

Universidad Peruana de Ciencias Aplicadas



INFORME DEL TRABAJO FINAL

CURSO INTELIGENCIA ARTIFICIAL

Carrera de Ciencias de la Computación

Sección

CC61

Profesor

Reyes Silva, Patricia Daniela

Alumnos:	
Código	Apellidos y Nombres
U20181B618	Garcia Chavez, Nikole Scarlet
U202014068	Garcia Napuri, Andrea Fabiana
U20191b984	Anco Galvez, Jorge Martin
U20201A917	Roque Ponce, Christian Alonso

Lima, junio de 2024

Contenido

1. DESCRIPCIÓN DEL PROBLEMA.....	3
2. DESCRIPCIÓN Y VISUALIZACIÓN DEL CONJUNTO DE DATOS.....	4
2.1. Origen de datos	4
2.2. Análisis exploratorio de datos (EDA).....	4
Carga de Datos	4
Limpieza	5
Visualización de datos	6
3. PROPUESTA.....	12
4. DISEÑO DEL APLICATIVO	14
5. VALIDACIÓN DE RESULTADOS Y PRUEBAS	17
6. CONCLUSIONES	19
7. REFERENCIAS BIBLIOGRÁFICAS.....	20

1. DESCRIPCIÓN DEL PROBLEMA

La problemática de las personas desaparecidas en el Perú es una realidad angustiante que afecta a miles de familias en todo el país. Según el Ministerio del Interior (2023), más de 20,000 personas desaparecen anualmente en el país, siendo las mujeres en distintas etapas de su vida las más afectadas. Lima, la capital, registró 1,220 casos solo en los dos primeros meses del 2024 (Guardamino, 2024). La baja tasa de personas encontradas, con solo una de cada cuatro personas reportadas como desaparecidas siendo encontradas con vida (Alva, 2024), refleja la gravedad del problema y la necesidad de soluciones eficaces.

Las causas de este problema son multidimensionales y complejas. La escasez de recursos y tecnología, la falta de coordinación entre agencias gubernamentales, y la insuficiente conciencia pública sobre la magnitud del problema, complican los esfuerzos de búsqueda y recuperación. Además, la ausencia de una base de datos centralizada y confiable dificulta el seguimiento de personas desaparecidas, exacerbando el panorama.

Ante esta preocupante realidad, es crucial buscar soluciones innovadoras que aprovechen el potencial de la tecnología y los datos disponibles. Nuestro objetivo es desarrollar un sistema que permita a los usuarios ingresar características específicas o imágenes de la persona desaparecida, y que el sistema recomiende los individuos más similares de una base de datos de personas encontradas. Este enfoque es especialmente útil para grupos vulnerables como niños, personas de la tercera edad o discapacitados, quienes son más susceptibles a perderse y requieren un enfoque especializado.

El sistema propuesto utilizará técnicas avanzadas de inteligencia artificial, como redes neuronales convolucionales para el reconocimiento facial, y algoritmos de clasificación para analizar características físicas. Esto no solo optimizará los esfuerzos de búsqueda, sino que también mejorará la precisión y rapidez en la identificación de personas desaparecidas, proporcionando una herramienta vital para autoridades y familias en la lucha contra este problema.

2. DESCRIPCIÓN Y VISUALIZACIÓN DEL CONJUNTO DE DATOS

2.1. Origen de datos

El conjunto de datos utilizado en este proyecto proviene del CelebA Dataset, disponible en Kaggle. Este dataset fue originalmente recopilado por investigadores del MMLAB, The Chinese University of Hong Kong, y está destinado a fines de investigación no comerciales. El CelebA Dataset consta de las siguientes características:

- 202,599 imágenes faciales de diversas celebridades.
- 10,177 identidades únicas, aunque los nombres de las identidades no se proporcionan.
- 40 anotaciones binarias de atributos por imagen, que incluyen características faciales como color de cabello, expresiones faciales, uso de accesorios, etc.
- 5 ubicaciones de puntos de referencia en la imagen.

El dataset se compone de varios archivos, incluyendo imágenes faciales alineadas, información de cuadros delimitadores (bounding boxes), coordenadas de puntos de referencia y etiquetas de atributos para cada imagen.

Este dataset es ampliamente utilizado en tareas de reconocimiento facial, detección de rostros y clasificación de atributos faciales debido a su diversidad en términos de pose, fondo, y personas representadas.

2.2. Análisis exploratorio de datos (EDA)

El análisis exploratorio de datos (EDA) realizado para el conjunto de datos CelebA se llevó a cabo utilizando las bibliotecas pandas, matplotlib y seaborn en Python.

Carga de Datos

En primer lugar, se carga el archivo CSV, el cual contiene etiquetas de atributos para cada imagen. Luego se revisa las primeras filas del DataFrame para entender la estructura de los datos y verificar si hay columnas relevantes o si hay datos que necesitan limpieza adicional.

```
In [2]: # Cargar el archivo CSV
df_original = pd.read_csv('list_attr_celeba.csv')
```

```
In [3]: # Visualizar Las primeras filas del DataFrame
print(df_original.head())
```

	image_id	5_o_Clock_Shadow	Arched_Eyebrows	Attractive	Bags_Under_Eyes	\
0	000001.jpg	-1	1	1	-1	
1	000002.jpg	-1	-1	-1	1	
2	000003.jpg	-1	-1	-1	-1	
3	000004.jpg	-1	-1	1	-1	
4	000005.jpg	-1	1	1	-1	

