




Rodrigo Diaz, **Manuel Szewc**, Luis Agustin Nieto

Aprendizaje Automático

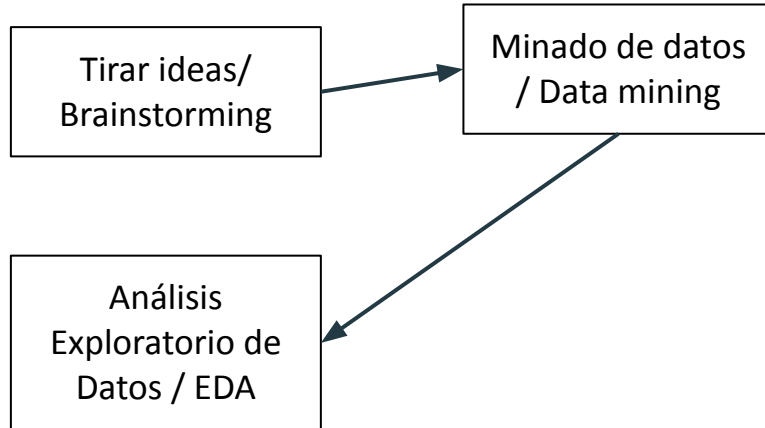
Jueves 11/03:
Un ejemplo casi de cabo a rabo



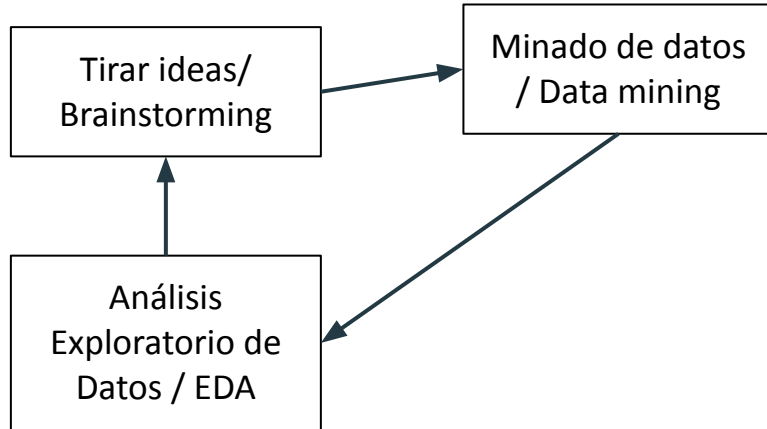
Pasos de un proyecto (a lo Geron)

1. Poner el proyecto en perspectiva
2. Conseguir los datos
3. Explorar los datos para desarrollar una intuición sobre los mismos
4. Preparar los datos para aplicar los distintos algoritmos
5. Seleccionar un modelo y entrenarlo
6. Encontrar la mejor versión del modelo para los datos
7. Presentar el modelo terminado
8. Lanzar, monitorear y mantener el sistema en el que está enmarcado el modelo.

