



códigofacilito

## — Agenda

- ¿Qué es Azure - AutoML?
- Hands On AutoML
- ¿Qué es MLFlow?
- HAndsOn MLFlow

Irving Uribe - ML Engineer





>\_



Búsqueda del mejor modelo de  
clasificación con Azure - AutoML

<https://learn.microsoft.com/es-es/training/modules/find-best-classification-model-automated-machine-learning/>

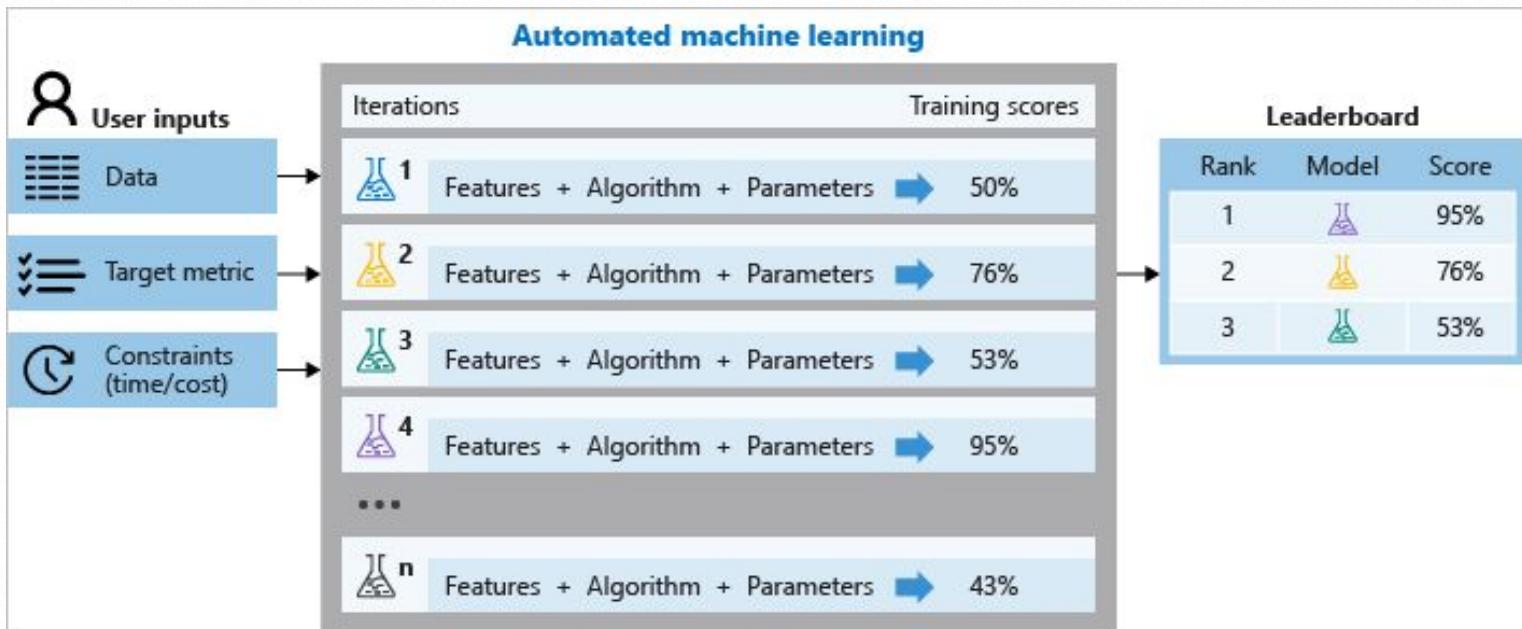




## ¿Qué es el Aprendizaje automático automatizado de Azure (Auto-ML)?



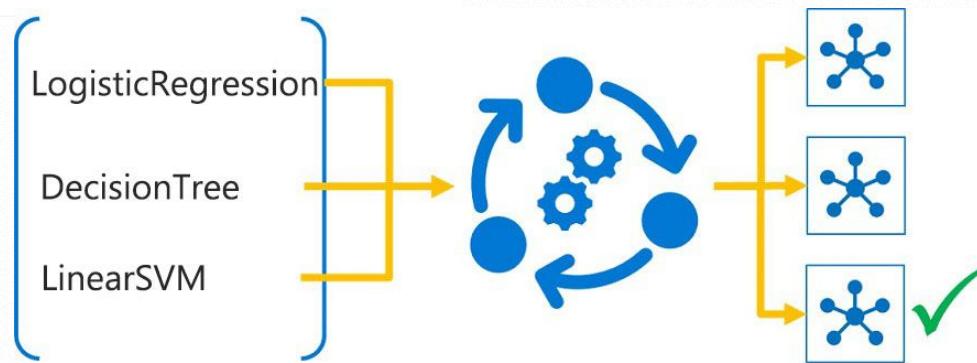
Servicio de Azure que permite, de manera **automatizada** crear y **comparar** modelos de ML, agilizando así el desarrollo





# Explicación de los modelos (comparativa)

Algorithm name	Explained	Responsible AI	Accuracy ↓
VotingEnsemble	<a href="#">View explanation</a>		0.95300
StackEnsemble			0.95280
MaxAbsScaler, LightGBM			0.95180
MaxAbsScaler, XGBoostClassifier			0.95180
MaxAbsScaler, ExtremeRandomTrees			0.83740