	Bald	Bangs	Big_Lips	Big_Nose	Black_Hair	...	Sideburns	Smiling	\
0	-1	-1	-1	-1	-1	...	-1	1	
1	-1	-1	-1	1	-1	...	-1	1	
2	-1	-1	1	-1	-1	...	-1	-1	
3	-1	-1	-1	-1	-1	...	-1	-1	
4	-1	-1	1	-1	-1	...	-1	-1	

	Straight_Hair	Wavy_Hair	Wearing_Earrings	Wearing_Hat	Wearing_Lipstick	\
0	1	-1	1	-1	1	
1	-1	-1	-1	-1	-1	
2	-1	1	-1	-1	-1	
3	1	-1	1	-1	1	
4	-1	-1	-1	-1	1	

	Wearing_Necklace	Wearing_Necktie	Young
0	-1	-1	1
1	-1	-1	1
2	-1	-1	1
3	1	-1	1
4	-1	-1	1

[5 rows x 41 columns]

Limpieza

Posteriormente, se realizó una inspección más detallada de los datos utilizando el método `info()` de pandas para identificar posibles valores nulos o faltantes para garantizar la integridad de los datos.

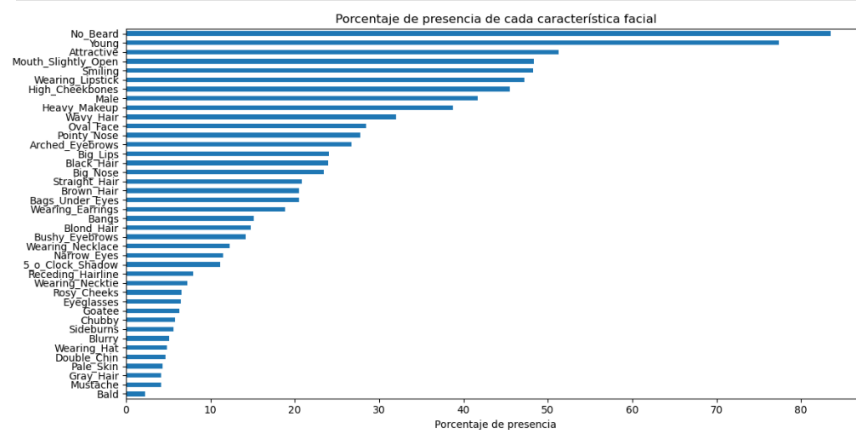
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 202599 entries, 0 to 202598
Data columns (total 41 columns):
#   Column                Non-Null Count  Dtype
---  -
0   image_id              202599 non-null object
1   5_o_Clock_Shadow      202599 non-null int64
2   Arched_Eyebrows      202599 non-null int64
3   Attractive            202599 non-null int64
4   Bags_Under_Eyes      202599 non-null int64
5   Bald                 202599 non-null int64
6   Bangs                202599 non-null int64
7   Big_Lips              202599 non-null int64
8   Big_Nose              202599 non-null int64
9   Black_Hair            202599 non-null int64
10  Blond_Hair            202599 non-null int64
11  Blurry                202599 non-null int64
12  Brown_Hair            202599 non-null int64
13  Bushy_Eyebrows        202599 non-null int64
14  Chubby                202599 non-null int64
15  Double_Chin           202599 non-null int64
16  Eyeglasses            202599 non-null int64
17  Goatee                202599 non-null int64
18  Gray_Hair             202599 non-null int64
19  Heavy_Makeup           202599 non-null int64
20  High_Cheekbones        202599 non-null int64
21  Male                  202599 non-null int64
22  Mouth_Slightly_Open    202599 non-null int64
23  Mustache              202599 non-null int64
24  Narrow_Eyes           202599 non-null int64
25  No_Beard              202599 non-null int64
26  Oval_Face             202599 non-null int64
27  Pale_Skin              202599 non-null int64
28  Pointy_Nose           202599 non-null int64
29  Receding_Hairline      202599 non-null int64
30  Rosy_Cheeks           202599 non-null int64
31  Sideburns             202599 non-null int64
32  Smiling               202599 non-null int64
33  Straight_Hair          202599 non-null int64
34  Wavy_Hair             202599 non-null int64
35  Wearing_Earrings      202599 non-null int64
36  Wearing_Hat           202599 non-null int64
37  Wearing_Lipstick       202599 non-null int64
38  Wearing_Necklace       202599 non-null int64
39  Wearing_Necktie        202599 non-null int64
40  Young                 202599 non-null int64
```

Visualización de datos

Luego, se calcularon y visualizaron diferentes métricas para comprender la distribución y la presencia de características faciales en las imágenes. Se crearon gráficos de barras e histogramas para mostrar el porcentaje de presencia de cada característica facial, así como la cantidad de características por imagen.

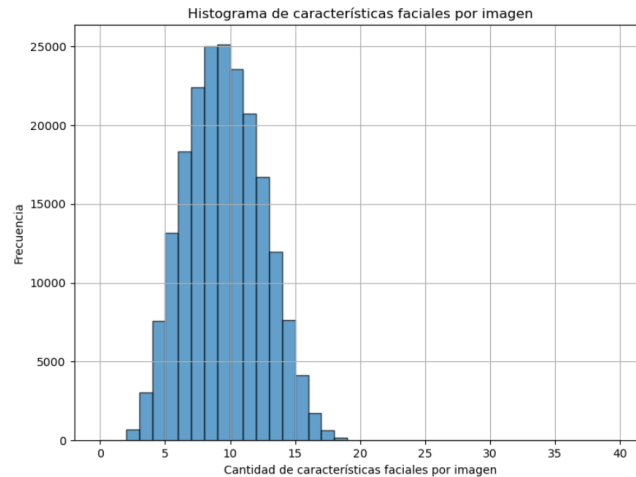
```
In [5]: # Calcular el porcentaje de presencia de cada característica
presence_percentages = df_original.drop(columns=['image_id']).apply(lambda x: (x == 1).mean() * 100)

