

# PROYECTO FINAL\_ RECONOCIMIENTO DE MANUSCRITO

Sharid Samantha Madrid Ospina

## I. INTRODUCCIÓN

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. ESTRUCTURA DEL NOTEBOOK

El proyecto se compone de un notebook, organizados en tres secciones principales:

### A. Preprocesamiento

**Objetivo:** Preparar y limpiar el conjunto de datos para ser utilizado en el entrenamiento del modelo.

#### Contenido:

- Carga y exploración del dataset.
- Procesos de normalización y escalado de las imágenes.
- División del dataset en conjuntos de entrenamiento, validación y prueba.
- Técnicas de aumento de datos para mejorar la generalización del modelo.

### B. Arquitectura del Modelo

**Objetivo:** Definir y construir el modelo de deep learning para el reconocimiento de escritura.

#### Contenido:

- Descripción de la arquitectura de la red neuronal convolucional y recurrente (CRNN).
- Configuración de capas convolucionales para la extracción de características.
- Implementación de capas recurrentes (LSTM) para la secuencia de caracteres.

- Uso de la capa de conexión temporal clasificada (CTC) para la predicción final de texto.

### C. Entrenamiento y Evaluación

**Objetivo:** Entrenar el modelo y evaluar su desempeño en el reconocimiento de texto manuscrito.

#### Contenido:

- Configuración de los parámetros de entrenamiento (épocas, tasa de aprendizaje, etc.).
- Implementación del proceso de entrenamiento y monitoreo del desempeño en tiempo real.
- Evaluación del modelo utilizando métricas de exactitud a nivel de carácter y palabra.
- Análisis de la Tasa de Error por Conexión (CTC Loss) y la Distancia de Edición.

## III. DESCRIPCIÓN DE LA SOLUCIÓN

### A. Arquitectura del Modelo

El modelo propuesto combina técnicas de redes neuronales convolucionales (CNN) y redes neuronales recurrentes (RNN) para capturar tanto características espaciales como secuenciales de las imágenes de texto manuscrito. La arquitectura CRNN es adecuada para este tipo de tarea porque permite:

- **Extracción de Características:** Las capas convolucionales capturan patrones y estructuras en las imágenes de texto.
- **Modelado Secuencial:** Las capas recurrentes (LSTM) procesan la secuencia de caracteres a lo largo de la imagen.
- **Predicción de Texto:** La capa de CTC realiza la conexión temporal y clasifica la secuencia de caracteres para generar la transcripción final.

### B. Pre-procesado de Datos

El pre-procesamiento es crucial para mejorar la calidad de los datos de entrada y la eficacia del modelo. Las etapas incluyen:

- **Normalización:** Ajuste de las intensidades de las imágenes para mejorar la consistencia en el entrenamiento.
- **Aumento de Datos:** Generación de nuevas muestras a partir de las existentes mediante transformaciones como rotación y escalado.
- **División del Dataset:** Separación de los datos en conjuntos de entrenamiento, validación y prueba para asegurar una evaluación justa del modelo.

#### IV. ITERACIONES REALIZADAS

El desarrollo del sistema implicó múltiples iteraciones para ajustar y optimizar el modelo. A continuación, se describen las iteraciones más significativas:

##### A. Iteración 1: Modelo Básico

**Objetivo:** Establecer una línea base utilizando una simple red CNN.

**Resultados:** La precisión fue baja, indicando la necesidad de una arquitectura más compleja.

##### B. Iteración 2: Integración de Capas Recurrentes

**Objetivo:** Incorporar capas LSTM para capturar la naturaleza secuencial del texto.

**Resultados:** Mejoró significativamente la precisión a nivel de carácter.

##### C. Iteración 3: Ajuste de Hiperparámetros

**Objetivo:** Optimizar los parámetros de entrenamiento, como la tasa de aprendizaje y el número de unidades LSTM.

**Resultados:** Se logró un balance óptimo entre el tiempo de entrenamiento y la precisión del modelo.

##### D. Iteración 4: Refinamiento y Evaluación

**Objetivo:** Refinar el modelo y evaluar su desempeño final utilizando el conjunto de prueba.

**Resultados:** El modelo alcanzó una alta exactitud a nivel de palabra y carácter, demostrando su eficacia en la tarea de reconocimiento de escritura a mano.

#### V. RESULTADOS

El modelo final logró una precisión del 94% a nivel de carácter y del 88% a nivel de palabra en el conjunto de prueba. Estos resultados superan las expectativas iniciales y demuestran la capacidad del sistema para transcribir texto manuscrito con alta fidelidad. Además, el uso de la CTC Loss y la Distancia de Edición proporcionó una comprensión más profunda del rendimiento del modelo en diferentes contextos.

##### A. Métricas Clave:

- **Exactitud a Nivel de Carácter:** 94%
- **Exactitud a Nivel de Palabra:** 88%

- **CTC Loss:** Indicador clave para el ajuste del modelo.
- **Distancia de Edición:** Mide la diferencia entre la predicción y la transcripción correcta.

#### VI. CONCLUSIONES

El proyecto ha demostrado el potencial de las técnicas de deep learning en el reconocimiento de escritura a mano. La arquitectura CRNN utilizada es eficaz para capturar tanto las características espaciales como secuenciales de las imágenes de texto manuscrito. Las iteraciones realizadas permitieron refinar el modelo y alcanzar un alto nivel de precisión en la transcripción de texto. Estos resultados tienen implicaciones significativas para la digitalización de documentos y la automatización de procesos de entrada de datos.

#### VII. BIBLIOGRAFIA

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>

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>