

Proyecto MLOps: Parte FastAPI

1. Instalacion de dependencias desde el shell:

```

Downloading networkx-3.5-py3-none-any.whl.metadata (6.3 kB)
Requirement already satisfied: Jinja2 in c:\Users\orid\appdata\local\programs\python\python313\lib\site-packages (from torch) (3.1.6)
Requirement already satisfied: MarkupSafe in c:\Users\orid\appdata\local\programs\python\python313\lib\site-packages (from Jinja2->torch) (3.0.2)
Collecting setuptools (from torch)
  Downloading setuptools-80.9.0-py3-none-any.whl.metadata (6.6 kB)
Requirement already satisfied: mpmath<1.4, >=1.1.0 in c:\Users\orid\appdata\local\programs\python\python313\lib\site-packages (from sympy>=1.13.3->torch) (1.3.0)
Requirement already satisfied: MarkupSafe in c:\Users\orid\appdata\local\programs\python\python313\lib\site-packages (from Jinja2->torch) (3.0.2)
Downloading torch-2.7.1-cp313-cp313-win_amd64.whl (216.1 MB)
216.1/216.1 MB 10.9 MB/s eta 0:00:00
Downloading networkx-3.5-py3-none-any.whl (2.0 MB)
2.0/2.0 MB 9.7 MB/s eta 0:00:00
Downloading setuptools-80.9.0-py3-none-any.whl (1.2 MB)
1.2/1.2 MB 19.0 MB/s eta 0:00:00
Installing collected packages: setuptools, networkx, torch
Successfully installed networkx-3.5 setuptools-80.9.0 torch-2.7.1

[notice] A new release of pip is available: 25.0.1 -> 25.1.1
[notice] To update, run: python.exe -m pip install --upgrade pip

C:\Users\orid\Desktop\Programacion\ML\Ops\mllops-main\Projecto>uvicorn main_hf_api:app --reload
INFO: Will watch for changes in these directories: ['C:\Users\orid\Desktop\Programacion\ML\Ops\mllops-main\Projecto']
INFO: Uvicorn running on http://127.0.0.1:8080 (Press CTRL+C to quit)
INFO: Started reload process [46608] using Setuptools
INFO: model was supplied, defaulted to distilbert/distilbert-base-uncased-finetuned-sst-2-english and revision 714e60f (https://huggingface.co/distilbert/distilbert-base-uncased-finetuned-sst-2-english).
INFO: Using a pipeline without specifying a model name and revision in production is not recommended.
INFO: hf_xet Storage is enabled for this repo, but the 'hf_xet' package is not installed. Falling back to regular HTTP download. For better performance, install the package with: 'pip install huggingface_hub[hf_xet]' or 'pip install hf_xet'
INFO: model.safetensors: 100% | 265M/268M [00:23<00:00, 11.0MB/s]
C:\Users\orid\Desktop\Programacion\ML\Ops\mllops-main\Projecto>python hf_api.py
INFO: UserWarning: 'huggingface_hub' cache-system uses symlinks by default to efficiently store duplicated files but your machine does not support them in C:\Users\orid\AppData\Local\Programs\Python\Python313\lib\site-packages\huggingface_hub\models--distilbert--distilbert-base-uncased-finetuned-sst-2-english. Caching files will still work but in a degraded version that might require more space on your disk. This warning can be disabled by setting the 'HF_HUB_DISABLE_SYMLINKS_WARNING' environment variable. For more details, see https://huggingface.co/docs/huggingface_hub/how-to-cache#limitations.
INFO: To support symlinks on Windows, you either need to activate Developer Mode or to run Python as an administrator. In order to activate developer mode, see this article: https://docs.microsoft.com/en-us/windows/apps/get-started/enable-your-device-for-development
INFO: warnings.warn(message)
INFO: tokenizer_config.json: 100% | 48.0/48.0 [00:00<00:00, 284KB/s]
INFO: vocab.txt: 232kB [00:00, 11.7MB/s]
INFO: device set to use cpu
INFO: model was supplied, defaulted to sshleifer/distilbart-cnn-12-6 and revision a48f83e (https://huggingface.co/sshleifer/distilbart-cnn-12-6).
INFO: Using a pipeline without specifying a model name and revision in production is not recommended.
INFO: config.json: 1.80kB [00:00, 11.0MB/s]
C:\Users\orid\Desktop\Programacion\ML\Ops\mllops-main\Projecto>python hf_api.py
INFO: UserWarning: 'huggingface_hub' cache-system uses symlinks by default to efficiently store duplicated files but your machine does not support them in C:\Users\orid\AppData\Local\Programs\Python\Python313\lib\site-packages\huggingface_hub\models--sshleifer--distilbart-cnn-12-6. Caching files will still work but in a degraded version that might require more space on your disk. This warning can be disabled by setting the 'HF_HUB_DISABLE_SYMLINKS_WARNING' environment variable. For more details, see https://huggingface.co/docs/huggingface_hub/how-to-cache#limitations.
INFO: To support symlinks on Windows, you either need to activate Developer Mode or to run Python as an administrator. In order to activate developer mode, see this article: https://docs.microsoft.com/en-us/windows/apps/get-started/enable-your-device-for-development
INFO: warnings.warn(message)
INFO: pytorch_model.bin: 100% | 1.22G/1.22G [01:45<00:00, 11.6MB/s]
INFO: tokenizer_config.json: 100% | 26.0/26.0 [00:00<00:00, 224KB/s]
INFO: vocab.txt: 890kB [00:00, 14.6MB/s]
INFO: merges.txt: 456kB [00:00, 2.40MB/s]
INFO: | 0.00/26.0 [00:00<?, ?B/s]
INFO: model.safetensors: 3% | 31.5M/1.22G [00:04<02:39, 7.46MB/s]
INFO: Device set to use cpu
INFO: Started server process [25884]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: | 1.22G/1.22G [01:47<00:00, 11.4MB/s]
INFO: model.safetensors: 100%
INFO: 127.0.0.1:59803 - "GET /docs HTTP/1.1" 200 OK
INFO: 127.0.0.1:59803 - "GET /openapi.json HTTP/1.1" 200 OK