## ¿En dónde estamos en la certificación dp-100?



### Exploración de datos y entrenamiento de modelos (de 35 a 40 %)

#### Uso del aprendizaje automático automatizado para explorar modelos óptimos

- Uso del aprendizaje automático automatizado para datos tabulares
- Uso del aprendizaje automático automatizado para Computer Vision
- Uso del aprendizaje automático automatizado para el procesamiento de lenguaje natural
- Selección y comprensión de las opciones de entrenamiento, incluido el preprocesamiento y los algoritmos
- Evaluación de una ejecución de aprendizaje automático automatizado, incluidas las directrices de inteligencia artificial responsable





## Pasos para crear modelos con Auto ML (Ejemplo PLN)



1.- Crear y configurar un Data Asset

<https://learn.microsoft.com/en-us/azure/machine-learning/how-to-create-data-assets?view=azureml-api-2&tabs=cli>

2.- Iniciar entorno de ejecución

3.- Conexión con el espacio de trabajo

4.- Preparación de la Data (Transformación)

5.- Configurar el job de ejecución

6.- Ejecutar el ML Job

7.- Explorar resultados





# 1.- Crear y configurar un Data Asset

## Data

Data assets   Datastores   Dataset monitors PREVIEW   Data import PREVIEW

Data assets are immutable references to your data that can be created from datastores, local cannot be deleted, but you can up-version or archive them for easy referencing and reuse in delete the data asset and all metadata. [Learn more about data assets](#)

+ Create   Refresh   Archive   Reset view   Create data asset

1 Data type  
2 Data source

Set the name and type for your data asset

Name \*  (G)

Description  (G)

Type \*  (i)



# 1.- Crear y configurar un Data Asset

## Choose a source for your data asset

Choose the data source you want to create your asset from. A data source can be from a local storage location on your computer, from an attached datastore, from Azure storage, or from a publicly available web location.

**From Azure storage**

Create a data asset from registered data storage services including Azure Blob Storage, Azure file share, and Azure Data Lake.

**From local files**

Create a data asset by uploading files from your local drive.

**From SQL databases**

Create a dataset from Azure SQL database and Azure PostGreSQL database.

**From web files**

Create a data asset from a single file located at a public web URL.



- ✓ Data type
- ✓ Data source
- 3 Destination storage type
- 4 File or folder selection
- 5 Settings
- 6 Schema
- 7 Review



# 1.- Crear y configurar un Data Asset



## Analisis\_Sentimientos

Version: 1 (latest) ▾



Details   Consume   **Explore**   Models   Jobs



Refresh



Generate profile

366M

Preview   Profile

Number of columns: 3   Number of rows: 50 (of 9999)

original	sentimiento	reseña
One of the ...	positive	Uno de los otros críticos han mencionado que después de ver sólo 1 episodio Oz estará enganchado. Tienen razón, ya que esto es exacta...
A wonderf...	positive	Una producción pequeña y maravillosa.     La técnica de filmación es muy unassumming- muy de moda en tiempo pasado de la...
I thought t...	positive	Pensé que esto era una maravillosa manera de pasar el tiempo en un fin de semana muy caluroso de verano, sentado en el teatro al aire a...
Basically th...	negative	Básicamente hay una familia en la que un niño pequeño (Jake) piensa que hay un zombi en su armario y sus padres están luchando todo e...
Petter Matt...	positive	Petter Mattei de "amor en los tiempos del dinero" es una película visualmente impresionante de ver. El Sr. Mattei nos ofrece un vivo retrat...
Probably m...	positive	Probablemente mi película favorita de todos los tiempos, una historia de desinterés, sacrificio y dedicación a una causa noble, pero no es ...
I sure woul...	positive	Estoy seguro que le gustaría ver una resurrección de una serie de hasta Seahunt anticuado, con la tecnología que tienen hoy sería traer de...
This show ...	negative	Este espectáculo fue una increíble, fresca y idea innovadora en los años 70 cuando se emitió por primera vez. Los primeros 7 u 8 años era...
Encourage...	negative	Alentados por los comentarios positivos sobre esta película de aquí me quedé con ganas de ver esta película. Mal error. He visto de 950 p...
If you like ...	positive	Si te gusta la risa desgarradora original que va a gustar esta película. Si eres joven o viejo, entonces le encantará esta película, diablos, incl...

## 2.- Iniciar entorno de ejecución (En un notebook)

### Notebooks

Files Samples



- > Logs
- ▽ Users
- ▽ ayuda
- ▽ AutoML
  - 📄 InferenciaModelo\_Diabetes.ipynb
  - 📄 PruebaAutoML.ipynb
  - 📄 PruebaAutoML\_Diabetes.ipynb
- ☆ ○ PruebaAutoML\_PLN.ipynb ...
- > MLFlow

Welcome to the Azure Machine Learning terminal

Enter "git clone [url]" to clone a repo

Enter "git --help" to learn about Git CLI. To learn more about integrating Git with the Azure Machine Learning terminal, navigate here:

<https://learn.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration#clone-git-repositories-into-your-workspace-file-system>

Enter "az ml --help" to learn about Azure ML CLI v2

Note: Use "az login --identity" instead of "az login" to avoid device code authentication

