

ACÀMICA

¡Bienvenidas/os a Data Science!



Agenda

¿Cómo anduvieron?

Explicación: Puesta en Producción

Sentiment Analysis

Hands-on training

Break

Pipelines - Sklearn

Cierre



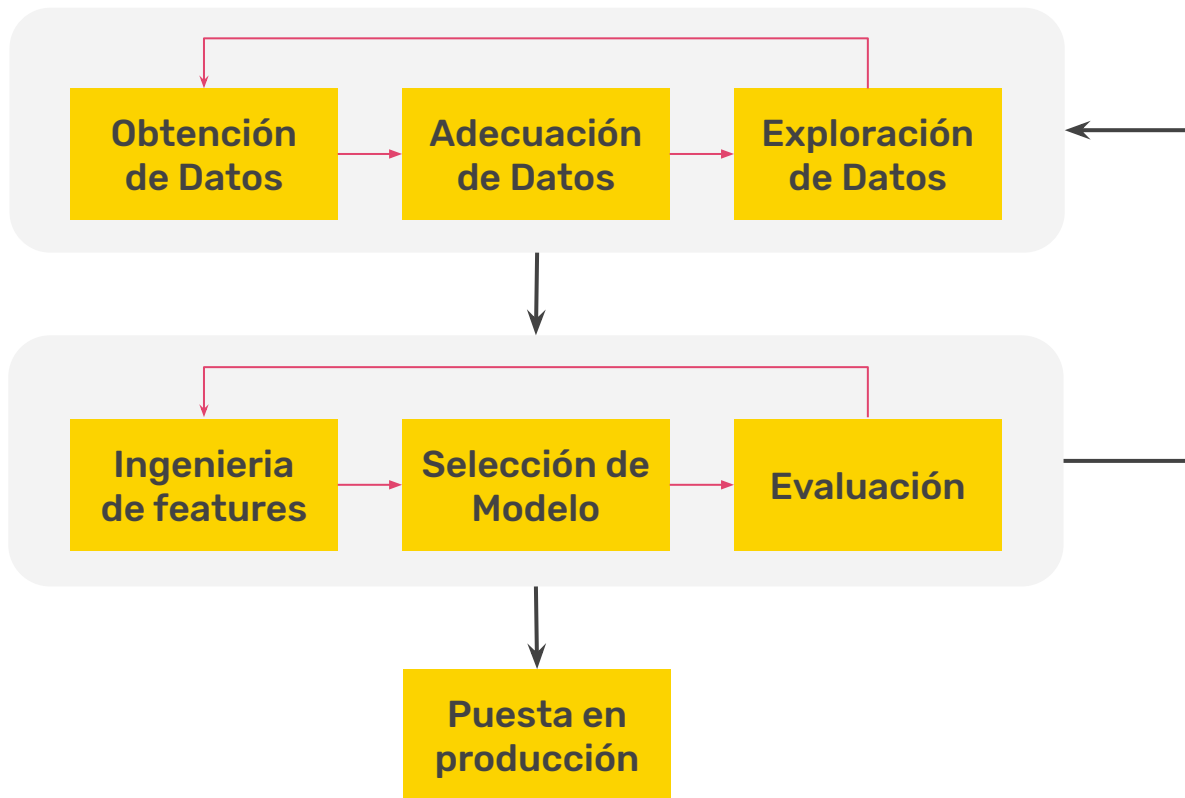
¿Cómo anduvieron?



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 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!

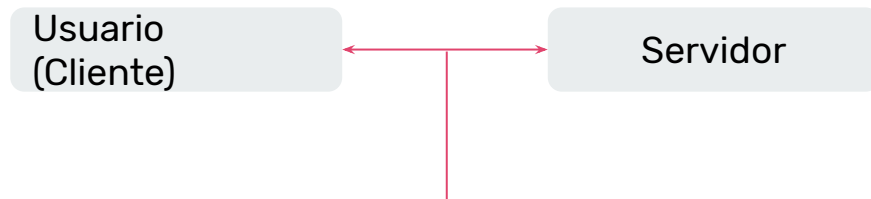


Para el servidor **existen dos posibilidades** dependiendo del uso que le queramos dar:

- 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

API de Watson: IBM cloud

IBM Watson • ¿Qué es?



