

## Fases de un proyecto ML

- Necesito un sistema con ML?
- Qué hago si mi sistema No funciona bien?

## Ejemplo: Detector de Spam (Email Classification)

7 pasos del ML

60% SVM  
40% x → Reg los  
→ 50% x → SVM x

- 1: Obtener datos \*
- 2: Analizar los datos (Limpieza)
- 3: Dividir conjuntos de datos para train/test/validation
- 4: Definir especificaciones del sistema
- 5: Entrenar un modelo (el más simple que funcione) \*
- 6: Evaluar y Medir rendimiento del modelo //
- 7: repetir (Todo)

### 1: Obtener Datos

→ Datos Reales, representativo

Correos SPAM  
no SPAM  
→ Dominio sobre el cual trabajo

→ aspectos éticos y legales  
→ Nunca es demasiado \*

### 2: Analizar los datos

→ Ver el contenido  
→ Construir herramientas → (Jupyter)  
→ Necesidad de expertos \*  
→ Multidisciplinario

### 3: Particionar datos



entrenar el modelo      Evaluar hiperparámetros      Medir el rendimiento del modelo

### 4: Especificaciones

- plasmar la visión del sistema
  - requisitos específicos
- Evitar Ambigüedades
- Definir el comportamiento
  - Qué es SPAM?

### 5 Implementar el modelo más simple posible

- No sobre optimizar
  - C++
- Entender los datos más que los modelo
- Consideren el costo !!

spec:  $\geq 60\%$   
300\$ ← SVM  $\approx 70\%$  ✓ 1ms,  $\sim 20$  MB  
4500\$ ← ANN  $\approx 85\%$  30ms,  $\sim 200$  MB  
60 req/min

### 6: Evaluar y Medir

- Visualizaciones \*
  - ver el funcionamiento
  - identificar bugs
- Curvas de Aprendizaje
- Monitoreo constante
- Post mortem.
  - Variables correlacionadas

### 7: Repetir

- mejorar en cada iteración