```

2. Visualizacion de Endpoints desde web:



FastAPI con pipelines HF 1.0 OAS 3.1

/openapi.json

default

GET	/ Root
GET	/hello Hello
GET	/add Add
GET	/sentiment Sentiment
GET	/summary Summary
GET	/length Text Length

- Summarize:

GET

/summary Summary

Cancel

Parameters

Name	Description
text <small>* required</small>	Text to summarize
string (query)	<div>El deployment de algoritmos en MLOps (Mac</div> <div>maxLength: 1000</div>

ExecuteClear

Responses

Curl

```
curl -X 'GET' \
  'http://127.0.0.1:8000/summary?text=ElX20deploymentX20deX20algoritmosX20enX20MLOpsX20enX20MachineX20LearningX20operationsX29X20seX20refiereX20alX20procesoX20deX20ponerX20unX20modeloX20deX20MachineX20LearningX20
  -H 'accept: application/json'
```

Request URL

```
http://127.0.0.1:8000/summary?
text=ElX20deploymentX20deX20algoritmosX20enX20MLOpsX20enX20MachineX20LearningX20operationsX29X20seX20refiereX20alX20procesoX20deX20ponerX20unX20modeloX20deX20MachineX20LearningX20enX20producciX3B3nX2CK20es
X20decirX2CK20hacerX20queX20funcioneX20enX20unX20entornoX20realX20dondeX20usuariosX20oX20sistemasX20puedanX20accederX20aX20elX3A9IX20yX20usarloX20paraX20hacerX20predicciones.X20X20F0X9FX94A7X20X2CK20F0uX2C
3X20X20implicarX3FX20EmpaquetadoX20delX20modeloX3AX20convertirX20elX20modeloX20entrenadoX20X20porX20ejemploX2CK20unX20archivoX20.pk1X20X20.ptX29X20enX20unX20componenteX20reutilizable.X20X20exposiciX3B3nX2
20comoX20servicioX3AX20generalmentex20mediantex20unaX20APIX20RESTX20X20porX20ejemploX2CK20usandoX20FastAPIX20oX20FlaskX29.X20X20IntegraciX3B3nX20conX20aplicacionesX3AX20ConectarloX20aX20unaX20appX20webX2
CK20X2CK20X3B3villX2CK20oX20unX20sistemaX20deX20backend.X20X20MonitoreoX20enX20tiempoX20realX3AX20EvaluarX20suX20rendimientoX20enX20producciX3B3nX20X20porX20ejemploX2CK20detectarX20driftX20enX20losX20datosX2
9.X20X20AutomatizaciX3B3nX20delX20cicloX20deX20vidaX3AX20usarX20herramientasX20comoX20DockerX2CK20CX20FX20CX20KubernetesX2CK20etc.
```