# Visualizar el porcentaje de presencia de cada característica
plt.figure(figsize=(12, 6))
presence_percentages.sort_values().plot(kind='barh')
plt.xlabel('Porcentaje de presencia')
plt.title('Porcentaje de presencia de cada característica facial')
plt.tight_layout()
plt.show()
```

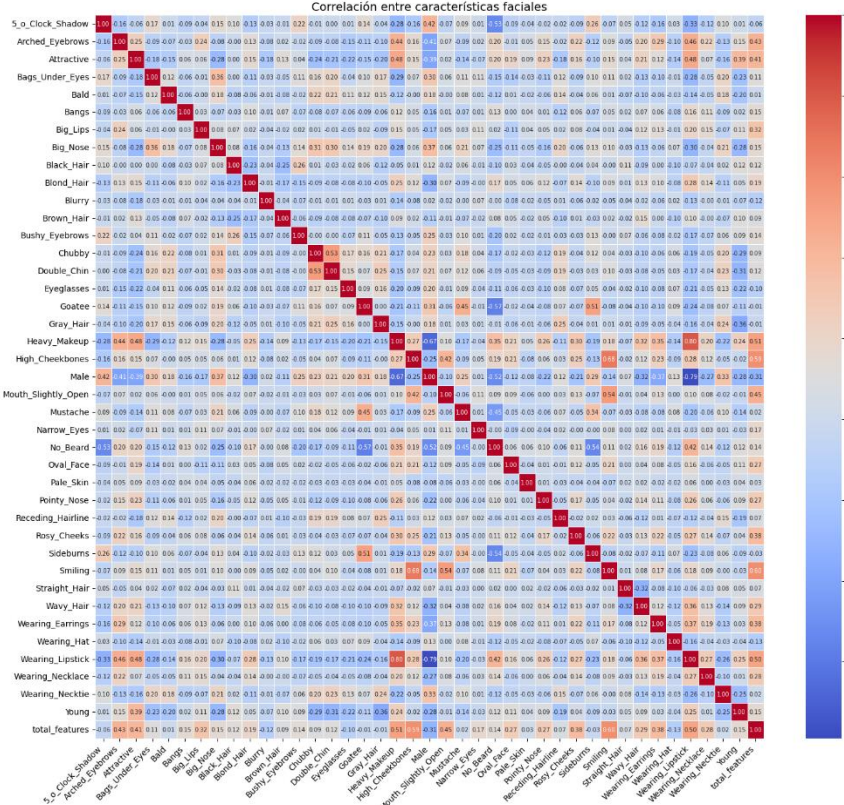


El gráfico muestra que las características más prevalentes en el dataset son la ausencia de barba y la juventud, presentes en aproximadamente el 80% y 75% de las imágenes, respectivamente. Esto indica un sesgo claro hacia individuos jóvenes y sin barba. Esto puede influir en el rendimiento del modelo, haciéndolo menos efectivo para identificar individuos de grupos demográficos menos representados, como personas mayores o con barba.

