

# Primera entrega de proyecto

Sharid Samantha Madrid Ospina

## I. CONTEXTO DE APLICACIÓN

Este proyecto se centra en el desarrollo de un sistema de reconocimiento de escritura a mano utilizando técnicas avanzadas de DeepLearning. El reconocimiento de texto manuscrito es un desafío significativo en el campo de la visión por computadora y el procesamiento del lenguaje natural, dado que implica interpretar una variedad de estilos de escritura y convertir imágenes de texto en texto legible por máquina. Este sistema tiene aplicaciones potenciales en la digitalización de documentos históricos, automatización de la entrada de datos, y mejora de interfaces hombre-máquina.

## II. OBJETIVO DE MACHINE LEARNING

El objetivo principal de este proyecto es desarrollar un modelo de deep learning que pueda predecir texto a partir de imágenes de escritura a mano. Utilizando una red neuronal convolucional y recurrente (CRNN), el sistema busca transformar imágenes de texto manuscrito en cadenas de texto digital, mejorando la precisión y eficiencia en la interpretación de escritura a mano.

## III. DATASET

Handwriting Recognition [kaggle.com](https://www.kaggle.com/datasets/sharidsamantamadridospina/handwriting-recognition)

El dataset está compuesto por más de 400,000 imágenes de nombres y apellidos escritos a mano, divididos en 206,799 nombres y 207,024 apellidos. Las imágenes están distribuidas en tres conjuntos: entrenamiento, prueba y validación. `written_name_test_v2.csv`

### A. Detalles del Dataset

- Conjunto de Entrenamiento: 331,059 imágenes.
- Conjunto de Prueba: 41,328 imágenes.
- Conjunto de Validación: 41,328 imágenes.

### B. Estructura de Archivos

El dataset se organiza en tres archivos CSV correspondientes a cada conjunto de datos:

- `written_name_test_v2.csv`
- `written_name_train_v2.csv`
- `written_name_validation_v2.csv`

Donde cada archivo tiene 2 columnas: FILENAME (URL de la imagen) e IDENTITY (Transcripción del nombre escrito a mano).

### C. Almacenamiento de Imágenes

Las imágenes están almacenadas en tres carpetas separadas, correspondientes a los conjuntos de entrenamiento, prueba y validación:

- `test_v2`
- `train_v2`
- `validation_v2`

### D. Tipo de Datos:

- **Imágenes:** Archivos de imagen que representan nombres y apellidos escritos a mano.
- **Etiquetas:** Transcripciones de texto de los nombres y apellidos.

### E. Tamaño en Disco

1 GB

## IV. MÉTRICAS DE DESEMPEÑO

Como métrica de machine learning, utilizaremos principalmente la Exactitud a Nivel de Carácter y la Exactitud a Nivel de Palabra para evaluar el rendimiento de nuestro modelo de reconocimiento de escritura a mano.

Estas métricas permiten evaluar con precisión la eficiencia de nuestro modelo en la predicción de texto manuscrito. El objetivo principal es maximizar el valor de estas exactitudes, dado que un valor más alto indica un mejor rendimiento. En un entorno donde la precisión en la interpretación de la escritura a mano es crucial, estas métricas se convierten en elementos críticos para medir la calidad de nuestras predicciones, permitiéndonos tomar decisiones informadas basadas en el rendimiento del modelo.

Además, considero el uso de la Tasa de Error por Conexión (CTC Loss) y la Distancia de Edición para proporcionar una comprensión más detallada del rendimiento del modelo en diferentes niveles de granularidad.

## V. MÉTRICAS DE NEGOCIO

El proyecto se considerará exitoso en términos de métricas de negocio si el modelo proporciona una mejora significativa en la eficiencia y la reducción de costos en la transcripción de textos manuscritos. Así pues, buscamos:

**Reducir el Tiempo de Procesamiento:** Cuanto más efectivo sea el modelo en la transcripción rápida de texto manuscrito, mayor será la reducción en el tiempo de procesamiento.

**Aumentar la Eficiencia Operativa:** Una mejora en la capacidad para procesar volúmenes mayores de texto manuscrito sin incremento proporcional en los recursos.

**Minimizar la Pérdida de Ingresos por Errores de Transcripción:** Al aumentar la precisión del modelo,

reducimos los costos asociados con los errores de transcripción y las correcciones manuales.

## VI. REFERENCIAS Y RESULTADOS

- A. *"An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition"*<sup>1</sup> - Baoguang Shi, Xiang Bai, Cong Yao.

Este estudio presenta una red neuronal que puede entrenarse de extremo a extremo para el reconocimiento de secuencias basadas en imágenes, aplicable al texto en escenas naturales. La eficacia de la red se demuestra a través de su aplicación en reconocimiento de texto en imágenes, un problema similar al reconocimiento de escritura a mano.

- B. *"Handwriting Recognition of Historical Documents with Few Labeled Data"*<sup>2</sup> - Vincent Christlein, Mathias Seuret, Andreas Maier.

Este trabajo se enfoca en el reconocimiento de escritura a

mano en documentos históricos con datos etiquetados limitados, utilizando técnicas de aprendizaje profundo. Es relevante analizar cómo los autores abordan el desafío de los datos limitados, las estrategias de entrenamiento que utilizan, y cómo estas podrían influir en tu enfoque, especialmente si tu dataset tiene limitaciones similares.

## VII. BIBLIOGRAFIA

[1] B. Shi, X. Bai, y C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," arXiv preprint arXiv:1507.05717, 2015. Disponible en: <https://arxiv.org/abs/1507.05717>

[2] V. Christlein, M. Seuret, y A. Maier, "Handwriting Recognition of Historical Documents with Few Labeled Data," arXiv preprint arXiv:1811.07768, 2018. Disponible en: <https://arxiv.org/abs/1811.07768>