
Pre TP1: Data Mining en Ciencia y Tecnología

José Saint Germain
joesg998@gmail.com

1 Introducción

El procesamiento de imágenes resulta desafiante por su alta dimensionalidad. La estructura de una **imagen digital** consiste en una **matriz de $N \times M$** , en donde la subunidad constituyente de la matriz es un **pixel** que codifica información para un color particular. Cada pixel representa la intensidad de luz en ese punto, que generalmente varía entre **[0,255]**, lo que es equivalente a **8 bits**.

Para representar imágenes a colores, se utiliza un modelo de percepción humana, en donde el color resulta a través de un sistema aditivo. El modelo se basa en la teoría de los componentes primarios del color que son Rojo, Verde y Azul (**RGB Red, Green and Blue**, por sus siglas en inglés). Por consiguiente, para representar digitalmente una imagen color, se necesitan **3 matrices de $N \times M$** . Una para el Rojo, otra para el Verde y otra para el Azul.

2 Objetivos

Familiarizarse con el procesamiento de imágenes. Para ello, se proponen diferentes manipulaciones que permitirán preparar el dataset para la detección y exploración de agrupamientos naturales.

3 Estructura de los datos

A partir del siguiente link, se obtendrán las imágenes a color de **210 flores** pertenecientes a **10 especies diferentes**. Cada imagen consiste en un archivo **.PNG** de 128 pixeles de ancho por 128 pixeles de profundidad (**128x128x3**). Adicionalmente, se encuentra el archivo **.CSV** con las etiquetas (labels) que corresponden a la especie de cada imagen.

4 Procesamiento de imágenes

Cargar el dataset de imágenes y sus respectivas etiquetas. Es importante asegurarse que las imágenes sean comparables en color, valor, rango y tamaño.

Explorar y graficar los subconjuntos de imágenes que representan flores de la misma especie.



Figure 1: Exploración de especies de flores

23 5 Manipulación de datos

24 5.1 Cambio de brillo

25 Cambiar la intensidad de una de las imágenes en escala de grises, transformarla en una imagen con
 26 mucho y otra con poco brillo.



Figure 2: Ajuste de brillo

27 5.2 Imagen en blanco y negro

28 Convertir una de las imágenes a blanco y negro (binario). ¿Es la única manera? Si existen otras
 29 transformaciones mostrar más de una conversión

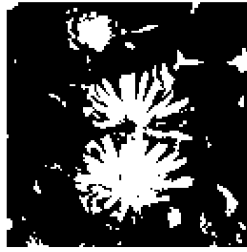


Figure 3: Imagen binaria

30 5.3 Imagen recortada

31 Recortar una parte significativa de la imagen, quedándose sólo con el círculo central de la misma.



Figure 4: Recorte de imagen

32 5.4 Imágenes mezcladas

33 Generar dos imágenes random: una imagen mezclando los pixels y otra mezclando partes de diferentes
34 imágenes.

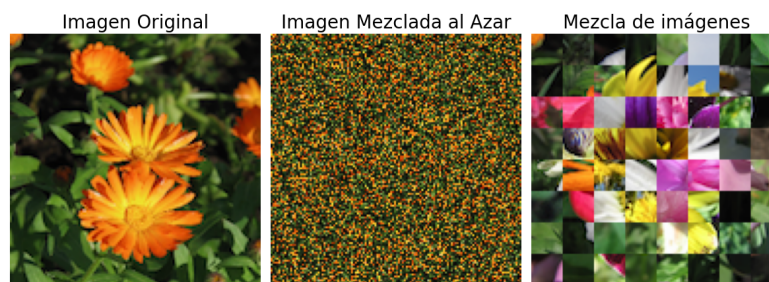


Figure 5: Imagenes con píxeles mezclados e intercambiados

35 5.5 Filtros de imagen

36 Aplicar dos tipos diferentes de filtros sobre una imagen, explique en qué casos conviene usar cada
37 uno.



Figure 6: Imagenes con filtros

38 5.6 Imágenes promedio

39 Calcular imagen promedio global y el promedio entre las distintas especies. ¿Se pueden distinguir
40 los promedios? ¿Cómo quedan los promedios si consideran las imágenes en blanco y negro?