```
In [8]: # Histograma de características por imagen
df_original['total_features'] = df_original.drop(columns=['image_id']).apply(lambda x: (x == 1).sum(), axis=1)
plt.figure(figsize=(8, 6))
plt.hist(df_original['total_features'], bins=range(0, 41), edgecolor='black', alpha=0.7)
plt.xlabel('Cantidad de características faciales por imagen')
plt.ylabel('frecuencia')
plt.title('Histograma de características faciales por imagen')
plt.grid(True)
plt.tight_layout()
plt.show()
```



El histograma muestra la cantidad de características faciales presentes por imagen en el dataset. La mayoría de las imágenes tienen entre 7 y 13 características faciales, con un pico alrededor de 10 características. Esto indica que la mayoría de las imágenes en el dataset tienen un número moderado de características faciales, lo que sugiere una buena variedad y complejidad en las representaciones faciales. Esta distribución equilibrada puede ayudar a entrenar un modelo más robusto al proporcionar suficientes ejemplos de imágenes con múltiples características faciales.



El mapa de calor muestra las correlaciones entre diferentes características faciales en el dataset. Las correlaciones más altas (cercanas a 1 o -1) indican una fuerte relación entre características, mientras que las más cercanas a 0 indican poca o ninguna relación.

- Correlaciones positivas destacadas: Por ejemplo, `Heavy_Makeup` y `Wearing_Lipstick` (0.63), lo que sugiere que las personas con mucho maquillaje también suelen usar lápiz labial.
- Correlaciones negativas destacadas: Por ejemplo, `No_Beard` y `Mustache` (-0.54), indicando que es raro encontrar personas sin barba pero con bigote.

En general, el gráfico revela algunas relaciones esperadas entre características faciales, lo que puede ayudar en la selección de atributos relevantes para el modelado y análisis posterior. La mayoría de las correlaciones son bajas, lo que sugiere que las características faciales en el dataset son bastante independientes entre sí.

Por otro lado, para facilitar la preparación del conjunto de datos y asegurar la calidad de las características utilizadas en el entrenamiento del modelo, se realizó un exhaustivo proceso de selección. Por ello, de las 40 columnas originales, se retuvieron 30 más relevantes para identificar

a personas perdidas y fueron guardadas en un archivo llamado 'list_attr_celeba_limpio.csv', con los datos listos para entrenar al modelo. A continuación, se detallan estas características seleccionadas junto con una breve descripción:

- **5_o_Clock_Shadow:** Indica la presencia de una sombra de barba.
- **Arched_Eyebrows:** Cejas arqueadas.
- **Bags_Under_Eyes:** Bolsas debajo de los ojos.
- **Bald:** Indica si la persona es calva.
- **Bangs:** Presencia de flequillo.
- **Big_Lips:** Labios grandes.
- **Big_Nose:** Nariz grande.
- **Black_Hair:** Cabello negro.
- **Blond_Hair:** Cabello rubio.
- **Brown_Hair:** Cabello castaño.
- **Bushy_Eyebrows:** Cejas pobladas.
- **Chubby:** Indica si la persona es regordeta.
- **Double_Chin:** Presencia de doble mentón.
- **Eyeglasses:** Indica si la persona usa gafas.
- **Goatee:** Presencia de perilla.
- **Gray_Hair:** Cabello gris.
- **High_Cheekbones:** Pómulos altos.
- **Male:** Indica si la persona es masculina.
- **Mustache:** Presencia de bigote.
- **Narrow_Eyes:** Ojos estrechos.
- **No_Beard:** Indica si la persona no tiene barba.
- **Oval_Face:** Rostro ovalado.
- **Pale_Skin:** Piel pálida.
- **Pointy_Nose:** Nariz puntiaguda.
- **Receding_Hairline:** Presencia de entradas en retroceso.
- **Rosy_Cheeks:** Mejillas sonrosadas.
- **Sideburns:** Presencia de patillas.

- **Straight_Hair:** Cabello liso.
- **Wavy_Hair:** Cabello ondulado.
- **Young:** Indica si la persona es joven.

Estas características se seleccionaron por su relevancia para la identificación facial y su potencial para entrenar eficazmente el algoritmo seleccionado. La eliminación de las características consideradas irrelevantes nos permite enfocarnos en aquellos rasgos físicos distintivos que son más útiles para nuestro propósito específico.

Finalmente, para establecer una estructura adecuada de entrenamiento, validación y prueba del modelo, se dividieron las imágenes en tres conjuntos distintos. Según las recomendaciones proporcionadas en el archivo "list_eval_partition.csv", se asignaron las siguientes cantidades de imágenes a cada conjunto:

- Conjunto de entrenamiento: Imágenes 1-162,770.
- Conjunto de validación: Imágenes 162,771-182,637.
- Conjunto de prueba: Imágenes 182,638-202,599.

Continuando con el procesamiento de datos, se ordenó el dataset a manera que se creen las carpetas características con sus debidas imágenes. Esto se realizó mediante la lectura de datos del CSV, luego realizamos la función donde se corrobora que imágenes pertenecen a cada característica, tal como se ve en la imagen.

```
Característica: 5_o_Clock_Shadow  
000007.jpg  
000008.jpg
```

```
Característica: Arched_Eyebrows  
000001.jpg  
000005.jpg  
000006.jpg  
000008.jpg  
000009.jpg
```

```
Característica: Attractive  
000001.jpg  
000004.jpg  
000005.jpg  
000006.jpg  
000007.jpg  
000009.jpg
```

```
Característica: Bags_Under_Eyes  
000002.jpg  
000007.jpg  
000008.jpg
```

```
Característica: Bald
```

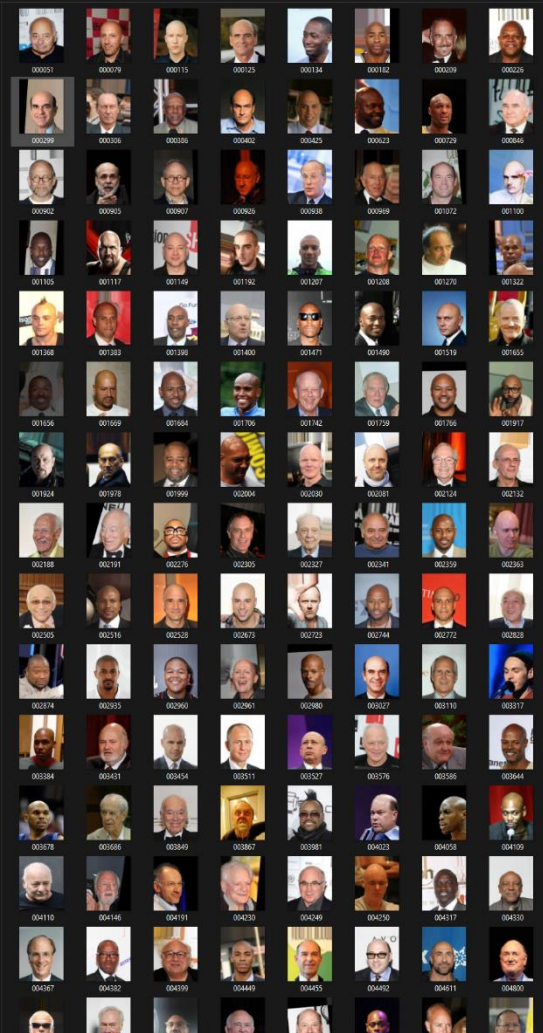
```
Característica: Bangs  
000009.jpg
```

```
Característica: Big_Lips  
000003.jpg  
000005.jpg  
000006.jpg  
000007.jpg  
000008.jpg  
000009.jpg
```