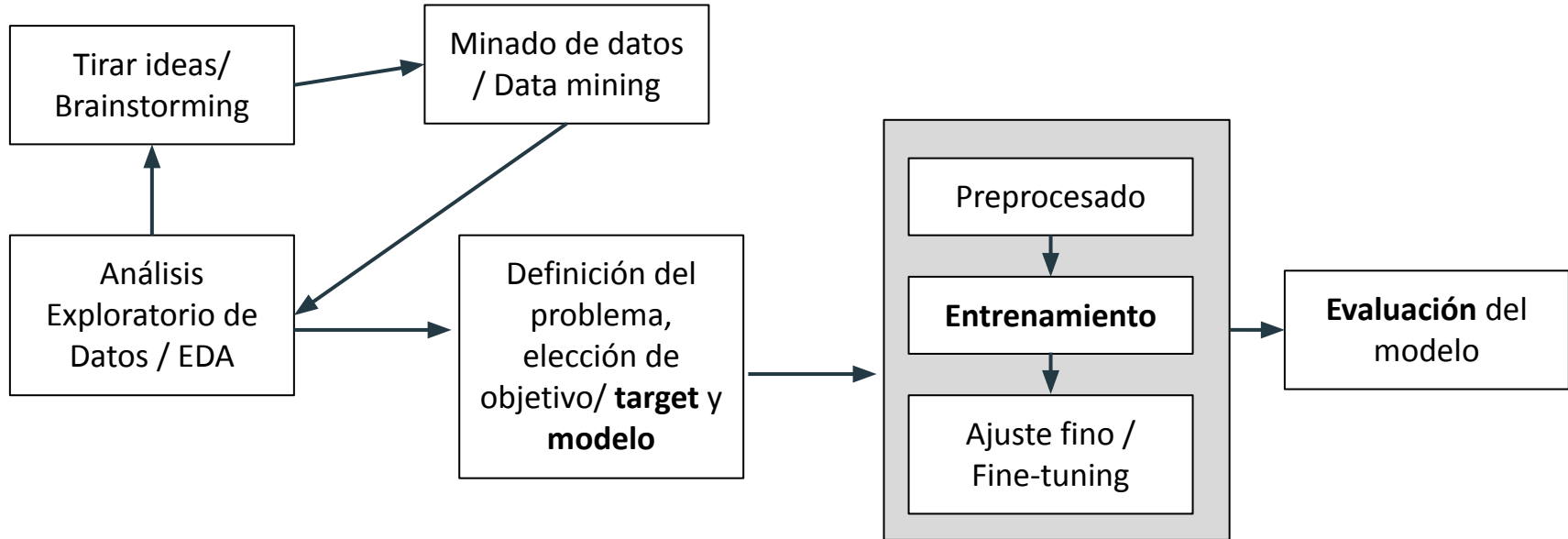
Pasos de un proyecto (a lo Geron)



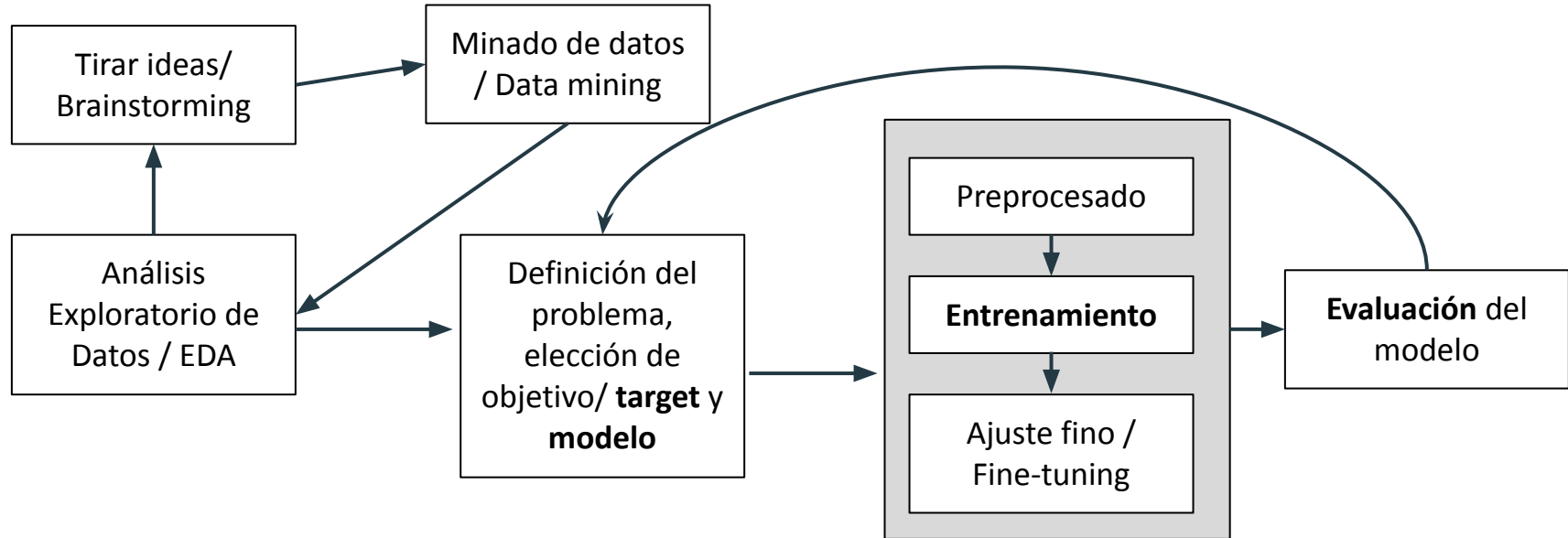
Pasos de un proyecto (a lo Geron)