(azureml\_py38) azureuser@instanciaautomltest:~/cloudfiles/code/Users/ayuda\$ █



Edit in VS Code	Compute: InstanciaAutoMLTest - Starting	...	No kernel connected
Edit in VS Code	Compute: InstanciaAutoMLTest - Running	...	Python 3.10 - SDK v2

397M



## 2.- Iniciar entorno de ejecución (En un notebook)

>\_

```
Note: Use "az login --identity" instead of "az login" to avoid device code authentication
```

```
(azureml_py38) azureuser@instanciaautomltest:~/cloudfiles/code/Users/ayuda/AutoML$ conda env list
# conda environments:
#
base                  /anaconda
azureml_py310_sdkv2    /anaconda/envs/azureml_py310_sdkv2
azureml_py38           * /anaconda/envs/azureml_py38
azureml_py38_PT_TF     /anaconda/envs/azureml_py38_PT_TF
jupyter_env            /anaconda/envs/jupyter_env
```

```
(azureml_py38) azureuser@instanciaautomltest:~/cloudfiles/code/Users/ayuda/AutoML$ conda activate azureml_py310_sdkv2
(azureml_py310_sdkv2) azureuser@instanciaautomltest:~/cloudfiles/code/Users/ayuda/AutoML$ pip install nltk
Collecting nltk
  Using cached nltk-3.8.1-py3-none-any.whl (1.5 MB)
Requirement already satisfied: click in /anaconda/envs/azureml_py310_sdkv2/lib/python3.10/site-packages (from nltk) (8.0.4)
Requirement already satisfied: joblib in /anaconda/envs/azureml_py310_sdkv2/lib/python3.10/site-packages (from nltk) (1.2.0)
Collecting regex>=2021.8.3 (from nltk)
  Downloading regex-2024.5.10-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (774 kB)
   774.1/774.1 kB 19.0 MB/s eta 0:00:00
Requirement already satisfied: tqdm in /anaconda/envs/azureml_py310_sdkv2/lib/python3.10/site-packages (from nltk) (4.55.0)
```



### 3.- Conexión con el espacio de trabajo

>\_

Azure ML CLI v2

Note: Use "az login --identity" instead of "az login" to avoid device code authentication

```
(azureml_py38) azureuser@instanciaautomltest:~/cloudfiles/code/Users/ayuda/AutoML$ pip install mltable
Collecting mltable
  Downloading mltable-1.6.1-py3-none-any.whl (189 kB)
    |██████████| 189 kB 4.9 MB/s
Requirement already satisfied: pyyaml<7.0.0,>=5.1.0 in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (from mltable) (6.0)
Requirement already satisfied: azure-mgmt-core<2.0.0,>=1.3.0 in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (from mltable) (1.4.0)
Requirement already satisfied: msrest>=0.6.18 in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (from mltable) (0.7.1)
Requirement already satisfied: PyJWT<3.0.0 in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (from mltable) (2.4.0)
Requirement already satisfied: azure-core!=1.22.0,<2.0.0,>=1.8.0 in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (from mltable) (1.26.4)
Requirement already satisfied: cryptography!=1.9,!>2.0.*,!>2.1.*,!>2.2.* in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (from mltable) (38.0.4)
```





### 3.- Conexión con el espacio de trabajo



## Connect to your Worksace

```
1 # Librerías para conexión con el Workspace desde cuenta Azure
2 from azure.identity import DefaultAzureCredential, InteractiveBrowserCredential
3 from azure.ai.ml import MLClient
4
5 try:
6     # Se intenta obtener el token por defecto desde el workspace actual
7     credential = DefaultAzureCredential()
8     credential.get_token("https://management.azure.com/.default")
9 except Exception as ex:
10     # En caso de que NO funcionen las credenciales por defecto, se abrirá el
11     # InteractiveBrowserCredential (para autenticarte desde el navegador)
12     credential = InteractiveBrowserCredential()
```



```
1 # Cree una instancia cliente para administrar el Worspace
2 ml_client = MLClient.from_config(credential=credential)
```



## 4.- Preparación de la Data



Welcome to the Azure Machine Learning terminal

Enter "git clone [url]" to clone a repo

Enter "git --help" to learn about Git CLI. To learn more about integrating Git with the Azure Machine Learning terminal, navigate here:

<https://learn.microsoft.com/en-us/azure/machine-learning/concept-train-model-git-integration#clone-git-repositories-into-your-workspace-file-system>

Enter "az ml --help" to learn about

Azure ML CLI v2

Note: Use "az login --identity" instead of "az login" to avoid device code authentication

```
(azureml_py38) azureuser@instanciaautomltest:~/cloudfiles/code/Users/ayuda$ pip install nltk
Requirement already satisfied: nltk in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (3.8.1)
Requirement already satisfied: tqdm in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (from nltk) (4.65.0)
Requirement already satisfied: click in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (from nltk) (8.1.3)
Requirement already satisfied: regex>=2021.8.3 in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (from nltk)
(2023.5.5)
Requirement already satisfied: joblib in /anaconda/envs/azureml_py38/lib/python3.8/site-packages (from nltk) (1.2.0)
(azureml_py38) azureuser@instanciaautomltest:~/cloudfiles/code/Users/ayuda$ █
```





