

## Generación de paleta de colores a partir de imágenes con técnicas de machine learning no supervisado

Una paleta de colores es un conjunto de tonos que se utilizan juntos para crear un efecto armonioso, coherente y atractivo en las representaciones visuales (diseño gráfico, cine, fotografía, pintura, etc.), así como para transmitir mensajes y generar emociones determinados, de acuerdo con la intencionalidad del autor de la obra.

En un contexto de amplia demanda de herramientas que puedan asistir a diseñadores gráficos, directores de arte, pintores y creadores de contenido, entre otros, para la selección rápida y adecuada de los colores apropiados para sus obras, una aplicación para configurar paletas de colores a partir de imágenes sería de mucha utilidad.

Una vía para lograr una herramienta como esta es utilizar técnicas de machine learning no supervisado sobre imágenes para visualizar la distribución de los colores presentes en estas y generar, de manera automática, modelos de paletas de colores. En esta primera aproximación el método va a permitir identificar los colores en una imagen para construir un muestrario por similitud de píxeles, lo cual puede ser muy útil para estudios de marketing, psicología, medicina, arte, ambiente, entre otros. Por ejemplo, se podrían extraer los colores de las diferentes superficies de la Tierra en una imagen satelital para estudiar la distribución de la vegetación o de la contaminación.

### A. Objetivo.

- Desarrollar un método, basado en técnicas de agrupación, que permita extraer los tonos de una imagen y generar un muestrario de los colores presentes en esta.

### B. Conjunto de datos

Los datos están asociados con imágenes de obras de arte. Pueden ser descargados a partir de este [enlace](#).

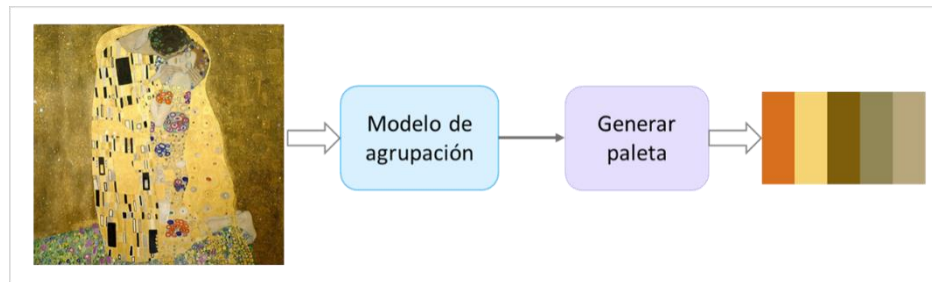
### C. Actividades para realizar.

1. Recopilación de las imágenes a partir del repositorio. La idea es seleccionar un conjunto diverso de muestras en diferentes estilos artísticos (no más de 10).
2. Preparación de las imágenes para el entrenamiento y prueba del modelo. Para este paso construir un pipeline que integre las transformaciones que se consideren adecuadas.
3. Desarrollo del modelo de agrupación para identificar los colores presentes en las imágenes.

4. Creación de una función que transforme los grupos de colores identificados en un muestrario representativo. Adicionalmente, se debe mostrar la distribución de los colores de la imagen en un espacio de dos dimensiones utilizando t-SNE.

#### D. Consideraciones.

Un diagrama general del método que se quiere desarrollar se muestra en la siguiente figura:



El algoritmo de agrupación a utilizar queda a consideración de cada grupo, pero es importante justificar la elección.

El número de colores del muestrario depende de la imagen. Sin embargo, la recomendación es que el muestrario tenga un máximo siete (7) colores.

Se debe evidenciar el desempeño del método construido mostrando la paleta para al menos cuatro (4) imágenes de diferentes estilos, con la visualización de la distribución de colores en el espacio de dos dimensiones.

#### D. Entregable.

Notebook (\*.ipynb y \*.html) del método desarrollado. El Notebook debe estar documentado con las justificaciones de las decisiones tomadas en cada paso. Además, deben ser visibles las ejecuciones de cada celda.

#### F. Criterios de evaluación.

Actividad	Puntos
Recopilación de un conjunto diverso de imágenes en diferentes estilos artísticos. Estas imágenes deben ser mostradas al inicio del notebook como parte del proceso de recolección de datos (no más de 10 imágenes).	10

Preparación de los datos, justificando las decisiones tomadas.	10
Construcción del pipeline de preparación de datos.	20
Construcción del modelo de agrupación con el algoritmo seleccionado con búsqueda de hiperparámetros, validándolo con medidas de evaluación adecuadas. Se justifica la selección del algoritmo y las métricas empleadas.	35
Construcción de la función que transforma los grupos de colores identificados en un muestrario representativo	15
Evidencia del desempeño del método construido mostrando la paleta generada para al menos cuatro (4) imágenes de diferentes estilos, con la visualización de la distribución de colores en el espacio de dos dimensiones utilizando t_SNE.	30