```
Característica: Big_Nose  
000002.jpg  
000007.jpg
```

Después de verificar esta información, se procedió a crear una carpeta denominada “Características”, la cual contiene una copia de cada imagen del dataset original. De esta manera, preparamos nuestro dataset para el entrenamiento.

Nombre	Fecha de modificación	Tipo	Tamaño
5_o_Clock_Shadow_images	27/4/2024 04:58	Carpeta de archivos	
Arched_Eyebrows_images	27/4/2024 04:58	Carpeta de archivos	
Attractive_images	27/4/2024 04:58	Carpeta de archivos	
Bags_Under_Eyes_images	27/4/2024 04:58	Carpeta de archivos	
Bald_images	27/4/2024 04:58	Carpeta de archivos	
Bangs_images	27/4/2024 04:58	Carpeta de archivos	
Big_Lips_images	27/4/2024 04:58	Carpeta de archivos	
Big_Nose_images	27/4/2024 04:58	Carpeta de archivos	
Black_Hair_images	27/4/2024 04:58	Carpeta de archivos	
Blond_Hair_images	27/4/2024 04:58	Carpeta de archivos	
Blurry_images	27/4/2024 04:58	Carpeta de archivos	
Brown_Hair_images	27/4/2024 04:58	Carpeta de archivos	
Bushy_Eyebrows_images	27/4/2024 04:58	Carpeta de archivos	
Chubby_images	27/4/2024 04:58	Carpeta de archivos	
Double_Chin_images	27/4/2024 04:58	Carpeta de archivos	
Eyeglasses_images	27/4/2024 04:58	Carpeta de archivos	
Goatee_images	27/4/2024 04:58	Carpeta de archivos	
Gray_Hair_images	27/4/2024 04:58	Carpeta de archivos	
Heavy_Makeup_images	27/4/2024 04:58	Carpeta de archivos	
High_Cheekbones_images	27/4/2024 04:58	Carpeta de archivos	
Male_images	27/4/2024 04:58	Carpeta de archivos	
Mouth_Slightly_Open_images	27/4/2024 04:58	Carpeta de archivos	
Mustache_images	27/4/2024 04:58	Carpeta de archivos	
Narrow_Eyes_images	27/4/2024 04:58	Carpeta de archivos	
No_Beard_images	27/4/2024 04:58	Carpeta de archivos	
Oval_Face_images	27/4/2024 04:58	Carpeta de archivos	
Pale_Skin_images	27/4/2024 04:58	Carpeta de archivos	
Pointy_Nose_images	27/4/2024 04:58	Carpeta de archivos	
Receding_Hairline_images	27/4/2024 04:58	Carpeta de archivos	
Rosy_Cheeks_images	27/4/2024 04:58	Carpeta de archivos	
Sideburns_images	27/4/2024 04:58	Carpeta de archivos	
Smiling_images	27/4/2024 04:58	Carpeta de archivos	
Straight_Hair_images	27/4/2024 04:58	Carpeta de archivos	
Wavy_Hair_images	27/4/2024 04:58	Carpeta de archivos	
Wearing_Earrings_images	27/4/2024 04:58	Carpeta de archivos	
Wearing_Hat_images	27/4/2024 04:58	Carpeta de archivos	
Wearing_Lipstick_images	27/4/2024 04:58	Carpeta de archivos	
Wearing_Necklace_images	27/4/2024 04:58	Carpeta de archivos	
Wearing_Necktie_images	27/4/2024 04:58	Carpeta de archivos	
Young_images	27/4/2024 04:58	Carpeta de archivos	



3. PROPUESTA

El objetivo de nuestra propuesta es desarrollar un sistema de inteligencia artificial que permita a los usuarios identificar y localizar personas desaparecidas mediante el uso de reconocimiento facial y análisis de características físicas. Este sistema busca mejorar la eficacia y precisión de las búsquedas, proporcionando una herramienta esencial para la recuperación de personas, especialmente aquellas vulnerables como niños, personas de la tercera edad y personas con discapacidades.

Para lograr este objetivo, hemos decidido utilizar técnicas avanzadas de inteligencia artificial. El sistema se basa en dos componentes principales:

1. **Reconocimiento facial con Redes Neuronales Convolucionales (CNNs):** Estas redes serán entrenadas para identificar y verificar rostros a partir de imágenes, aprovechando su capacidad para extraer características faciales de manera efectiva.
2. **Clasificación de características físicas mediante algoritmos de Machine Learning:** Utilizaremos técnicas como Random Forest y Support Vector Machines (SVM) para analizar características físicas adicionales como el color de cabello, uso de gafas, etc.

El sistema permitirá a los usuarios ingresar hasta 30 características físicas de la persona desaparecida, lo cual incluirá, pero no se limitará a:

- Color de cabello (negro, rubio, castaño, gris)
- Presencia de barba o bigote
- Uso de gafas
- Rasgos faciales específicos (nariz grande, labios grandes, pómulos altos)
- Presencia de cicatrices o tatuajes
- Edad aproximada
- Altura aproximada
- Complexión (delgada, media, robusta)
- Tipo de peinado (liso, ondulado, con flequillo)
- Presencia de accesorios (aretes, gorros)

Estas características han sido seleccionadas por su relevancia para la identificación facial y su potencial para entrenar eficazmente el algoritmo.

La metodología que guía nuestro proyecto es CRISP-DM (Cross-Industry Standard Process for Data Mining). Este marco proporciona una estructura robusta y sistemática para el desarrollo del proyecto, asegurando una implementación eficiente y efectiva.

4. DISEÑO DEL APLICATIVO

Para desarrollar un aplicativo que facilite la búsqueda de desaparecidos con un sistema de reconocimiento visual usando redes convolucionales, se aplica el marco de trabajo CRISP-DM (Cross-Industry Standard Process for Data Mining). Este marco consta de seis fases principales: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue.

Comprensión del negocio:

- En esta fase, se comprende la problemática de la búsqueda de desaparecidos y cómo el reconocimiento visual puede contribuir a resolverla.
- Se identifican los objetivos del aplicativo, como facilitar la identificación de individuos desaparecidos a través de imágenes.
- Se definen los criterios de éxito, como la precisión del sistema en la identificación de individuos.

Comprensión de los datos:

- Se recopilan datos relevantes para el entrenamiento del modelo de reconocimiento visual. Esto incluye imágenes de individuos desaparecidos y otras imágenes de referencia.
- Se realiza un análisis exploratorio de los datos para comprender su calidad y características.

Preparación de los datos:

- En esta fase, se preprocesan y se limpian los datos. Esto incluye la eliminación de imágenes de baja calidad o con información irrelevante.
- Se lleva a cabo la división del conjunto de datos en conjuntos de entrenamiento, validación y prueba.

Modelado:

- Se selecciona y se entrena un modelo de reconocimiento visual basado en redes convolucionales, como CNN (Convolutional Neural Network).
- Se ajustan los hiperparámetros del modelo para mejorar su rendimiento.

- Se realiza la validación cruzada para evaluar la capacidad del modelo para generalizar a datos no vistos.

Evaluación:

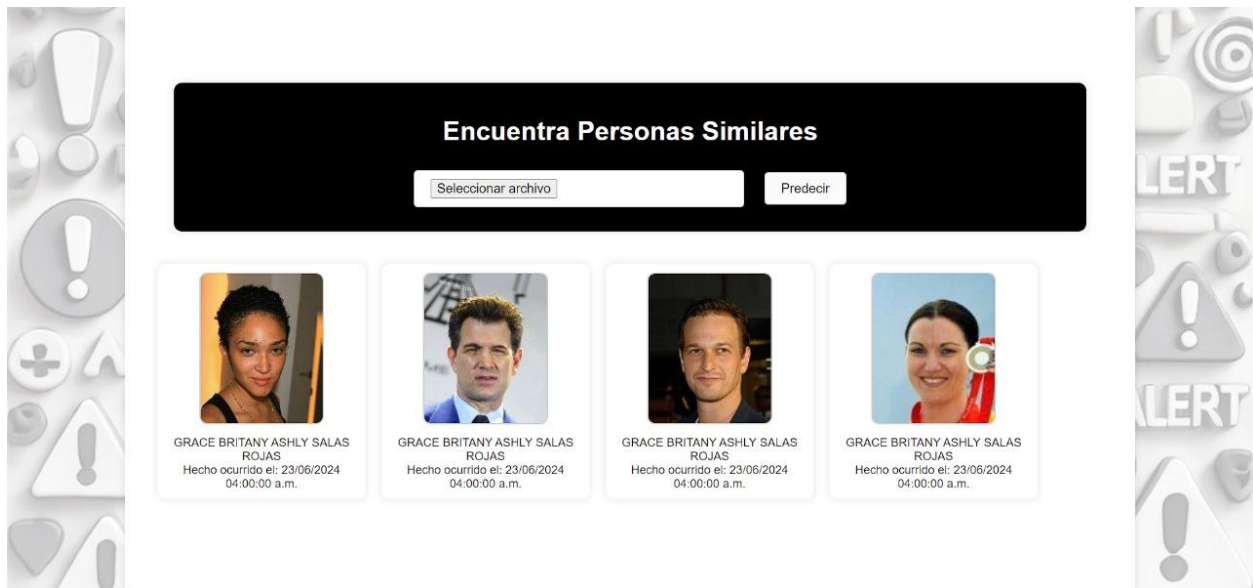
- Se evalúa el rendimiento del modelo utilizando métricas relevantes, como la precisión, la sensibilidad y la especificidad.
- Se identifican posibles áreas de mejora del modelo y se consideran estrategias para abordarlas.

Despliegue:

- Se implementa el aplicativo que utiliza el modelo entrenado en un entorno de producción.
- Se realizan pruebas finales para garantizar que el sistema funcione correctamente.
- Se proporciona capacitación y soporte para los usuarios finales del aplicativo.

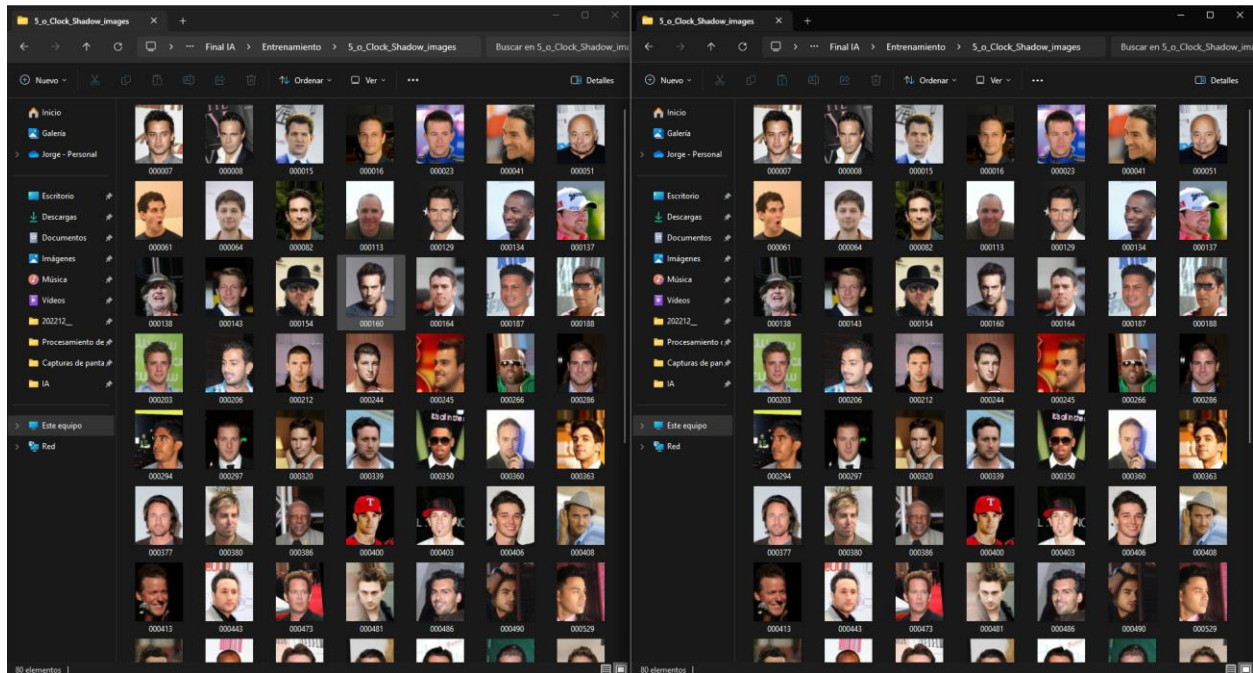
Este proceso sigue una metodología iterativa, donde cada fase se retroalimenta de las anteriores para garantizar el desarrollo efectivo del aplicativo de búsqueda de desaparecidos con reconocimiento visual.

El aplicativo se muestra de esta manera:



Entrenamiento:

Para el entrenamiento primero tuvimos que dividir la cada carpeta que obtuvimos de características, en una carpeta de imágenes para entrenamiento y otro para validación. Esto se debe para garantizar que los modelos de aprendizaje automáticos sean capaces de generalizar bien a los nuevos datos, y no solo memorizar los datos de entrenamiento.

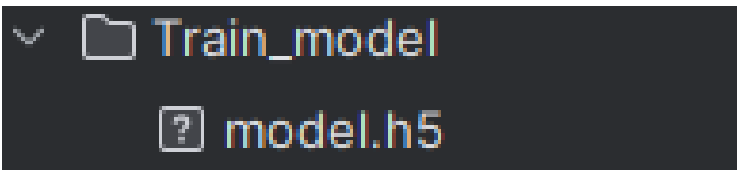


Una vez obtenido ambas carpetas de entrenamiento y validación, se genera el archivo de entrenamiento model.h5, el cual nos apoyara para determinar que al ingresar una imagen este tenga como resultado sus características.