## 4.- Preparación de la Data

```
1 from azure.ai.ml.constants import AssetTypes
2 from azure.ai.ml import Input
3 import mltable
4
5 # Crea un Dataset basado en los archivos en la carpeta de datos local
6 nombre_tabla = "Analisis_Sentimientos"
7 version = "1"
8
9 # Cargamos los datos desde el Asset de datos y los leemos como DataFrame
10 data_asset = ml_client.data.get(nombre_tabla, version=version)
11 tbl = mltable.load(f'azureml://{{data_asset.id}}')
12 df = tbl.to_pandas_dataframe()
13 df
```

		original	sentimiento	reseña
0	One of the other reviewers has mentioned that ...	positive	Uno de los otros críticos han mencionado que d...	
1	A wonderful little production.   The...	positive	Una producción pequeña y maravillosa.   <...	
2	I thought this was a wonderful way to spend ti...	positive	Pensé que esto era una maravillosa manera de p...	
3	Basically there's a family where a little boy ...	negative	Básicamente hay una familia en la que un niño ...	
4	Petter Mattei's "Love in the Time of Money" is...	positive	Petter Mattei de "amor en los tiempos del dine...	
...	...	...	...	
9994	First off, this is the worst movie I've ever s...	negative	En primer lugar, esta es la peor película que ...	
9995	Fun, entertaining movie about WWII German spy ...	positive	Diversión, entretenimiento película sobre la S...	
9996	Give me a break. How can anyone say that this ...	negative	Dáme un respiro. ¿Cómo puede alguien decir que...	
9997	This movie is a bad movie. But after watching ...	negative	Esta película es una película mala. Pero despue...	
9998	This is a movie that was probably made to ente...	negative	Esta es una película que probablemente fue hec...	





## 5.- Configurar el Job de ejecución



### Configure ML Job

```
1  from azure.ai.ml.constants import AssetTypes
2  from azure.ai.ml import Input
3  import mltable
4
5  # creates a dataset based on the files in the local data folder
6  training_data_input = Input(type=AssetTypes.MLTABLE, path="azureml:sentiments:1")
```



```
1  from azure.ai.ml import automl
2
3  # configure the classification job
4  classification_job = automl.classification(
5      compute="InstanciaAutoMLTest",                      # Nombre del cluster de cómputo
6      experiment_name="AnalisisSentimientosExperimento", # Nombre que le quieres dar al experimento
7      training_data=training_data_input,                  # Data / Dataset de input
8      target_column_name="sentiment",                   # Columna objetivo de la clasificación
9      primary_metric="accuracy",                        # Metrica de evaluación
10     n_cross_validations=5,                           # Número de validaciones cruzadas
11     enable_model_explainability=True                 # Permite las explicaciones de las predicciones
12 )
```





## 6.- Ejecutar el ML Job

>\_

### Run an automated machine learning job

```
1 # Submit the AutoML job
2 returned_job = ml_client.jobs.create_or_update(
3     classification_job
4 )
5
6 # submit the job to the backend
7 aml_url = returned_job.studio_url
8 print("Monitor your job at", aml_url)
```

]



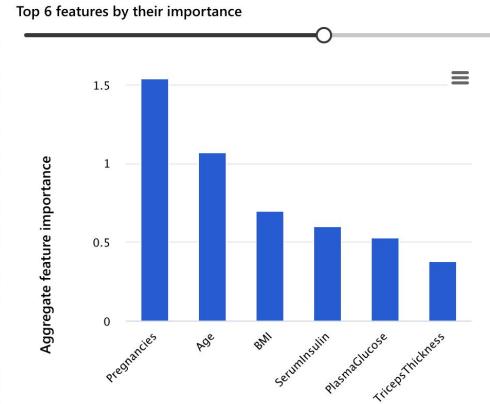
Monitor your job at [https://ml.azure.com/runs/wheat\\_pot\\_1prbtty2h0?wsid=/subscriptions/4e4f15fc-cca7-492a-935d-d9171a888c0d/resourcegroups/rg-dp100-lb49c6ecb55584a48b0/workspaces/taller\\_auto\\_ml&tid=500f314c-66f1-4fa8-9741-e645a10a0d52](https://ml.azure.com/runs/wheat_pot_1prbtty2h0?wsid=/subscriptions/4e4f15fc-cca7-492a-935d-d9171a888c0d/resourcegroups/rg-dp100-lb49c6ecb55584a48b0/workspaces/taller_auto_ml&tid=500f314c-66f1-4fa8-9741-e645a10a0d52)





## 7.- Explorar resultados

VotingEnsemble	<a href="#">View explanation</a>	0.95300
StackEnsemble		0.95280
MaxAbsScaler, LightGBM		0.95180
MaxAbsScaler, XGBoostClassifier		0.95180
MaxAbsScaler, ExtremeRandomTrees		0.83740





## 8 (Adicional).- Inferencia del modelo

&gt;\_

```
import urllib.request
import json
import os
import ssl

def allowSelfSignedHttps(allowed):
    # bypass the server certificate verification on client side
    if allowed and not os.environ.get('PYTHONHTTPSVERIFY', '') and getattr(ssl, '_'
        | | _create_default_https_context = ssl._create_unverified_context

allowSelfSignedHttps(True) # this line is needed if you use self-signed certifica

# Request data goes here
# The example below assumes JSON formatting which may be updated
# depending on the format your endpoint expects.
# More information can be found here:
# https://docs.microsoft.com/azure/machine-learning/how-to-deploy-advanced-entry-
data = {
    "input_data": {
        "columns": [
            "PatientID",
            "Pregnancies",
            "PlasmaGlucose",
            "DiastolicBloodPressure",
            "TricepsThickness",
            "Insulin",
            "BMI",
            "Age"
        ]
    }
}
```



