

UNIVERSIDAD INTERNACIONAL DEL ECUADOR

MAESTRIA BIG DATA

Grupo 6

Integrantes

Lilian Amán

Santiago Borja

Renato Jacome

Rubén Recalde

Link proyecto Github:

[EvilWayz79/BigDataModelosYaprendizajesGrupo6: Grupo 6 Maestria UIDE \(github.com\)](https://github.com/EvilWayz79/BigDataModelosYaprendizajesGrupo6)

OJO el archivo de training por restricciones de tamaño de GitHub fue recortado en 10k Líneas

Paso 1: Obtener los datos: mediante las librerías básicas: pandas

```
train_data=pd.read_csv("mnist_train.csv") test_data=pd.read_csv("mnist_test.csv")
```

Paso 2: Exploración de los datos: `train_data.info()` #el resultado me indica que tiene 60,000 filas y 785 columnas, todas con valores enteros de 64 bits, y ocupa alrededor de 359.3 megabytes de memoria en tu sistema.

Paso 3.

Después, creamos nuevas características como la suma de los valores de intensidad de los píxeles y las agregamos a las características originales. Esto puede mejorar el rendimiento del modelo.

Paso 4.

Construimos un modelo de clasificación utilizando Random Forest Classifier y lo entrenamos utilizando las características originales más las nuevas características.

Paso 5.

Finalmente, evaluamos el modelo en el conjunto de prueba y visualizamos la matriz de confusión para interpretar el rendimiento del modelo en la clasificación de los números escritos a mano.