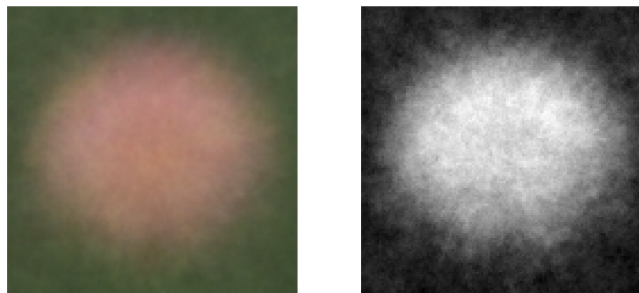


Figure 7: Imagenes promedio color y blanco y negro

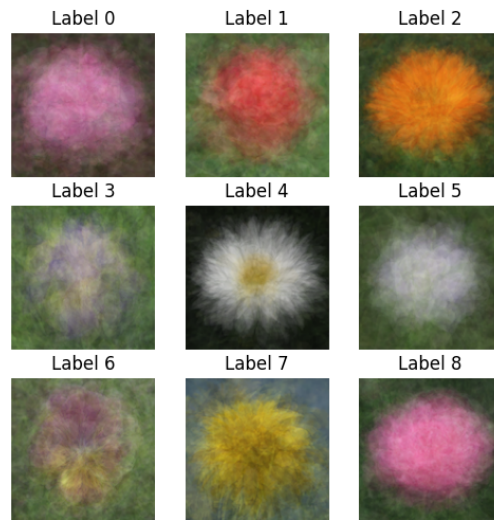


Figure 8: Imagenes promedio color y blanco y negro

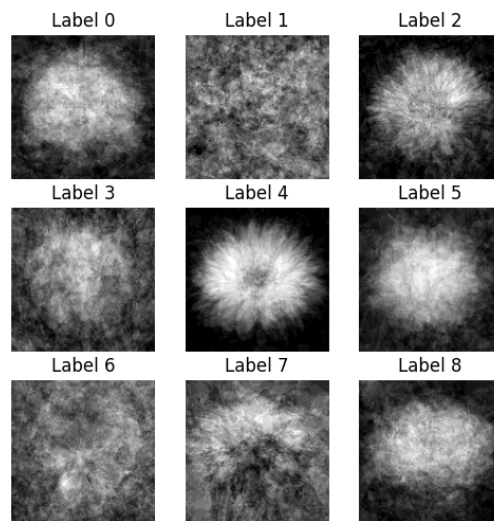


Figure 9: Imagenes promedio color y blanco y negro

41 6 Búsqueda de features

42 Analizar las distribuciones de valores de pixels por cada especie. ¿Se puede distinguir una especie en
43 algún rango de color?