Con base en los siguientes datos:

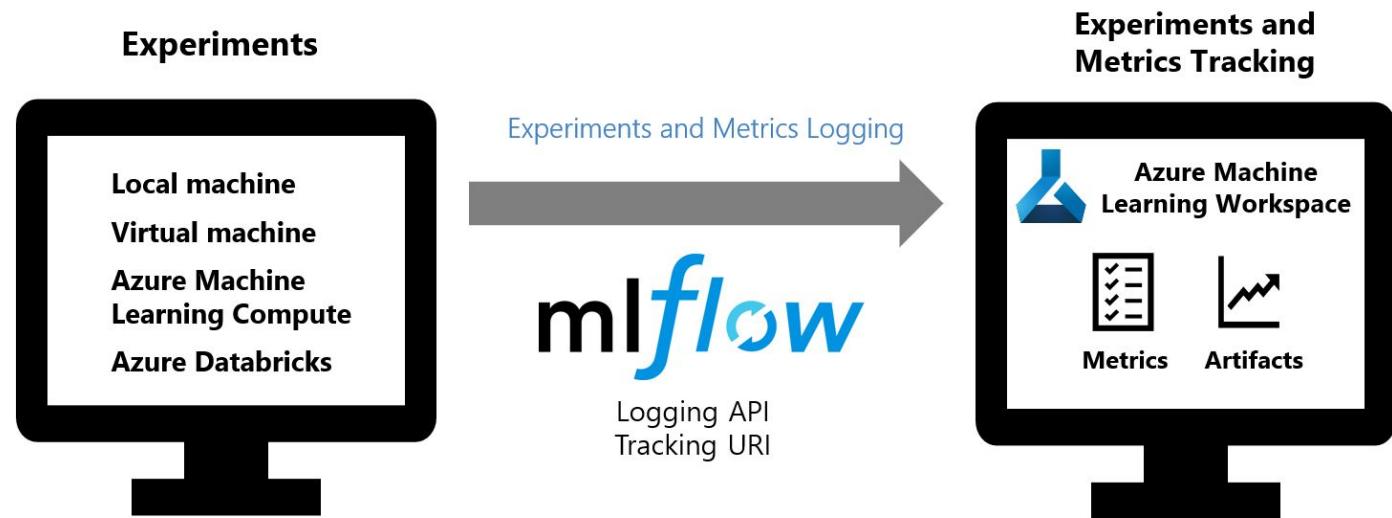
Embarazos: 0  
Glucosa: 171  
PresionDiastolica: 80  
GrosorTriceps: 34  
InsulinaSerica: 23  
IMC: 24.5  
PedigreeDiabetes: 0.22  
Edad: 21  
El diagnóstico es:  
No Diabetes



## ¿Qué es Azure ML - Flow?



Imagina que como científico de datos te es extrema importancia tener **reproducibilidad** y un **track** en todo el **entrenamiento** del modelo

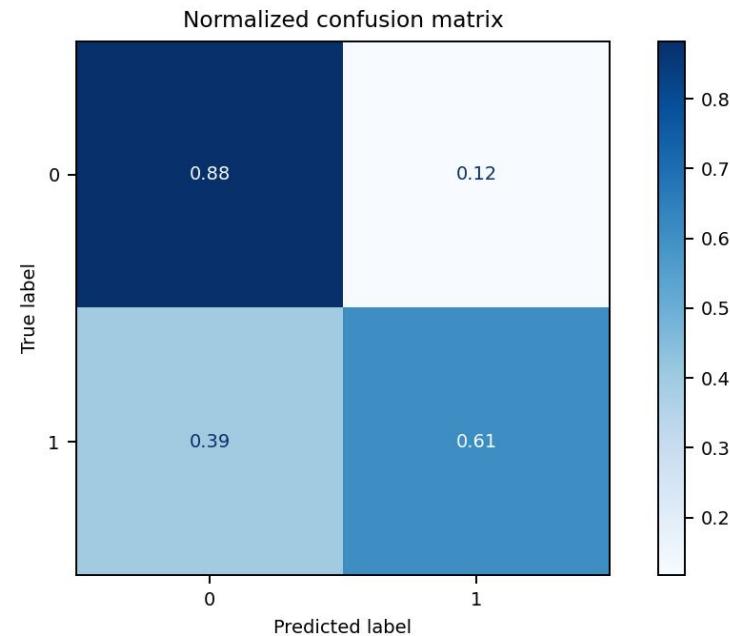
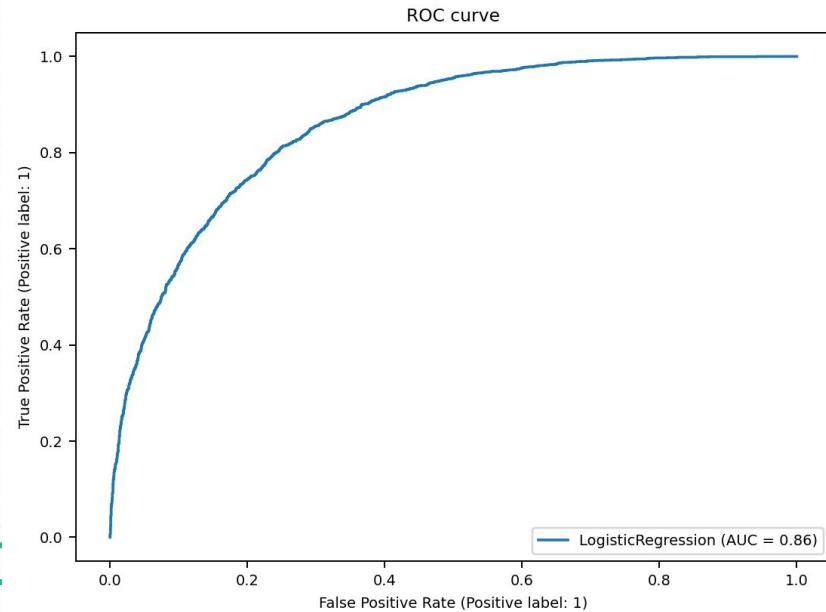




