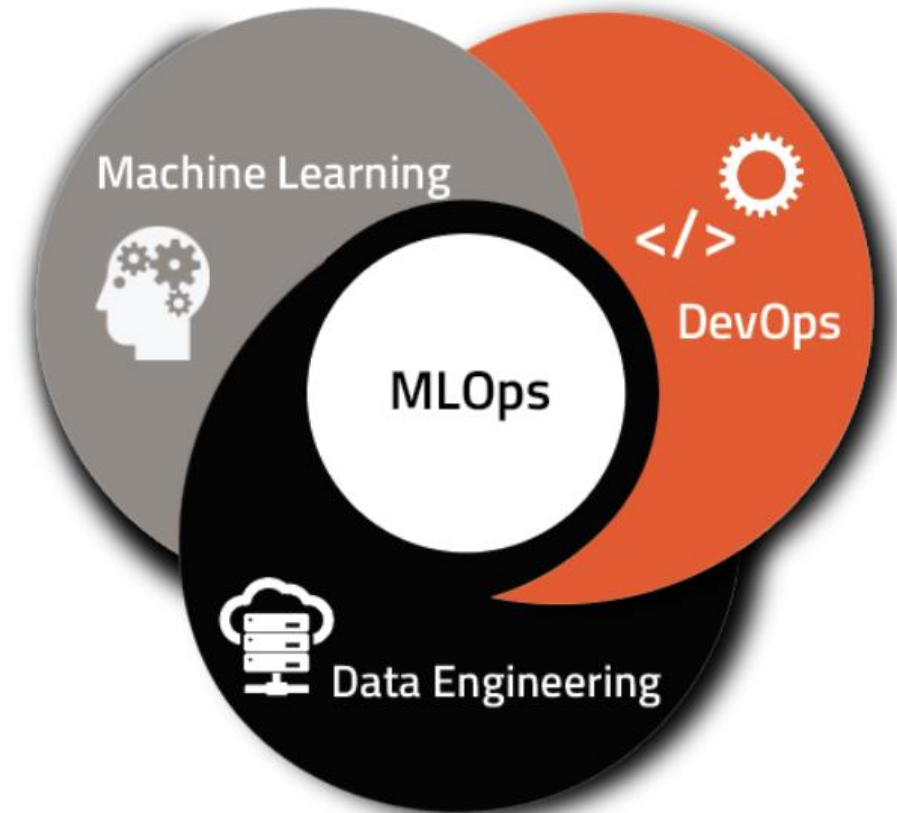
Gobierno y cumplimiento regulatorio

Escalabilidad

Colaboración

Casos de negocio

Monitoreo y gestión



Gracias



rsandova@ing.puc.cl

rodrigo@RSolver.com



@RSandovalSolver



/in/RodrigoSandoval

www.RodrigoSandoval.net

www.RSolver.com