

ACÀMICA

¡Bienvenidas/os a Data Science!



Agenda

¿Cómo anduvieron?

Data Science en mi vida

Repaso: Puesta en producción

Explicación: Pipelines de Scikit Learn y Deploy

Break

Hands-on training

Lanzamiento Entrega 07

Cierre



¿Dónde estamos?



¿Cómo anduvieron?



Hoja de ruta

fase	ADQUISICIÓN Y EXPLORACIÓN		MODELADO				DEPLOY
	Exploración de datos	Feature Engineering	Machine Learning: Clasificación y Regresión	Optimización de parámetros	Procesam. del lenguaje natural	Sistema de recomendación	Publicación de modelos
tiempo	SEM 1	SEM 5	SEM 8	SEM 12	SEM 14	SEM 18	SEM 22
	SEM 2	SEM 6	SEM 9	SEM 13	SEM 15	SEM 19	SEM 23
	SEM 3	SEM 7	SEM 10		SEM 16	SEM 20	SEM 24
	SEM 4		SEM 11		SEM 17	SEM 21	



Cronograma

Usted
Está Aquí

SEM 20

- PCA
- Sistemas de recomendación

SEM 21

- Sistemas de recomendación
- Trabajar en la entrega 6 + Empleabilidad

SEM 22

- Trabajar en la entrega 6 + Empleabilidad
- Puesta en producción

SEM 23

- Puesta en producción
- Trabajar en la entrega 7

SEM 24

- Cierre de cursada

Entrega 6

Entrega 7



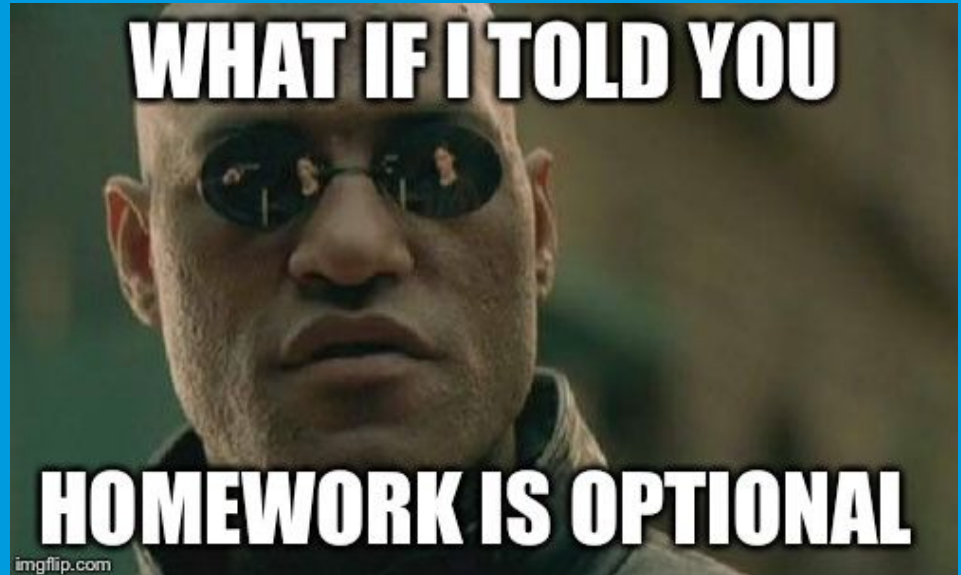
Importante Anuncio:

Entrega 07

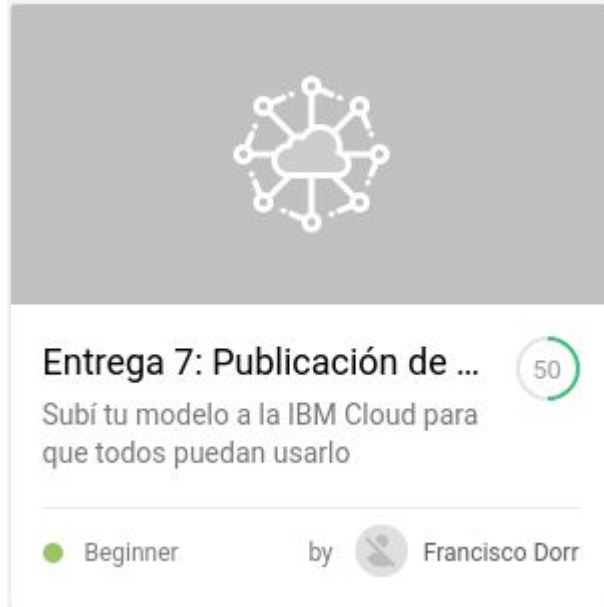


Importante Anuncio:

Entrega 07



Proyecto 3: Publicación de Modelos (Entrega 07)



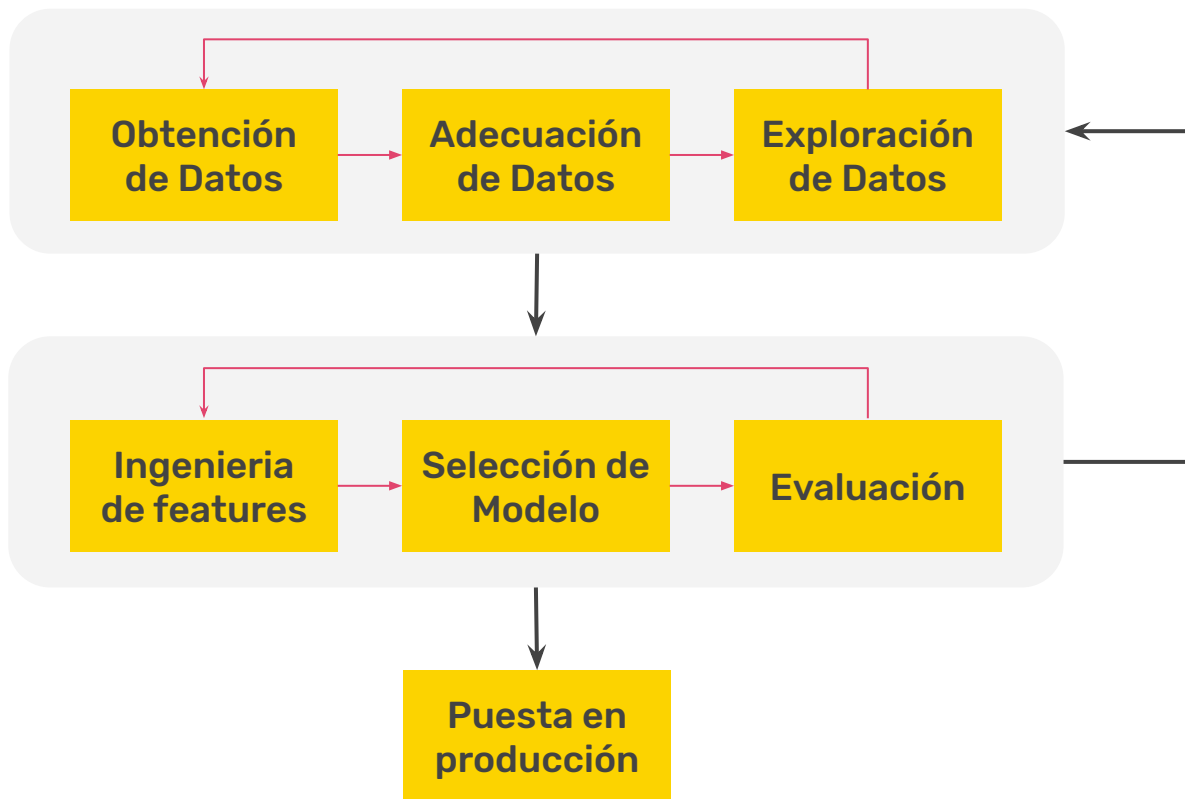
1. Bajar los materiales.
2. Leer la Checklist
3. ¡Empezar a trabajar en la entrega!

OPCIONAL !!

Repaso: Puesta en producción



Recorrido completo



¿Por qué Puesta en producción?



Queremos que el modelo que nosotros creamos pueda ser usado por otros usuarios. Con este fin vamos a tener que tener en cuenta **tres aspectos principales** que pueden resultar problemáticos:

- Acceso
- Compatibilidad (Lenguajes, Hardware, Librerías, etc.)
- Escala

SOLUCIÓN 1

Súper usuarios

Todos tienen los modelos y los recursos para correrlos



SOLUCIÓN 2

Division of labor!

