

**DEPARTAMENTO:** Ciencias de la Ingeniería

**CARRERA:** Sistemas de Información

**ESTUDIANTE:** Lasso Valdivieso Stefanie Alejandra

Prado Imbacuán Giovanni Marcelo

Singaña Vergara Josue Alexander

**CURSO:** Noveno **PARALELO:** "A "

**DESCRIPCIÓN:** Concurso Innovación Tecnológica

**SEMESTRE:** 2025A

**TEMA:** Resolver el caso desarrollando un sistema de clasificación automática implementando los requerimientos de innovación tecnológica.

## Reporte

1. Tabla comparativa: "Hadoop vs. Dask vs. Pandas" (ventajas/desventajas en su solución).

Criterio	Hadoop	Dask	Pandas
RAM usada	Alta ( $\geq 8$ GB)	Media ( $\leq 4$ GB)	Baja ( $\leq 2$ GB)
Velocidad	Rápido en clúster	Bueno en CPU	Lento si $> 5k$ filas
Escalabilidad	Alta	Media/Alta	Baja
Instalación	Compleja	Sencilla (pip)	Muy sencilla
Ideal para	Big Data real	CPU multiproceso	Datasets pequeños

☒ Dask fue seleccionado como framework de procesamiento paralelo por su ligereza, escalabilidad y capacidad de ejecutar operaciones distribuidas o multinúcleo sin depender de ecosistemas pesados como Hadoop, facilitando así el procesamiento eficiente de imágenes.

☒ En el entrenamiento se usó TensorFlow Dataset con batch size para optimizar el manejo de memoria RAM.

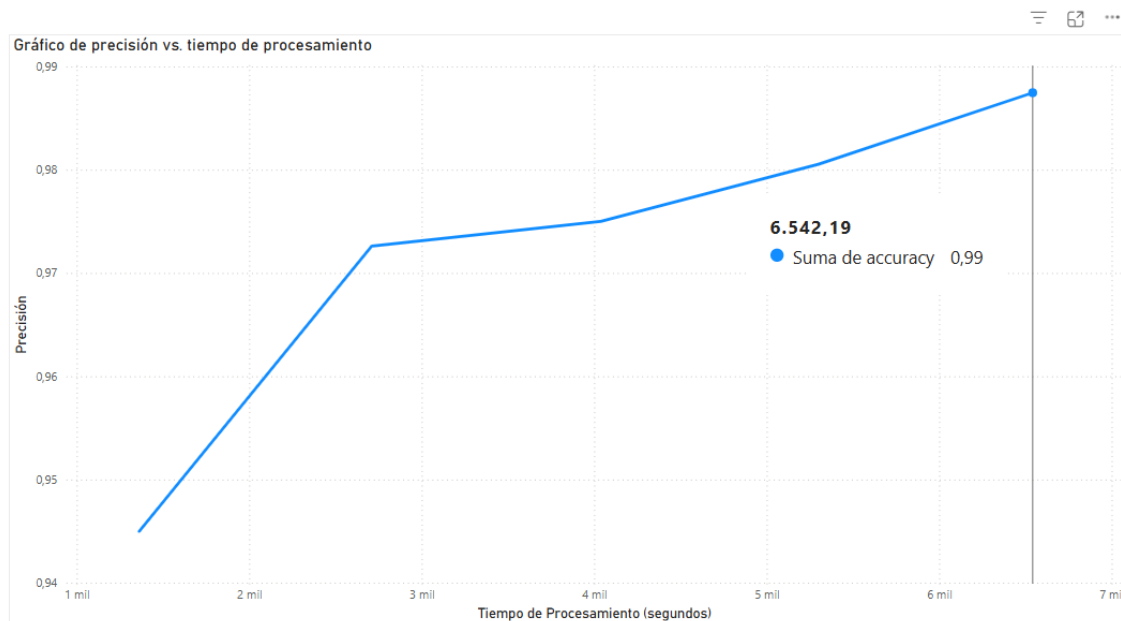
☒ Pandas se usó para obtener reporte Final en csv.



Hadoop no se utilizó, por optimizar recursos y debido al tiempo de desarrollo además mediante otras herramientas indicadas en el archivo del concurso permitió optimizar uso de memoria para procesamiento lo cual permitió no usar hadoop para nuestra solución.

## 2. Gráfico de precisión vs. tiempo de procesamiento.

Grafica generada en Power BI para Precisión vs tiempo de procesamiento.



Épocas ejecutadas durante el entrenamiento del modelo con sus respectivas métricas

```

Found 15153 files belonging to 3 classes.
Using 3030 files for validation.
Epoch 1/5
379/379 1362s 4s/step - accuracy: 0.9102 - loss: 0.2409 - val_accuracy: 0.7122 - val_loss: 3.7245
Epoch 2/5
379/379 1348s 4s/step - accuracy: 0.9686 - loss: 0.0955 - val_accuracy: 0.6713 - val_loss: 11.0043
Epoch 3/5
379/379 1329s 4s/step - accuracy: 0.9717 - loss: 0.0786 - val_accuracy: 0.6703 - val_loss: 11.9888
Epoch 4/5
379/379 1261s 3s/step - accuracy: 0.9775 - loss: 0.0671 - val_accuracy: 0.7901 - val_loss: 3.0100
Epoch 5/5
379/379 1242s 3s/step - accuracy: 0.9859 - loss: 0.0441 - val_accuracy: 0.7416 - val_loss: 4.9355

```

	accuracy	loss	val_accuracy	val_loss	epoch_time_sec	cumulative_time_sec
0	0.944981	0.157191	0.712211	3.724468	1362.204441	1362.204441
1	0.972614	0.080857	0.671287	11.004301	1348.084430	2710.288870
2	0.975006	0.073516	0.670297	11.988830	1328.519058	4038.807929
3	0.980533	0.059271	0.790099	3.010010	1261.147435	5299.955364
4	0.987462	0.039244	0.741584	4.935481	1242.236932	6542.192296