## ¿Qué puedo medir con ML Flow?



Puedes encontrar el .pkl del modelo entrenado, métricas como la AUC, ROC, o matriz de confusión





códigofacilito

## — Agenda

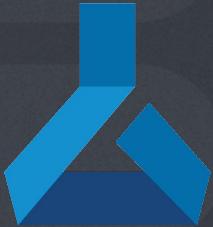
- Ambientación de instancias de cómputo
- ¿Cómo entrenar un modelo de NLP con redes neuronales en un Jupyter Notebook
- Inferencia del modelo (Creación de ChatBot)
- Creación de Job de entrenamiento
- Seguimiento de Jobs con MLFlow

Irving Uribe - ML Engineer





>\_



Ejecución de un script de entrenamiento como un trabajo  
de comando en Azure Machine Learning

<https://learn.microsoft.com/es-es/training/modules/run-training-script-command-job-azure-machine-learning/>





## Pasos para crear un job de entrenamiento con Azure ML

0.- Ambientar la instancia de procesamiento\*



1.- Desarrollar la lógica en un Jupyter Notebook

2(Opcional).- Pruebas de inferencia del modelo

3.- Creación de Script del Job

4.- Creación de la lógica para ejecutar el Job

5(Opcional).- Creación y configuración de imagen para procesamiento

6.- Ejecución del Job

7.- Variación de parámetros del Job de ejecución



8.- Explorar resultados



## ¿Qué es un ChatBot?

- > El principal componente de un ChatBot, , es un **Clasificador** que tiene la tarea de determinar qué quiere hacer el usuario.

Hola, soy tu asistente. ¿En qué te puedo ayudar?

Necesito que pongas una alarma para las 8 PM

Listo. Ya está tu alarma. ¿Necesitas algo más?

Si, por favor llámale a José

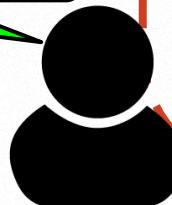
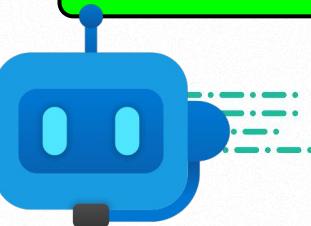
Ejemplos de acciones que puede realizar este sistema