Usuario
(Cliente)

Sólo mando
consultas

Servidor

Los modelos
corren acá



SOLUCIÓN 2

Division of labor!

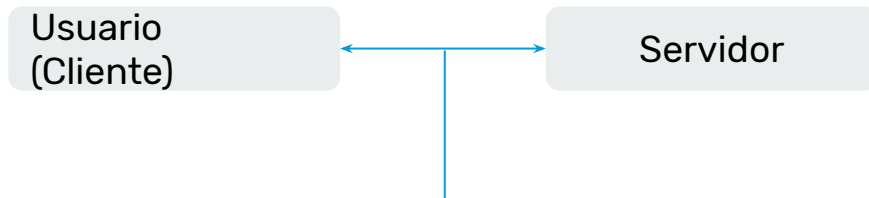


Para el servidor **existen dos posibilidades** dependiendo del uso que le queramos dar:

- Servidor Local (en una empresa o institución) - Red local o intranet
- Nube - Internet

SOLUCIÓN 2

Division of labor!



API:

Application programming interface

Es una librería con una serie de funciones que nos permiten comunicarnos con el servidor

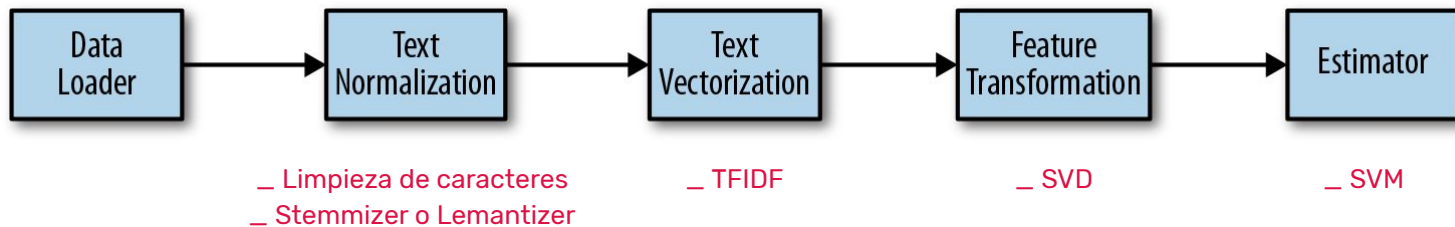
Pipelines de Scikit Learn



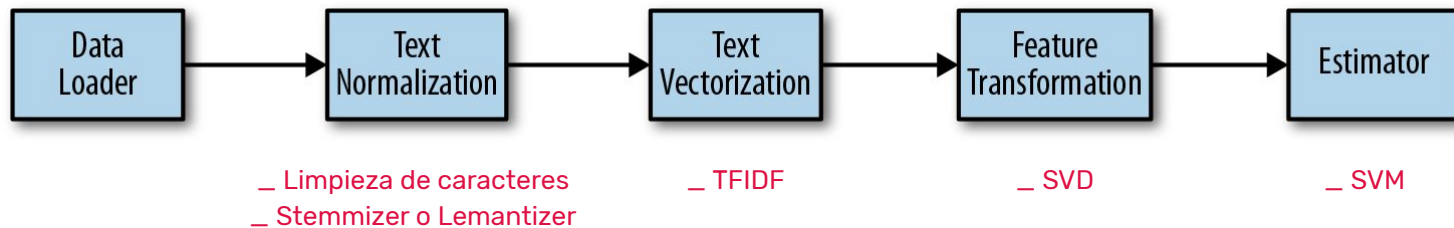
Workflow de trabajo

Desde la carga de los datos con los que vamos a trabajar hasta la salida del modelo, solemos aplicar una serie de pasos encadenados uno detrás del otro. A este camino se le llama “flujo de trabajo” (Workflow).