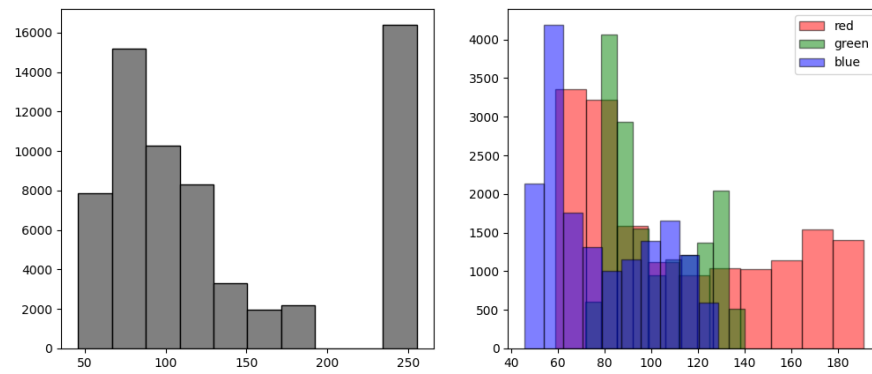


Figure 10: Distribución promedio de píxeles

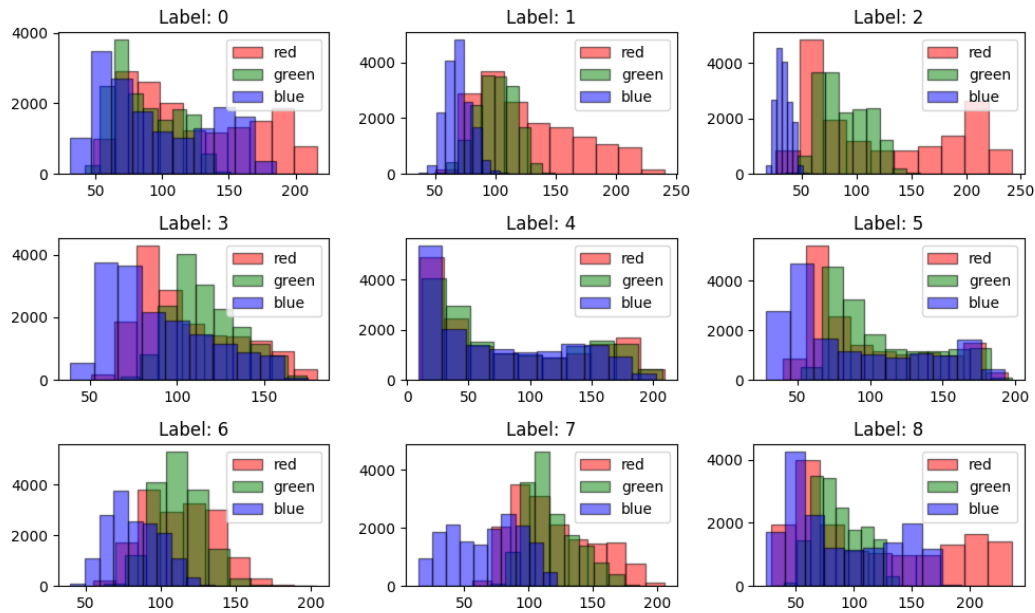


Figure 11: Distribución de píxeles de color por especie

6.1 Análisis de componentes principales

Realizar una inspección de las componentes principales del dataset y analizar si se pueden identificar las especies en esta representación.

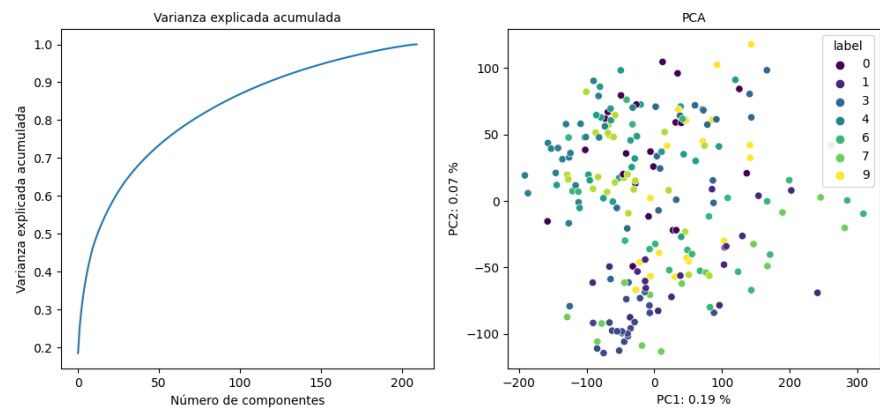


Figure 12: Análisis de componentes principales