Crear recordatorios

Programar alarmas

Llamar a persona

Contar un chiste

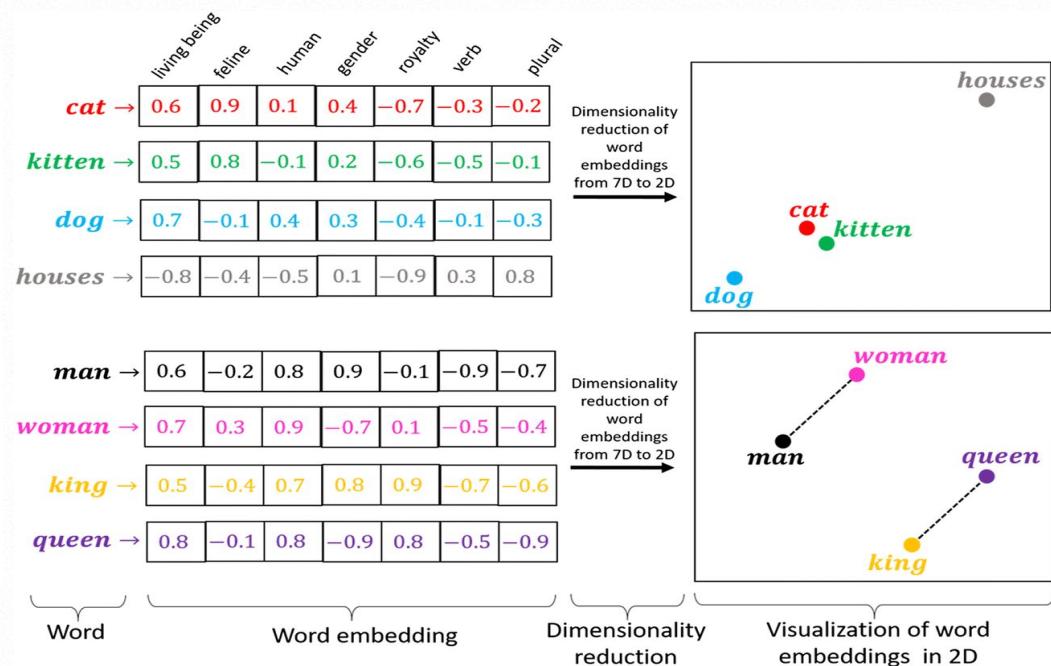
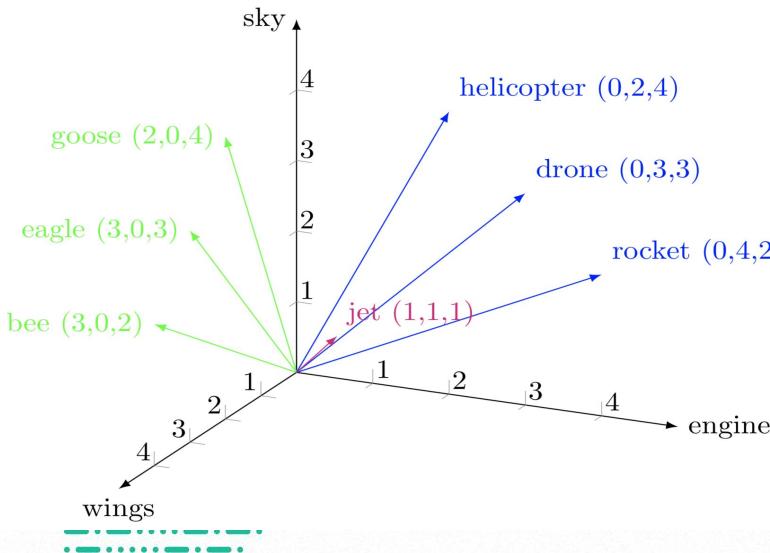




# ¿Qué son los Word embeddings?



Las palabras incrustadas (Embeddings) corresponden a palabras (**características**) con las que se evalúa a otra palabra. Cada embedding representa una dimensión diferente.