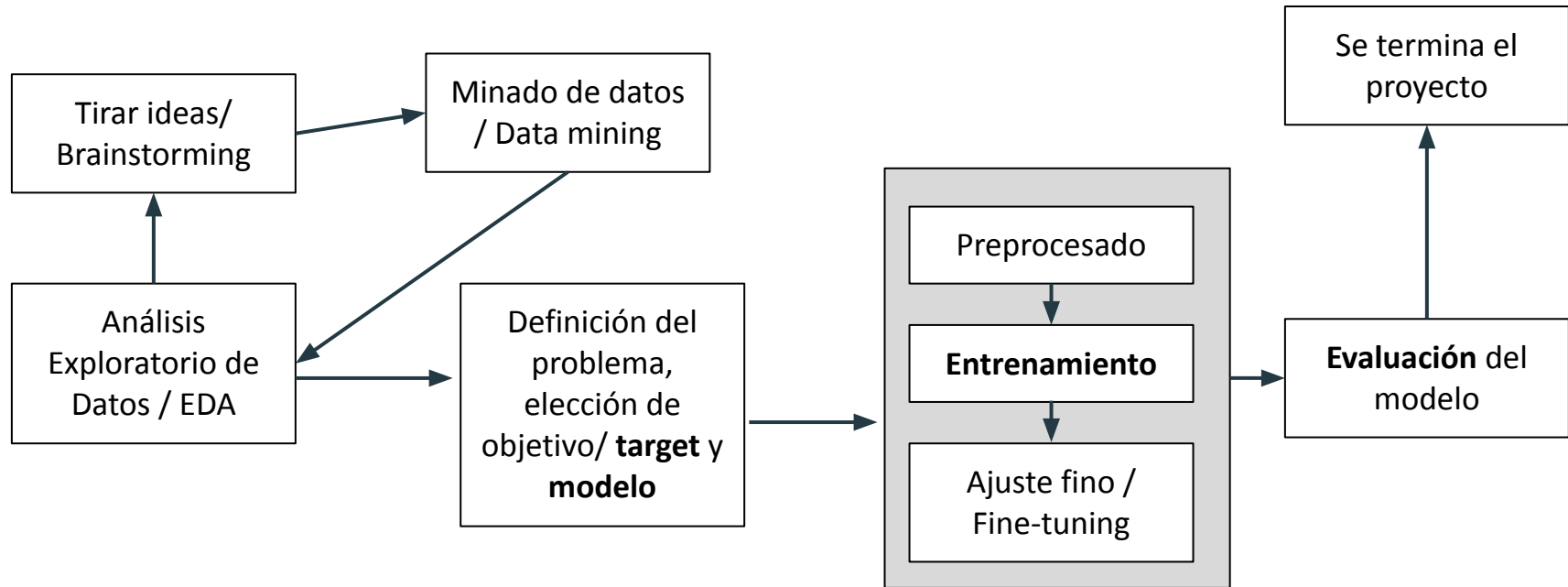
Pasos de un proyecto (a lo Geron)