Server response

Code	Details
200	<div>Response body</div> <div><pre>{ "summary": " El deployment de algoritmos en MLOps (Machine Learning Operations) se refiere al proceso de poner un modelo de Machine Learning en producción . Modelo ponera en un entorno real do nde usuarios o sistemas puedan acceder a él y usarlo para hacer predicciones ." }</pre></div> <div>Response headers</div> <div><pre>content-length: 274 content-type: application/json date: Mon, 07 Jul 2025 15:13:59 GMT server: uvicorn</pre></div>

- Add:

GET

/addAdd

Parameters

Name	Description
<div><div>a</div><div><div>*</div><div>required</div></div></div>	
<div>integer</div> <div>(query)</div>	<input type="text" value="1"/>
<div><div>b</div><div><div>*</div><div>required</div></div></div>	
<div>integer</div> <div>(query)</div>	<input type="text" value="2"/>

Execute

Responses

Curl

```
curl -X 'GET' \
'http://127.0.0.1:8000/add?a=1&b=2' \
-H 'accept: application/json'
```

Request URL

```
http://127.0.0.1:8000/add?a=1&b=2
```

Server response

Code	Details
200	<div><div>Response body</div><pre>{ "result": 3 }</pre></div>

- Sentiment:

GET **/sentiment** Sentiment

Parameters

Name	Description
text * required string (query)	Text to analyze sentiment <div>esta parte del bootcamp me gusta mucho</div>

Execute

Responses

Curl

```
curl -X 'GET' \
'http://127.0.0.1:8000/sentiment?text=esta%20parte%20del%20bootcamp%20me%20gusta%20mucho' \
-H 'accept: application/json'
```

Request URL

```
http://127.0.0.1:8000/sentiment?text=esta%20parte%20del%20bootcamp%20me%20gusta%20mucho
```

Server response

Code	Details
200	<div>Length: 47</div> <div>Response body</div> <pre>{ "label": "POSITIVE", "score": 0.9615129232406616 }</pre> <div>Response headers</div> <pre>content-length: 47 content-type: application/json date: Mon, 07 Jul 2025 15:12:27 GMT server: uvicorn</pre>

- Length:

GET **/length** Text Length

Parameters

Name	Description
text * required string (query)	Text to count length <div>**¿Qué es Machine Learning? ->** Es una ra</div>

Execute

Responses

Curl

```
curl -X 'GET' \  
  'http://127.0.0.1:8000/length?text=%2A%2A%C2%BFQu%C3%A9%20es%20Machine%20Learnin%3E%2A%2A%20Es%20una%20rama%20de%20la%20inteligencia%20artificial%20que%20permite%20expl%C3%ADcitamente%20programadas%20para%20cada%20tarea.%20En%20vez%20de%20escor%20relaciones%2A%2A%20para%20tomar%20decisiones%20o%20hacer%20predicciones.%20e%20los%20datos%2C%20sin%20ser%20programados%20expl%C3%ADcitamente%20para%20el%20Modelo%20de%20Aprendizaje%20autom%C3%A1tico%20m%C3%A1s%20apropiado.%2A%2A%2A'
```

Request URL

```
http://127.0.0.1:8000/length?text=%2A%2A%C2%BFQu%C3%A9%20es%20Machine%20Learnin%3E%2A%2A%20Es%20una%20rama%20de%20la%20inteligencia%20artificial%20que%20permite%20expl%C3%ADcitamente%20programadas%20para%20cada%20tarea.%20En%20vez%20de%20escor%20relaciones%2A%2A%20para%20tomar%20decisiones%20o%20hacer%20predicciones.%20e%20los%20datos%2C%20sin%20ser%20programados%20expl%C3%ADcitamente%20para%20el%20Modelo%20de%20Aprendizaje%20autom%C3%A1tico%20m%C3%A1s%20apropiado.%2A%2A%2A'
```