```

Epoch 1/20
1941/1941 ————— 2471s 1s/step - accuracy: 0.1397 - loss: 2.8552 - val_accuracy: 0.1422 - val_loss: 2.6322
Epoch 2/20
1941/1941 ————— 2439s 1s/step - accuracy: 0.1488 - loss: 2.6145 - val_accuracy: 0.1408 - val_loss: 2.6167
Epoch 3/20
1941/1941 ————— 2345s 1s/step - accuracy: 0.1501 - loss: 2.5630 - val_accuracy: 0.1310 - val_loss: 2.6177
Epoch 4/20
1941/1941 ————— 2356s 1s/step - accuracy: 0.1534 - loss: 2.5249 - val_accuracy: 0.1308 - val_loss: 2.6368
Epoch 5/20
1941/1941 ————— 2343s 1s/step - accuracy: 0.1594 - loss: 2.4847 - val_accuracy: 0.1185 - val_loss: 2.6723
Epoch 6/20
1941/1941 ————— 2344s 1s/step - accuracy: 0.1702 - loss: 2.4391 - val_accuracy: 0.1146 - val_loss: 2.7375
Epoch 7/20
1941/1941 ————— 2612s 1s/step - accuracy: 0.1848 - loss: 2.3869 - val_accuracy: 0.1062 - val_loss: 2.8189
Epoch 8/20
1941/1941 ————— 2457s 1s/step - accuracy: 0.2008 - loss: 2.3311 - val_accuracy: 0.0997 - val_loss: 2.9151
Epoch 9/20
1941/1941 ————— 2452s 1s/step - accuracy: 0.2189 - loss: 2.2755 - val_accuracy: 0.0980 - val_loss: 3.0105
Epoch 10/20
1941/1941 ————— 2466s 1s/step - accuracy: 0.2373 - loss: 2.2229 - val_accuracy: 0.0919 - val_loss: 3.0969
Epoch 11/20
1941/1941 ————— 2467s 1s/step - accuracy: 0.2529 - loss: 2.1733 - val_accuracy: 0.0965 - val_loss: 3.2755
Epoch 12/20
1941/1941 ————— 2464s 1s/step - accuracy: 0.2670 - loss: 2.1275 - val_accuracy: 0.0891 - val_loss: 3.4037
Epoch 13/20
1941/1941 ————— 2512s 1s/step - accuracy: 0.2827 - loss: 2.0823 - val_accuracy: 0.0885 - val_loss: 3.5944
Epoch 14/20
1941/1941 ————— 2456s 1s/step - accuracy: 0.2960 - loss: 2.0389 - val_accuracy: 0.0905 - val_loss: 3.8265
Epoch 15/20
1941/1941 ————— 2461s 1s/step - accuracy: 0.3101 - loss: 1.9985 - val_accuracy: 0.0847 - val_loss: 4.0306
Epoch 16/20
1941/1941 ————— 2488s 1s/step - accuracy: 0.3192 - loss: 1.9674 - val_accuracy: 0.0868 - val_loss: 4.1463
Epoch 17/20
1941/1941 ————— 2486s 1s/step - accuracy: 0.3318 - loss: 1.9263 - val_accuracy: 0.0823 - val_loss: 4.2203
Epoch 18/20
1941/1941 ————— 2479s 1s/step - accuracy: 0.3443 - loss: 1.8894 - val_accuracy: 0.0820 - val_loss: 4.5740
Epoch 19/20
1941/1941 ————— 2482s 1s/step - accuracy: 0.3544 - loss: 1.8599 - val_accuracy: 0.0821 - val_loss: 4.7661
Epoch 20/20
1941/1941 ————— 2482s 1s/step - accuracy: 0.3632 - loss: 1.8333 - val_accuracy: 0.0782 - val_loss: 4.6077

```



5. VALIDACIÓN DE RESULTADOS Y PRUEBAS

El sistema es una aplicación que recibe una imagen tipo carnet de una persona y devuelve las características faciales de la persona. Estas características faciales se identifican utilizando técnicas avanzadas de reconocimiento facial.

- Entrada:

Imagen tipo carnet: Una imagen frontal de la cara de la persona, en un formato estándar (JPEG, PNG, etc.), con buena iluminación y calidad suficiente para permitir un análisis preciso de las características faciales.

- Salida:

Características faciales: Una lista de atributos que describe las características faciales de la persona. Estos atributos pueden incluir:

- 5_o_Clock_Shadow
- Arched_Eyebrows
- Bags_Under_Eyes
- Bald
- Bangs
- Big_Lips
- Big_Nose
- Black_Hair
- Blond_Hair
- Brown_Hair
- Bushy_Eyebrows
- Chubby
- Double_Chin
- Eyeglasses
- Goatee
- Gray_Hair
- High_Cheekbones
- Male
- Mustache
- Narrow_Eyes
- No_Beard
- Oval_Face
- Pale_Skin
- Pointy_Nose
- Receding_Hairline

- Rosy_Cheeks
- Sideburns
- Straight_Hair
- Wavy_Hair
- Young

Interpretación de resultados y pruebas:

