



Escuela de ingeniería y ciencias.

Campus **Monterrey**

Unidad de formación **TC3006C.102**

Inteligencia artificial avanzada para la ciencia de datos II

Semana 5: CNN + Transfer Learning

Equipo **3**

Grupo: **102**

Regina Reyes Juárez - **A01275790**  
Nadia Salgado Alvarez - **A01174509**  
Gilberto Angel Camacho Lara - **A01613895**  
Santiago Miguel Lozano Cedillo - **A01198114**

28 de octubre

## Introducción

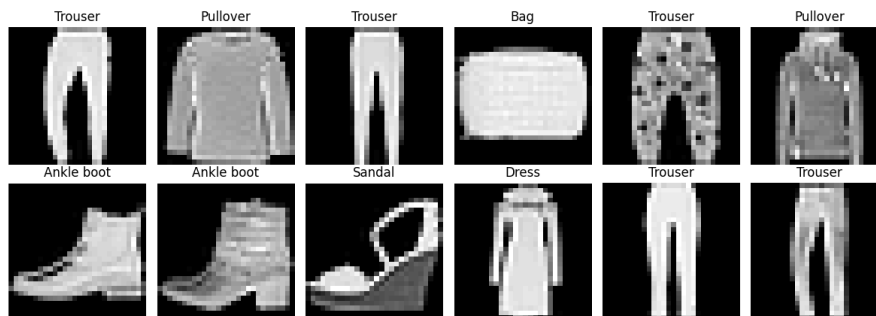
El presente análisis tiene como objetivo explorar y comparar distintas estrategias de clasificación de imágenes aplicadas al dataset Fashion-MNIST, que contiene 70,000 imágenes en escala de grises de prendas de vestir distribuidas en 10 categorías. Se buscó implementar un modelo de CNN desde cero como referencia (baseline) y, adicionalmente, aprovechar Transfer Learning mediante la arquitectura preentrenada MobileNetV2, evaluando su capacidad para generalizar y mejorar la precisión sobre las imágenes del conjunto de prueba. La preparación de los datos incluyó limpieza, normalización y transformación de las imágenes para garantizar la compatibilidad con los distintos modelos, así como la creación de pipelines eficientes para entrenamiento y validación.

El enfoque adoptado combina técnicas de aprendizaje profundo tradicionales y modernas: la CNN baseline fue diseñada con bloques Conv-BN-ReLU, MaxPooling y Dropout para capturar patrones locales y prevenir sobreajuste, mientras que MobileNetV2 permitió reutilizar características visuales previamente aprendidas en ImageNet, aplicando un fine-tuning ligero para optimizar la clasificación en las 10 categorías de Fashion-MNIST. La evaluación se realizó mediante métricas como accuracy, entropía cruzada categórica, precision, recall y F1-score, permitiendo un análisis detallado por clase y una comparación directa del desempeño entre las distintas estrategias de modelado.

## Exploración, explicación y limpieza de datos

Para este proyecto estaremos trabajando con la base de datos Fashion-MNIST, dónde para entrenar usaremos 54,000 imágenes de 28x28 píxeles y 1 canal, para evaluar 6,000 imágenes y para testear 10,000 imágenes.

En esa base de datos tenemos 10 diferentes clases como se puede observar en la *Imagen 1*, las cuales corresponden a: T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag y Ankle boot.



**Imagen 1.** Fotografías de categorías.

A partir de estos datos, se separaron las etiquetas de las imágenes y se reestructuraron los valores de los píxeles de cada muestra en una matriz de 28x28 con un solo canal, correspondiente a una imagen en escala de grises. Luego, se normalizaron los valores de los píxeles dividiendo entre 255 para que todos los valores quedaran en el rango  $[0,1]$ , lo que mejora la estabilidad numérica y la eficiencia del entrenamiento. Posteriormente, el conjunto de entrenamiento se dividió en subconjuntos de entrenamiento y validación de forma estratificada, garantizando que ambas particiones mantuvieran la misma distribución de clases.

Además, las etiquetas se transformaron a formato one-hot encoding, convirtiendo cada clase en un vector binario para poder utilizar funciones de pérdida como la entropía cruzada categórica. Finalmente, se creó un pipeline de datos con tf.data que incluye barajado, agrupamiento en lotes (batching) y prelectura (prefetching), optimizando el flujo de datos hacia la GPU. De manera opcional, se aplicó una función de augmentación ligera que permite generar pequeñas variaciones en las imágenes, contribuyendo a mejorar la generalización del modelo.