Watson es una inteligencia Artificial multipropósito desarrollada y mantenida por IBM. Está especialmente diseñada para lidiar bien con el lenguaje natural y tiene varias funcionalidades. Consta de una serie de modelos muy complejos (poco bias) entrenados con bases de datos muy grandes.

IBM Watson • ¿Qué es?



Watson es una inteligencia Artificial multipropósito desarrollada y mantenida por IBM. Está especialmente diseñada para lidiar bien con el lenguaje natural y tiene varias funcionalidades. Consta de una serie de modelos muy complejos (poco bias) entrenados con bases de datos muy grandes.

¡Se hizo famosa ganando al Jeopardy!

IBM Watson • ¿Qué es?



En concreto, cuando decimos Watson nos referimos a una serie de modelos muy complejos (poco bias), entrenados (o pre-entrenados) con bases de datos muy grandes que se encuentra disponible en IBM Cloud.

Otras alternativas son:

- Google Cloud Platform
- Microsoft Azure

IBM Watson • ¿Cómo usarlo?



Vamos a seguir los pasos detallados [en la plataforma de Acámica](#), y crearnos un usuario de IBM Cloud.

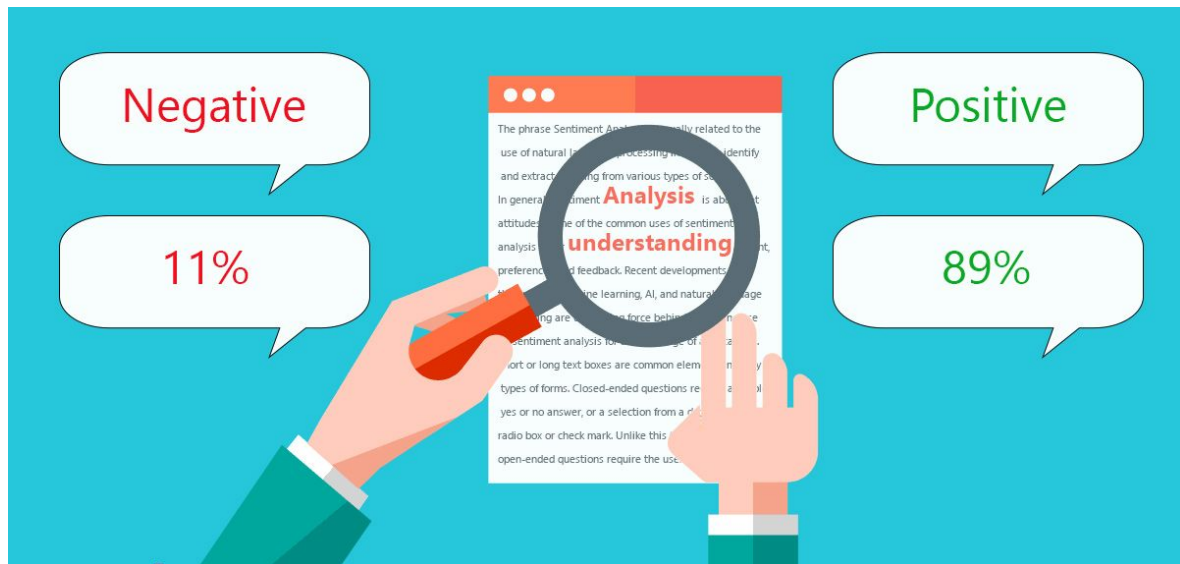
Debemos tener a mano [la documentación](#) para saber cómo conectarnos y usar las funciones disponibles.

Sentiment Analysis



Sentiment Analysis

Vamos a usar Watson para determinar si el contenido de un cuerpo de texto es positivo o negativo.

