

DESAFÍO

PIPELINE

RESULTS

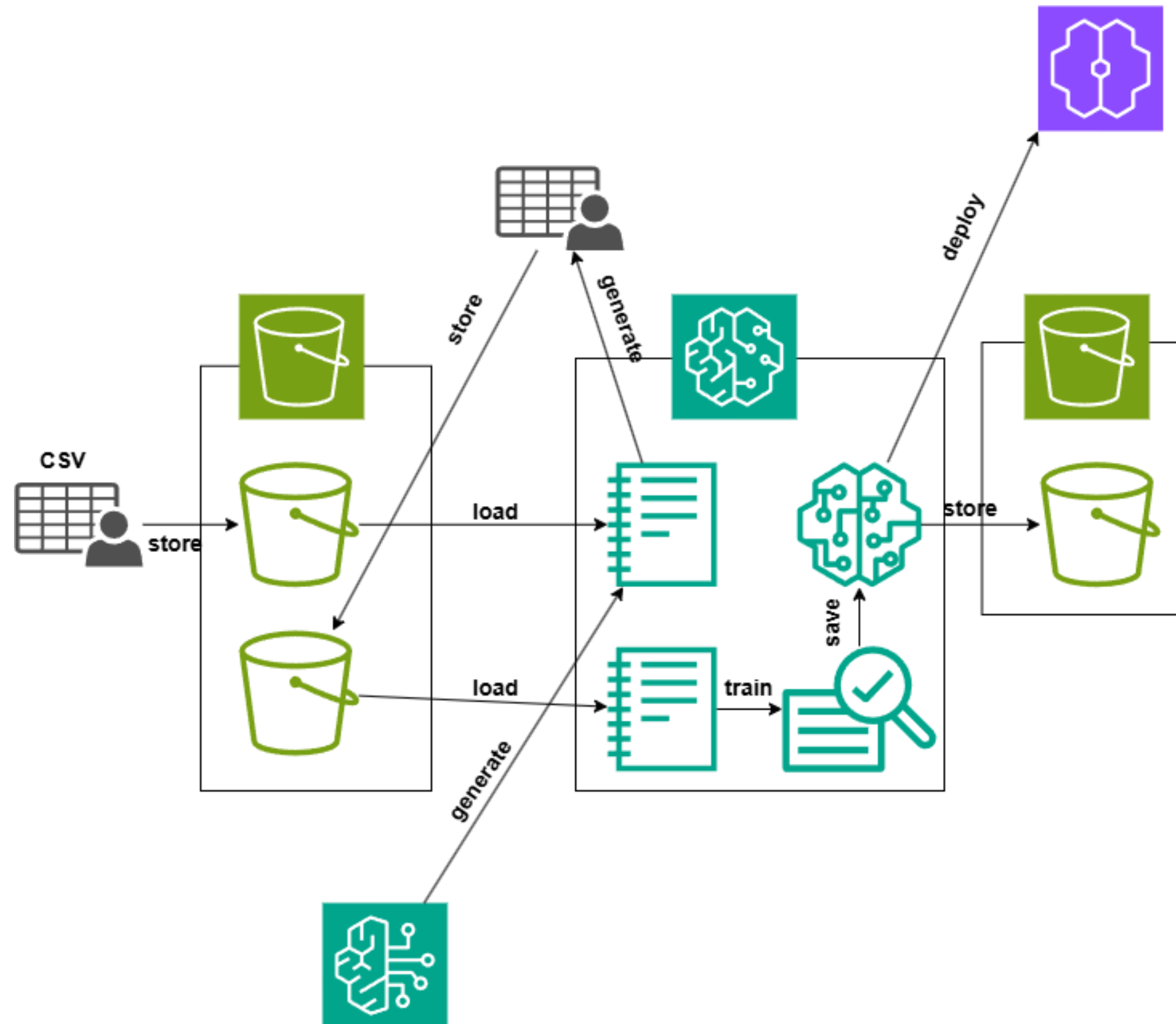
OTHERS



PREAMUTH CHALLENGE AI

Detección de Fraudes con Bedrock y SageMaker

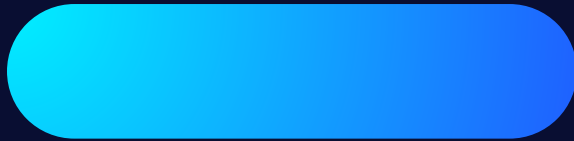
El presente desafío consistió en desarrollar una **solución integral de Machine Learning en AWS** para la **detección de riesgo crediticio**. Partiendo de un dataset tabular sin etiquetas, el objetivo era utilizar el poder de los modelos de lenguaje de Amazon Bedrock para **generar descripciones textuales y asignar automáticamente etiquetas** de "buen" o "mal" riesgo. Posteriormente, se debía **entrenar, optimizar y desplegar un modelo de clasificación** en Amazon SageMaker, creando un sistema de extremo a extremo **capaz de realizar predicciones en tiempo real** sobre nuevas transacciones.