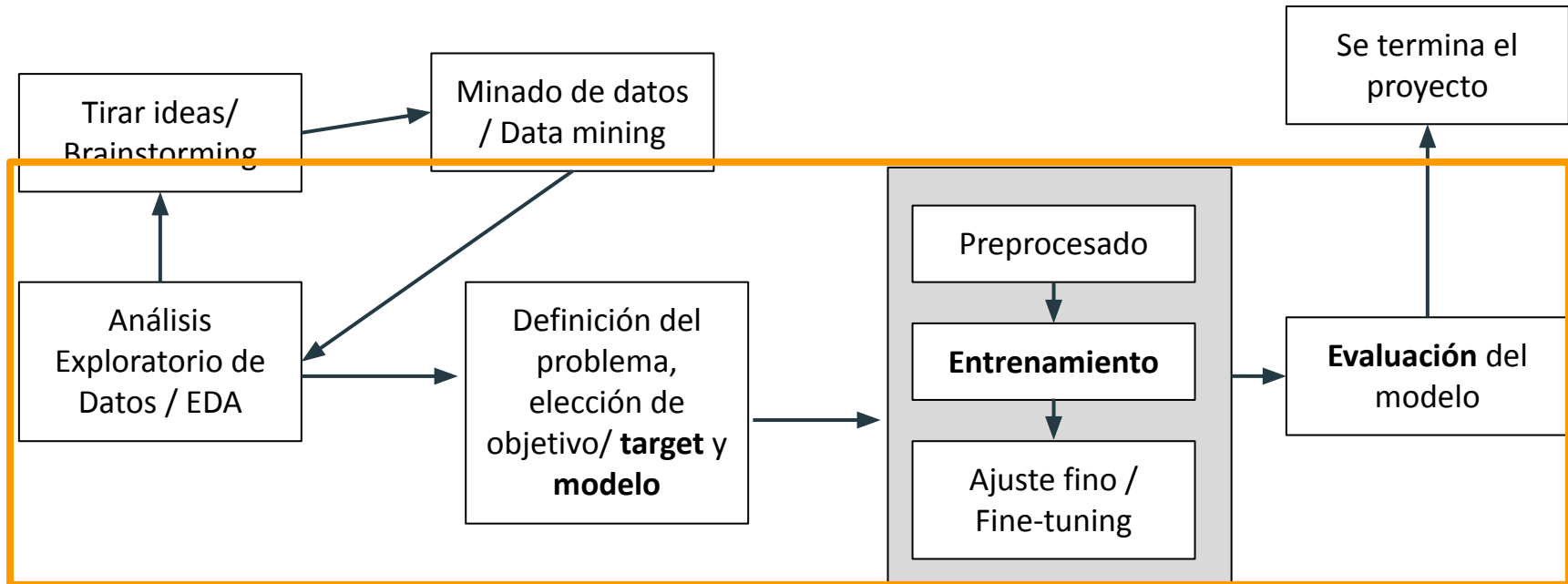
Pasos de un proyecto (a lo Geron)



Pasos de un proyecto (a lo Geron)



Pasos de un proyecto (a lo Geron)



Definición del problema, elección de objetivo/target y modelo

- Qué información tienen los datos?
- Queremos organizar los datos? Hacer predicciones?
- Que tipo de predicciones? Tenemos los datos adecuadamente etiquetados? Son etiquetas continuas (regresión) o categóricas (clasificación)?
- Cuán correlacionado está nuestro target con las variables (o **features**) que tenemos? Es un problema lineal? O es más complejo?
- Existen modelos que se adecuen al problema y al target definidos?

Definición del problema, elección de objetivo/ target y modelo

- Tenemos que separar en entrenamiento / validación / testeo (o implementar validación cruzada) para poder obtener una predicción lo más robusta posible a nuevos datos
- Tenemos que elegir o definir las métricas de nuestro problema. Cómo se mide el buen o mal funcionamiento de los modelos y se compara entre ellos?

Preprocesado

- Ingeniería de datos / feature engineering: Adecuamos los datos de manera tal que los modelos funcionen mejor
- Limpieza de datos: Decidir que se hace con datos faltantes y anomalías (o outliers) estadísticas.
- Preprocesado de datos: Estandarización, encodeo de variables categóricas.

Entrenamiento

- Un modelo simple que sirva como referencia / benchmark.
- Exploración de distintos modelos.
- Utilizar los resultados iniciales / intermedios para ir mejorando la selección de modelo.
- Elegir los mejores modelos y ajustarlos todavía más a los datos. Ojo con el sobreajuste!

Evaluación del modelo

- Evaluamos el modelo en datos no explorados hasta ahora (conjunto de testeo)
- Obtenemos el error estimado en datos nuevos
- Estudiamos el modelo para ganar interpretabilidad (si es posible)
- Encontrar donde no funciona el modelo
- De ser necesario, volver a empezar con la sabiduría ganada

AI notebook!