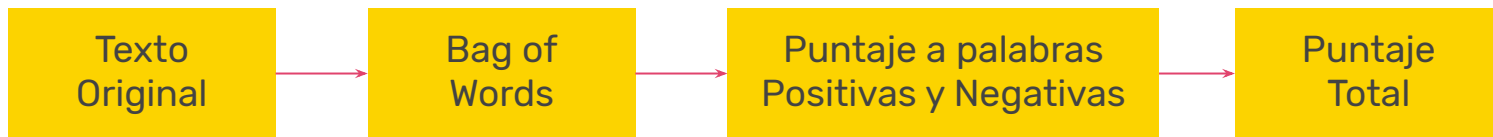


¿Cómo hacer Sentiment Analysis?



Sentiment Analysis • ¿Cómo hacerlo?

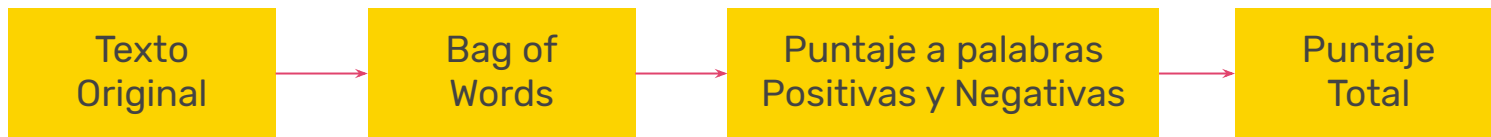
Está muy relacionado a lo que vimos de NLP. Una primera aproximación podría ser:



A pesar de ser muy simple, esto funciona (relativamente bien).

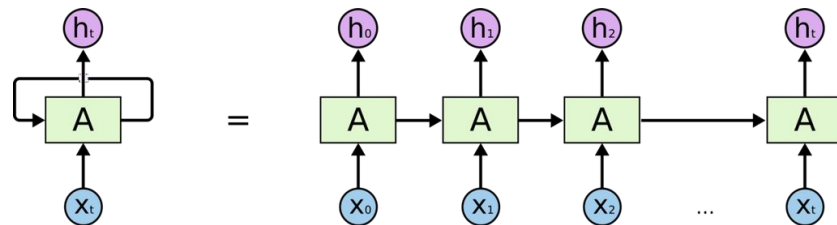
Sentiment Analysis • ¿Cómo hacerlo?

Está muy relacionado a lo que vimos de NLP. Una primera aproximación podría ser:



A pesar de ser muy simple, esto funciona (relativamente bien).

En el caso de Watson, se utilizan modelos más complejos de **Redes Recurrentes**. Esto permite tener en cuenta el orden de las palabras (estructura del texto) y no solo su aparición.



Hands-on training



Hands-on training

DS_Encuentro_44_Watson.ipynb



A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver fork are visible, though they are out of focus. The overall lighting is soft and even, highlighting the textures of the coffee and the smooth surface of the cup.

¡BREAK!