Server response

Code	Details
200	<div>Response body</div> <pre>{ "length": 646 }</pre>

WINE MODEL:

Wine Classifier API 1.0 OAS 3.1

[/openapi.json](#)

Predict wine class from chemical features

default

GET / Read Root

POST /predict Predict

Responses

Curl

```
curl -X 'GET' \
  'http://127.0.0.1:8000/' \
  -H 'accept: application/json'
```

Request URL

```
http://127.0.0.1:8000/
```

Server response

Code	Details
------	---------

200	
-----	--

Response body

```
{
  "message": "Welcome to the Wine Classifier API "
}
```

Response headers

```
content-length: 49
content-type: application/json
date: Mon, 07 Jul 2025 15:48:24 GMT
server: uvicorn
```

Responses

```
INFO:      Waiting for application startup.
INFO:      Application startup complete.
INFO: 127.0.0.1:61458 - "GET /docs HTTP/1.1" 200 OK
INFO: 127.0.0.1:61458 - "GET /openapi.json HTTP/1.1" 200 OK
INFO: 127.0.0.1:61551 - "POST /predict HTTP/1.1" 200 OK
INFO: 127.0.0.1:61630 - "POST /predict HTTP/1.1" 200 OK
```

- Clase 1 y clase 2:

Request body required

Edit Value | Schema

```
{  "alcohol": 1,  "malic_acid": 2,  "ash": 3,  "alkalinity_of_ash": 4,  "magnesium": 5,  "total_phenols": 0,  "flavanoids": 8,  "nonflavanoid_phenols": 0,  "proanthocyanins": 1,  "color_intensity": 1,  "hue": 1,  "od280_od315_of_diluted_wines": 0,  "proline": 0}
```

Execute

Responses

Curl

```
curl -X 'POST' \  'http://127.0.0.1:8000/predict' \  -H 'accept: application/json' \  -H 'Content-Type: application/json' \  -d '{  "alcohol": 1,  "malic_acid": 2,  "ash": 3,  "alkalinity_of_ash": 4,  "magnesium": 5,  "total_phenols": 0,  "flavanoids": 8,  "nonflavanoid_phenols": 0,  "proanthocyanins": 1,  "color_intensity": 1,  "hue": 1,  "od280_od315_of_diluted_wines": 0,  "proline": 0}'
```

Request URL

http://127.0.0.1:8000/predict

Server response

Code	Details
200	<div><div>Response body</div><div><pre>{ "predicted_class": "class_1"}</pre></div><div>Response headers</div><div><pre>content-length: 29 content-type: application/json date: Mon, 07 Jul 2025 15:35:32 GMT server: uvicorn</pre></div></div>

```
{  "alcohol": 14,  "malic_acid": 0,  "ash": 0,  "alkalinity_of_ash": 0,  "magnesium": 0,  "total_phenols": 0,  "flavanoids": 0,  "nonflavanoid_phenols": 0,  "proanthocyanins": 0,  "color_intensity": 14,  "hue": 0,  "od280_od315_of_diluted_wines": 0,  "proline": 0}
```

Execute

Responses

Curl

```
curl -X 'POST' \  'http://127.0.0.1:8000/predict' \  -H 'accept: application/json' \  -H 'Content-Type: application/json' \  -d '{  "alcohol": 14,  "malic_acid": 0,  "ash": 0,  "alkalinity_of_ash": 0,  "magnesium": 0,  "total_phenols": 0,  "flavanoids": 0,  "nonflavanoid_phenols": 0,  "proanthocyanins": 0,  "color_intensity": 14,  "hue": 0,  "od280_od315_of_diluted_wines": 0,  "proline": 0}'
```

Request URL

http://127.0.0.1:8000/predict

Server response

Code	Details
200	<div><div>Response body</div><div><pre>{ "predicted_class": "class_2"}</pre></div></div>