La correcta interpretación de estos resultados permite a los usuarios finales aplicar la información obtenida en una variedad de contextos. Un contexto especialmente relevante es la búsqueda de personas desaparecidas, donde las características faciales pueden ser cruciales para su identificación y reconocimiento. Este uso permite a las autoridades y organizaciones de búsqueda utilizar los datos proporcionados para encontrar y reconocer a las personas desaparecidas de manera más eficiente y precisa, mejorando así las probabilidades de reunificación con sus seres queridos.

6. CONCLUSIONES

En el análisis exploratorio de datos del conjunto de datos CelebA, se observó una estructura clara y una buena calidad inicial al visualizar las primeras filas del archivo CSV. La ausencia de valores nulos confirmada por métodos adicionales indicó una preparación de datos sólida para análisis más profundos.

La visualización de datos desveló patrones interesantes y variaciones en la distribución de características faciales. Se analizó la presencia y frecuencia de cada atributo facial mediante gráficos de barras e histogramas, lo que permitió una comprensión detallada de las características más prevalentes y su distribución.

Además, el análisis de correlación resaltó relaciones significativas entre diferentes características, proporcionando una base para la selección estratégica de atributos para el entrenamiento del modelo. Este proceso culminó con la refinación del conjunto de datos, asegurando que solo los atributos más útiles fueran retenidos para fases posteriores de modelado, optimizando así el conjunto de datos para el entrenamiento eficaz del modelo.

En conclusión, la implementación de redes neuronales convolucionales para la predicción de características faciales a partir de imágenes tipo carnet ha demostrado ser una herramienta valiosa

en el contexto de la búsqueda de personas desaparecidas. Este sistema permite identificar de manera precisa atributos faciales como la presencia de barba, el color del cabello, el uso de gafas, entre otros, lo cual es crucial para la identificación y reconocimiento de individuos. La capacidad del sistema para analizar y devolver estas características de manera eficiente mejora significativamente las posibilidades de localizar y reunir a personas desaparecidas con sus seres queridos, proporcionando una solución tecnológica avanzada para un problema humanitario crítico.

7. REFERENCIAS BIBLIOGRÁFICAS

- Alva, G. (2024, febrero 27). *Desaparecidos en el Perú: los peruanos que no volvieron a casa*. El Comercio Perú. Recuperado de <https://elcomercio.pe/ecdata/desaparecidos-en-el-peru-95907-peruanos-que-no-volvieron-a-casa-policia-nacional-te-estamos-buscando-menores-mujeres-alerta-amber-especdis-noticia/>
- Guardamino, B. (2024, 18 abril). *3.295 personas desaparecidas en Perú: alarmante cifra se registra solo en los dos primeros meses del 2024*. Infobae. Recuperado de <https://www.infobae.com/peru/2024/04/18/3295-personas-desaparecidas-en-peru-alarman-te-cifra-se-registra-solo-en-los-dos-primeros-meses-del-2024/#:~:text=En%20los%20primeros%20dos%20meses,que%20se%20reportaron%20541%20casos.>
- Jessica Li. (2024). *CelebFaces Attributes (CelebA) Dataset* [Datos de archivo]. Recuperado el 24 de abril de 2024, de <https://www.kaggle.com/datasets/jessicali9530/celeba-dataset>
- Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). *Deep Learning Face Attributes in the Wild*. En *Proceedings of International Conference on Computer Vision (ICCV)*. Recuperado de https://openaccess.thecvf.com/content_iccv_2015/html/Liu_Deep_Learning_Face_ICCV_2015_paper.html
- Ministerio del Interior. (2023). *DIAGNÓSTICO SOBRE EL FENÓMENO DE LA DESAPARICIÓN DE PERSONAS EN EL PERÚ*. Mininter.gob.pe. Recuperado de <https://cdn.www.gob.pe/uploads/document/file/4825911/Diagnostico%20Desaparici%C3%B3n%20Peru.pdf.pdf>