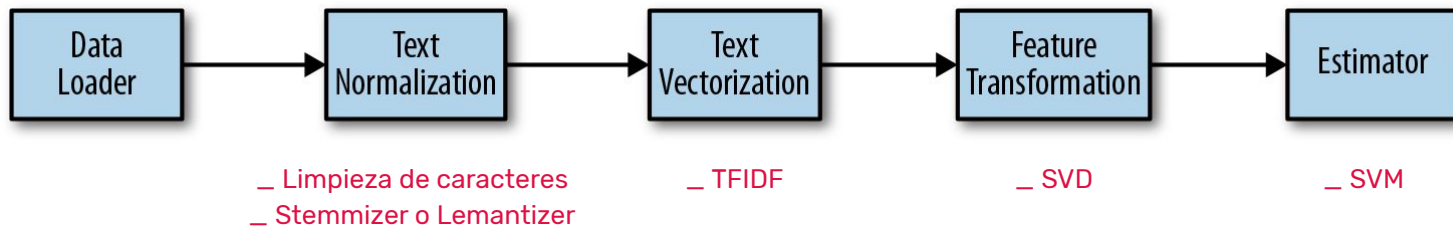
Pipelines de Scikit Learn



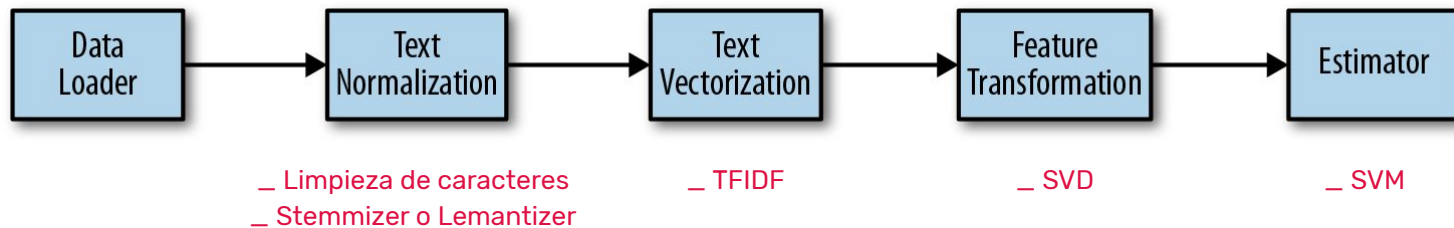
Workflow de trabajo

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



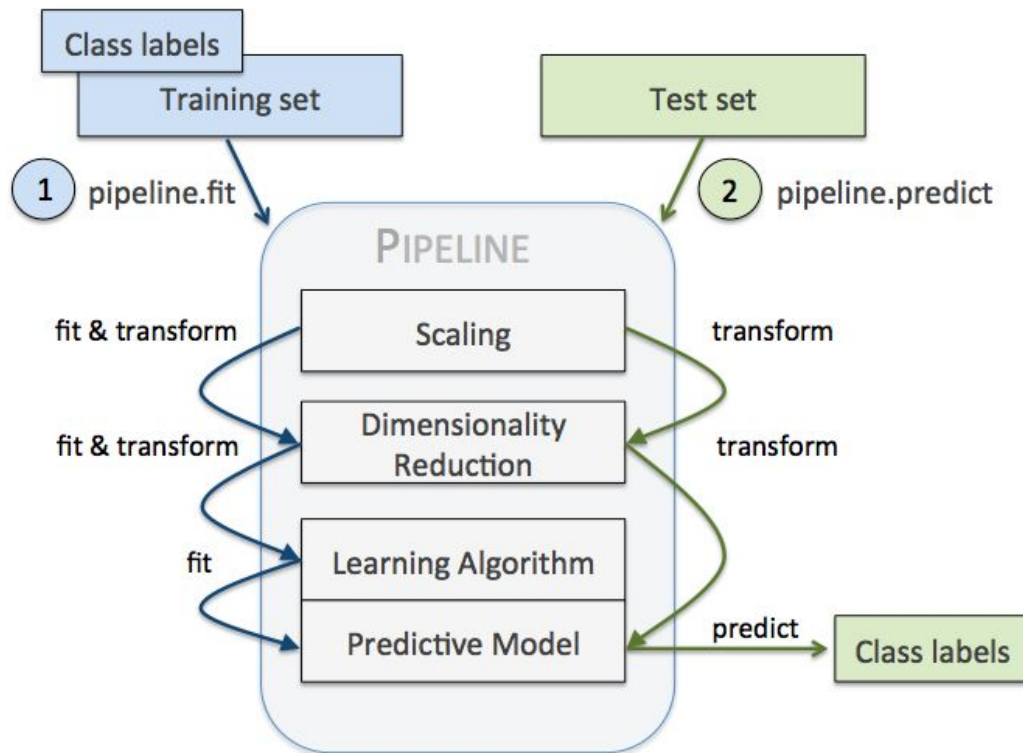
Notemos que tanto los datos del **Training Set** como los del **Test Set** deben realizar este mismo recorrido.

¿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)
```


Hands-on training

DS_Encuentro_44_Pipelines_y_Deploy.ipynb



Para la próxima

1. Terminar de ver los videos de puesta en producción.
2. Completar el notebook de hoy de Watson.
3. Terminar la entrega 06 si aún no lo hicieron.

ACÀMICA