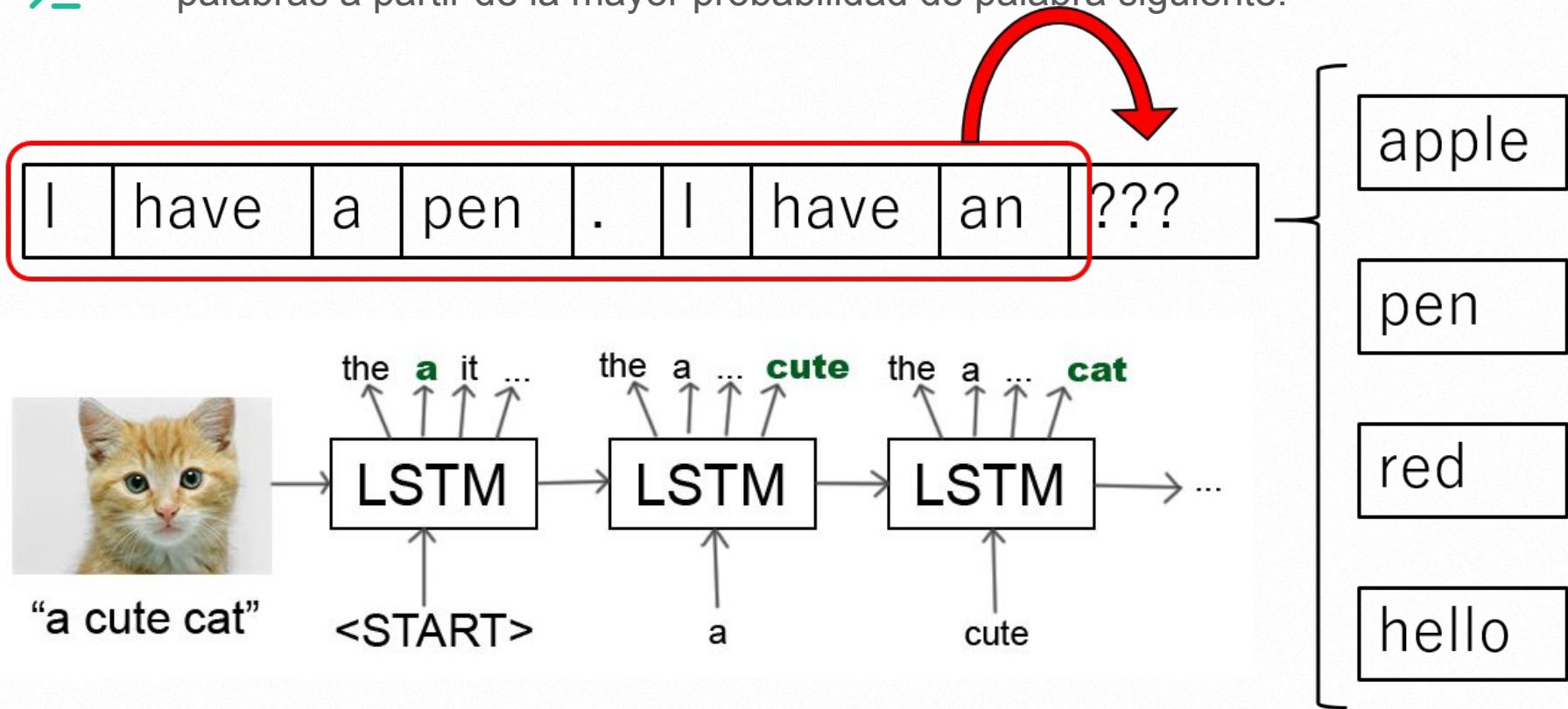




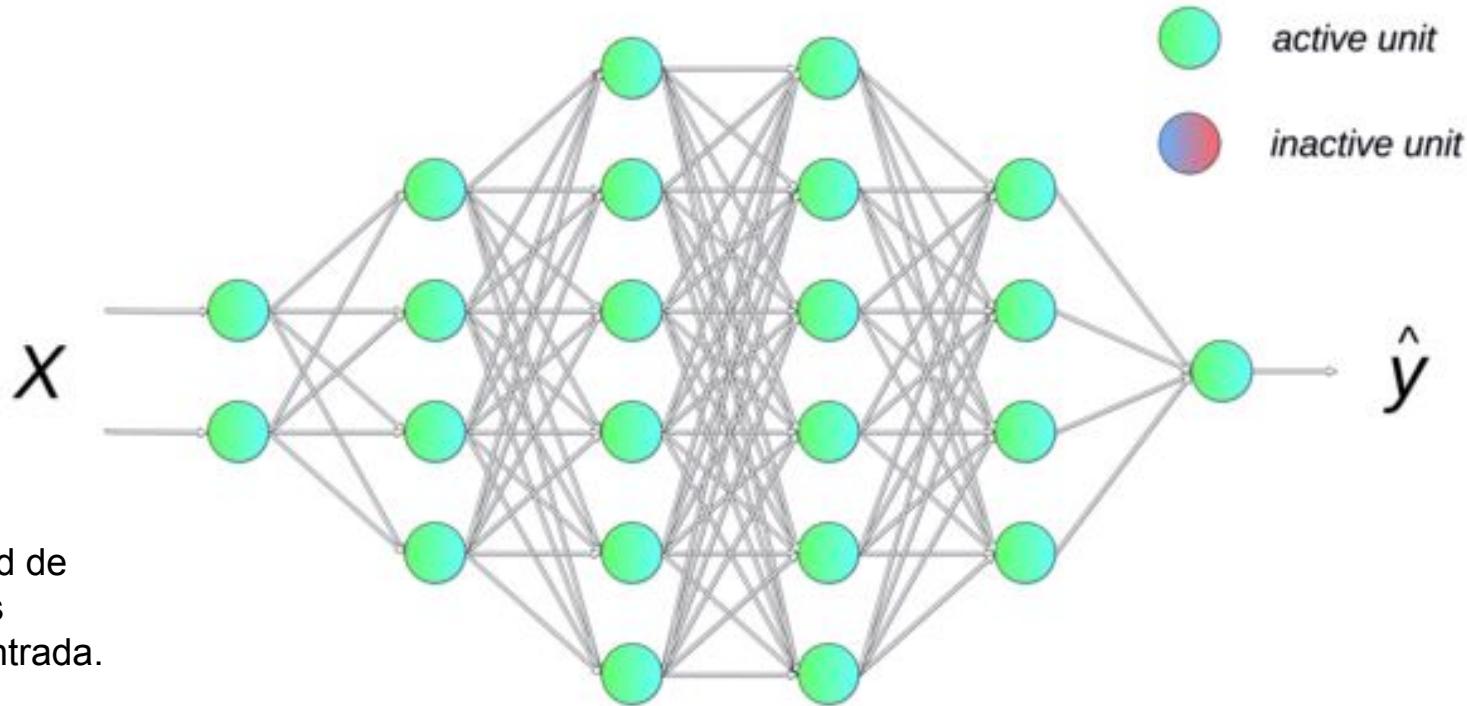
## ¿Por qué usar una red LSTM en NLP?



Son Redes recurrentes capaces de generar la secuencia más lógica de palabras a partir de la mayor probabilidad de palabra siguiente:



# ¿Qué es Dropout?



$$p^{[0]} = 0.0 \quad p^{[1]} = 0.0 \quad p^{[2]} = 0.5 \quad p^{[3]} = 0.0 \quad p^{[4]} = 0.25$$



## Ejercicio práctico:



Modifica los argumentos del comando para ejecutar el job de entrenamiento, para que te acepte diferentes hp de la red neuronal como por ejemplo, las épocas el batch size

```
1  from azure.ai.ml import command
2  from azure.ai.ml.entities import Environment
3
4  # configure job
5  job = command(
6      code="./",
7      command="python nlp_chatbot.py --training_data ./src/chat1.json",
8      environment=env,
9      compute="InstanciaAutoMLTest",
10     display_name="chatbot-script",
11     experiment_name="chatbot-job-training-chat1"
12 )
13
14 # submit job
15 returned_job = ml_client.create_or_update(job)
16 aml_url = returned_job.studio_url
17 print("Monitorea la url en:", aml_url)
...
# Cargar los datos de entrenamiento desde la ruta del archivo JSON
with open(args.training_data, 'r') as f:
    data = json.load(f)
df = pd.DataFrame(data)
```