MINST handwritten datasets are benchmark for deep learning models for handwritten digits. Each image is a grayscale image = (28 \* 28) = 784 pixels .

X\_train should be a matrix of size (number\_of\_features =N \* 28 \* 28) and similarly others.