Por ejemplo, para un problema de NLP, el flujo podría estar compuesto por las siguientes acciones:



Workflow de trabajo



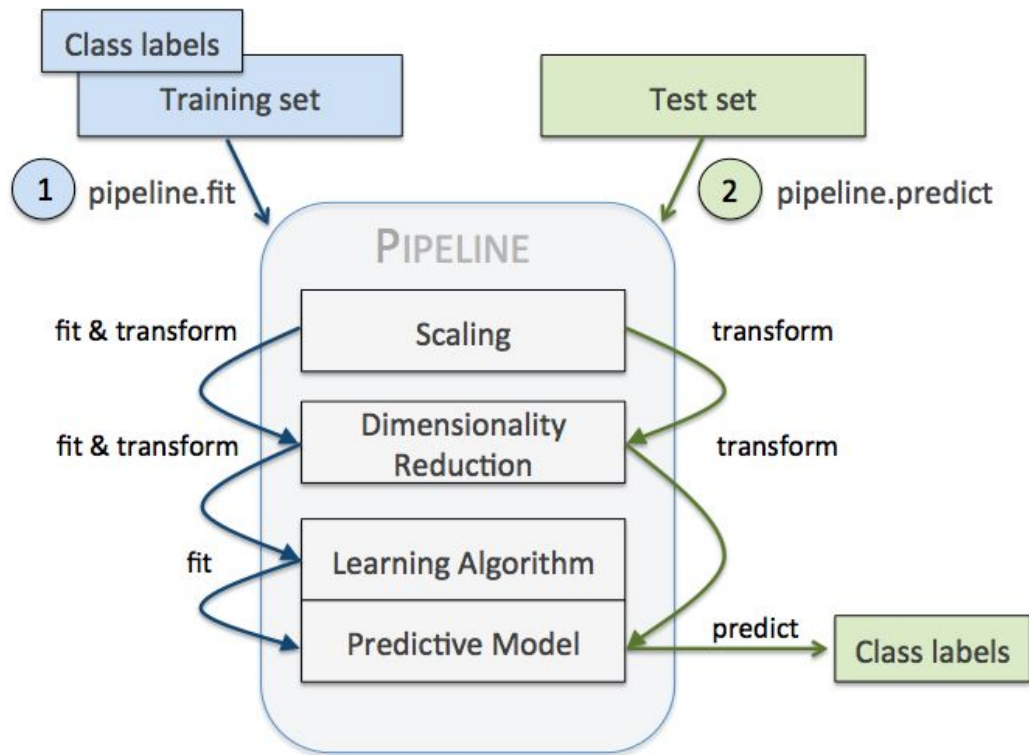
Notemos que tanto los datos del **Training Set** como los del **Test Set** deben realizar este mismo recorrido.

¿Qué es un Pipeline?



Pipeline • Definición

Es un único objeto que nos permite empaquetar todas estas acciones que van del preprocesamiento de los datos a la predicción del modelo.



Pipeline • ¿Para qué usarlo?

Es conveniente utilizarlo porque:

- Simplifica el proceso y aumenta la reproducibilidad
- Evita cometer errores (como saltarse algún preprocesamiento o mezclar datos del training set con datos del test set)
- Simplifica la implementación de cross-validation y la elección de hiperparámetros.

DOCUMENTACIÓN



EJEMPLOS



Pipeline • Ejemplo de Pipeline

```
from sklearn.feature_extraction.text import  
CountVectorizer,TfidfTransformer  
from sklearn.svm import LinearSVC  
from sklearn.pipeline import Pipeline
```

```
X_train,X_test,y_train,y_test = make_my_dataset()
```

```
vect = CountVectorizer()  
tfidf = TfidfTransformer()  
clf = LinearSVC()
```

```
pipeline = Pipeline([('vect',vect),('tfidf',tfidf),('clf',clf)])  
pipeline.fit(X_train,y_train)  
y_preds = pipeline.predict(X_test)
```


Deploy: Subir nuestro modelo a la nube



IBM Watson • ¿Cómo usarlo?



Vamos a seguir los pasos detallados [en la plataforma de Acámica](#), y crearnos un usuario de IBM Cloud (si ya lo hicieron el encuentro pasado, no lo hagan nuevamente).

Luego hay que crear un recurso de Machine learning siguiendo los pasos sugeridos detallados en [este link](#).

SKLEARN CON IBM CLOUD



Hands-on training



Hands-on training

DS_Encuentro_47_Pipelines_y_Deploy.ipynb



Para la próxima

1. Terminar de ver los videos de Puesta en producción.
2. Terminar el notebook de hoy.
3. Comenzar a trabajar en la Entrega 07

ACÀMICA