

## CONCLUSIONES DEL ANÁLISIS DEL DATASET

Integrantes:

- Espínola, Carla
- Gambarte, Antonella Nerea
- Torres, Dimas Ignacio

Nuestro grupo eligió el dataset de [PlantVillage](#) para resolver un problema de clasificación que consistirá en entrenar un modelo para identificar una enfermedad específica en una especie de planta.

### Carga del dataset y etiquetado

Al cargar el dataset notamos que se encuentra dividido en tres subcarpetas: color, grayscale y segmented. La primera contiene las imágenes en color, la segunda tiene las mismas imágenes pero en escala de grises y la última tiene las mismas imágenes pero sin el fondo.

Dentro de cada carpeta las imágenes se encuentran ya etiquetadas de la siguiente manera:

<Planta>\_\_<Enfermedad/healthy>

Donde:

- Planta: nombre de la planta.
- Enfermedad/healthy: nombre de la enfermedad. Si se trata de imágenes de plantas sanas, se etiqueta healthy.

### Cantidad de imágenes y distribución

Existen en total 162916 imágenes. Según el tipo, las cantidades son las siguientes:

Tipo	Cantidad
Color	54306
Grayscale	54305
Segmented	54305

No se detectaron archivos de imagen dañados o ilegibles.

## Tamaños de imágenes