El pipeline en AWS ilustra el **flujo de trabajo** completo desde la ingestión de datos hasta la inferencia.

A través de distintos servicios de AWS, los datos son procesados, transformados, clasificados, y utilizados para **entrenar y desplegar un modelo de machine learning**.

Todo el proceso está orquestado para asegurar eficiencia, escalabilidad y trazabilidad, permitiendo al usuario final interactuar con el modelo desplegado.



DESAFÍO

PIPELINE

RESULTS

OTHERS



Performance

Artifacts

Security

Hyperparameters

Configurations

Instances

Logs

Tags

Performance

Name	Value
ObjectiveMetric	0.8799999952316284
validation:accuracy	0.8799999952316284

Tras un **proceso de optimización** automatizada, se seleccionó un modelo con alto desempeño y capacidad de generalización.

El modelo final, elegido por su **precisión en validación (0.88)**, demuestra que es viable entrenar clasificadores efectivos utilizando **datos generados por modelos de lenguaje**.



The screenshot displays the AWS SageMaker Studio environment. The top navigation bar includes menus for File, Edit, View, Run, Kernel, Git, Tabs, Settings, and Help. The top right corner shows the default profile and a challenge ID: default-1753391593523 / challenge-preauth.

The left sidebar contains a file explorer with a tree view showing the following files and folders:

- user-default-efs (5m ago)
- deploy.ipynb (1h ago)
- generate_descripti... (8h ago)
- generate_targets.i... (5h ago)
- inference.py (2h ago)
- predict.py (2h ago)
- requirements.txt (4h ago)
- train_classifier.ipynb (2h ago)
- train.py (3h ago)
- untitled.py (5m ago)

The main terminal window shows the following commands and output:

```
sagemaker-user@default:~$ python predict.py "The 38-year-old male customer, a full-time government employee and homeowner, demonstrate s a strong credit risk profile. His consistent salary, substantial savings, long-standing checking account, and history of timely repa yments on previous loans all point toward a reliable financial standing, making him a low-risk borrower." --endpoint-name "credit-risk -classifier-endpoint-final"
Consultando al endpoint: credit-risk-classifier-endpoint-final

--- Resultado de la Clasificación ---
Predicción: good risk
Confianza: 0.8998
-----

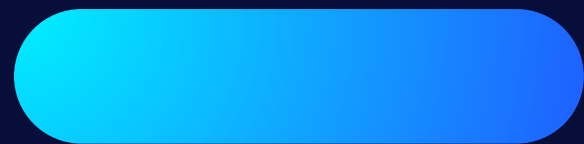
sagemaker-user@default:~$ python predict.py "The 29-year-old female customer, currently unemployed and living with relatives, reflects a high credit risk profile. Her lack of steady income, minimal savings, frequent overdrafts in her checking account, and recent appli cation for multiple short-term loans suggest significant financial vulnerability." --endpoint-name "credit-risk-classifier-endpoint-fi nal"
Consultando al endpoint: credit-risk-classifier-endpoint-final

--- Resultado de la Clasificación ---
Predicción: bad risk
Confianza: 0.1144
-----

sagemaker-user@default:~$ ls
deploy.ipynb          generate_targets.ipynb  predict.py            train.py              untitled.py
generate_descriptions.ipynb  inference.py            requirements.txt      train_classifier.ipynb  user-default-efs
sagemaker-user@default:~$
```




Se construyó un **pipeline de MLOps** totalmente funcional en AWS, integrando **generación y etiquetado automático de datos** mediante Amazon Bedrock, lo que permitió acelerar el desarrollo y reducir esfuerzo manual. El modelo final, **robusto y desplegado en un endpoint**, alcanzó una alta precisión en inferencia en tiempo real. A nivel técnico, se resolvieron desafíos clave como **la inestabilidad del entrenamiento** ajustando el espacio de búsqueda del HPO e implementando un scheduler de learning rate, además de optimizar el uso de recursos en la nube y **desarrollar un script de inferencia eficiente** y desacoplado para producción.



DESAFÍO

PIPELINE

RESULTS

OTHERS



GRACIAS

[linkedin.com/in/fabrizziomarinc/](https://www.linkedin.com/in/fabrizziomarinc/)

github.com/fabrizziomcl

sf.marinc@alum.up.edu.pe