## Desarrollo del Modelo de Deep Learning

Hiperparámetros de modelo base para transe learning usando el modelo **MobileNetv2**

Hiperparámetro	Modelo 1
IMG SIZE	128
BATCH_TL	64
LR_FEATURE	1e-3
LR_FINETUNE	1e-5
N_TRAINABLE _LAYERS	20
DROPOUT_RATE	0.3
EPOCHS_FEATURE	5
EPOCHS_FINETUNE	4
Accuracy	0.6946

**Tabla 1.** Hiperparámetros

Se entrenó un modelo sequential utilizando una CNN desde cero, que sirvió como baseline para nuestro análisis. Además, implementamos Transfer Learning aprovechando la arquitectura preentrenada MobileNetV2 para abordar el problema de clasificación de Fashion-MNIST.

Dado que MobileNetV2 espera imágenes RGB de tamaño 160x160 píxeles y nuestro dataset consiste en imágenes en escala de grises de 28x28x1, primero realizamos la conversión a tres canales y el redimensionamiento adecuado. Posteriormente, aplicamos un preprocesamiento consistente con la normalización de ImageNet, asegurando coherencia en las entradas del modelo. Después, congelamos el backbone de MobileNetV2 para reutilizar sus características visuales generales y evitar reentrenar toda la red, agregando una nueva cabeza densa para clasificar en 10 categorías. Inicialmente, entrenamos solo esta cabeza para prevenir sobreajuste, y luego realizamos un fine-tuning ligero descongelando las últimas capas, mejorando así la predicción sin comprometer la estabilidad del modelo.

En cuanto a la arquitectura de la CNN baseline, el modelo se diseñó con bloques Conv-BN-ReLU que capturan patrones locales de las imágenes, combinados con MaxPooling y Dropout para reducir el sobreajuste, y capas densas finales con softmax para realizar la clasificación multiclase.

## Resultados e Interpretación

Para el primer modelo de CNN con Transfer Learning basado en MobileNetV2, se logró que aproximadamente el 93% de las imágenes fueran clasificadas correctamente, y la entropía cruzada categórica alcanzó un valor de 0.19, lo que refleja un alto grado de confianza promedio en las predicciones del modelo.

Durante el entrenamiento, se observó que el uso de Early Stopping y Reduce LR on Plateau contribuyó a prevenir el sobreajuste, evitando que existiera divergencia significativa entre el rendimiento en el

conjunto de entrenamiento y el de validación, asegurando así una generalización más estable del modelo.

Época	Accuracy (train)	Accuracy (val)	Comentario
1	0.72	0.38	Inicio: el modelo apenas aprende
2	0.87	0.75	Aprendizaje fuerte inicial
4	0.90	0.91	Modelo generaliza mejor
6-10	≈0.93	≈0.93	Se estabiliza — excelente generalización

**Tabla 2.** Entrenamiento.

Al analizar las métricas por clase, se observa que las categorías con texturas y formas muy diferenciadas, como zapatos o bolsas, presentan un desempeño altamente preciso, mientras que las clases visualmente más similares, como camisas, suéteres o abrigos, muestran una mayor cantidad de errores de clasificación.

Sin embargo, al evaluar los promedios generales, tanto el macro average como el weighted average alcanzan un valor de 0.93, lo que indica un desempeño balanceado entre las clases y sugiere que el modelo no presenta sesgos significativos hacia ninguna categoría.

## Conclusión

La preparación de los datos —incluyendo normalización, one-hot encoding, estratificación y el uso de pipelines eficientes con `tf.data`— permitió entrenar modelos consistentes y optimizados para la clasificación del dataset Fashion-MNIST. El modelo baseline (CNN desde cero) sirvió como punto de referencia, mientras que el enfoque de Transfer Learning con MobileNetV2 (IMG\_SIZE=128, BATCH=64, LR\_FEATURE=1e-3, LR\_FINETUNE=1e-5, DROPOUT=0.3, 20 capas entrenables) alcanzó una precisión del 69.46 %. Aunque el resultado fue menor al rendimiento típico de MobileNetV2, refleja la capacidad del modelo para generalizar con recursos limitados. En conjunto, los experimentos confirman que el Transfer Learning mejora la estabilidad y eficiencia del entrenamiento frente a modelos desde cero, y que un ajuste adicional de los hiperparámetros podría elevar significativamente la precisión final.