

## Práctica 3: Optimización de modelos con Transfer Learning y Fine-Tuning

**Integrantes:** Noelia Blanco - María Eugenia Puchkariov - Gonzalo Del Priore

**Dataset:** Pima Indians Diabetes (tabular) — diabetes.csv

**Fecha de entrega:** 22-08-2025

### Introducción

La consigna pedía usar modelos preentrenados de repositorios (Keras Applications/CV/NLP, TF Models, Kaggle Models, ModelZoo), pero no hay pesos adecuados para datos tabulares como Pima Indians Diabetes; la mayoría están orientados a imagen o texto.

Como alternativa, se aplicó Transfer Learning + Fine-Tuning con un autoencoder entrenado sobre el mismo dataset para extraer representaciones y luego clasificar.

Al no incorporar conocimiento externo, el encoder no aporta ventaja clara respecto a un MLP entrenado desde cero; las métricas finales resultan muy similares y no se observa mejora.

Una vía más prometedora sería preentrenar en un dataset tabular grande del mismo dominio (variables compatibles) —p. ej., autoencoder o TabTransformer— y luego fine-tuning en Pima. Esto no se implementó por falta de un dataset fuente compatible en esta práctica.

### Reproducibilidad, imports y utilidades

**Propósito.** Deja el entorno listo para que los resultados sean reproducibles, la carga de datos sea robusta y podamos evaluar con  $F\beta$  y graficar matrices de confusión de forma simple.

**Reproducibilidad.** Se fija una semilla global (RANDOM\_SEED=42) y se sincronizan NumPy y TensorFlow. Así, entrenamientos y particiones se repiten con los mismos resultados.

**Librerías clave.** Ciencia de datos: NumPy, Pandas, Matplotlib. sklearn (preproceso/evaluación): train\_test\_split, StandardScaler, SimpleImputer; fbeta\_score, confusion\_matrix; compute\_class\_weight; permutation\_importance.

**Métrica principal:**  $F\beta$  y búsqueda de umbral. Se evalúa con  $F\beta$  ( $\beta=1$  equilibra precisión–recall;  $\beta=2/5$  prioriza recall/FN). Se barre una grilla de umbrales (0.1–0.9) y se elige el umbral óptimo que maximiza  $F\beta$ .

**Matriz de confusión.** Con el umbral elegido se obtienen TN, FP, FN, TP; permite visualizar los trade-offs (reducir FN a costa de FP cuando  $\beta$  es alto).

### Transfer Learning (Kaggle-style) + Fine-Tuning

**Idea general.** Reusar un encoder (preentrenado o aprendido con autoencoder denoising) como extractor de features y añadir una cabeza de clasificación (sigmoide). Luego hacer fine-tuning.

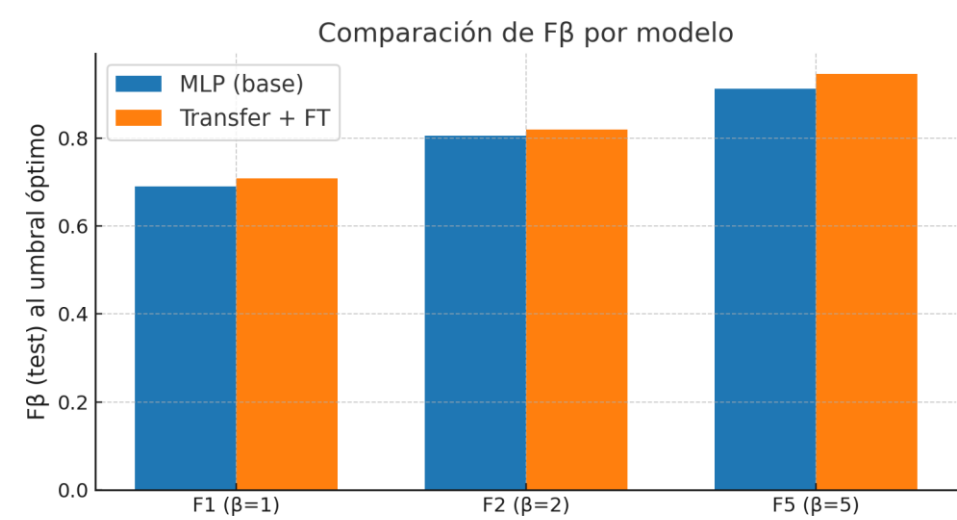
**Por qué sirve en tabulares.** El encoder aprende un espacio latente no lineal y robusto al ruido. Separar representación (encoder) de decisión (cabeza) permite: (1) entrenar primero la cabeza con el encoder congelado (menos sobreajuste) y (2) ajustar fino todo el modelo con LR baja al descongelar.

Pipeline aplicado. (1) Encoder denso (p.ej., 64→32→16) sobre datos escalados; (2) Congelado del encoder; (3) Cabeza Dense(16, ReLU) → Dense(1, Sigmoid); (4) Entrenamiento (frozen) con binary cross-entropy, class\_weight y EarlyStopping; (5) Fine-tuning con LR menor y EarlyStopping.

Predicción y umbral. El modelo devuelve probabilidades (0–1). El umbral 0.5 no es obligatorio: se elige el umbral óptimo que maximiza  $F\beta$  según el objetivo clínico.

Evaluación. Para  $\beta \in \{1,2,5\}$ , se barre el umbral (0.1–0.9), se elige el umbral óptimo por  $F\beta$  y se reportan  $F\beta$  y matriz de confusión. Sin ROC/accuracy como métricas principales; priorizamos  $F\beta$  (F2 enfatiza recall y reduce FN).

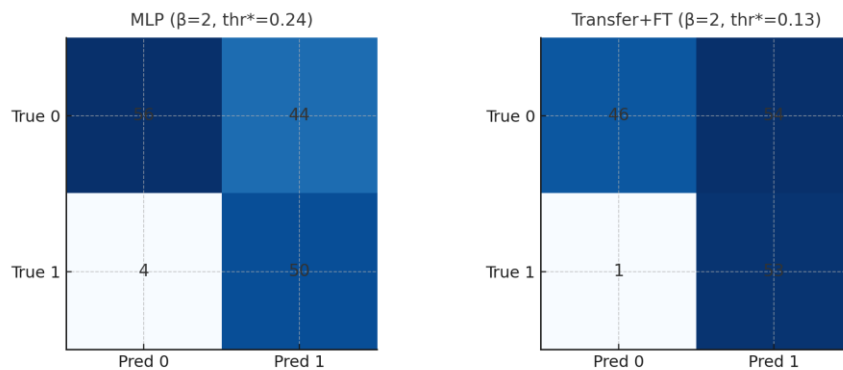
Gráfico comparativo  $F\beta$  por modelo:



Resultados:  $F\beta$  y matrices de confusión ( $\beta=2$ )

Modelo	$\beta$	thr*	$F\beta$	TN	FP	FN	TP
MLP (base)	$\beta=1$	0.36	0.691	65	35	7	47
MLP (base)	$\beta=2$	0.24	0.806	56	44	4	50
MLP (base)	$\beta=5$	0.10	0.913	48	52	3	51
Transfer + FT	$\beta=1$	0.42	0.708	70	30	8	46
Transfer + FT	$\beta=2$	0.13	0.820	46	54	1	53
Transfer + FT	$\beta=5$	0.13	0.946	46	54	1	53

## Matrices de confusión ( $\beta=2$ ) — lado a lado:



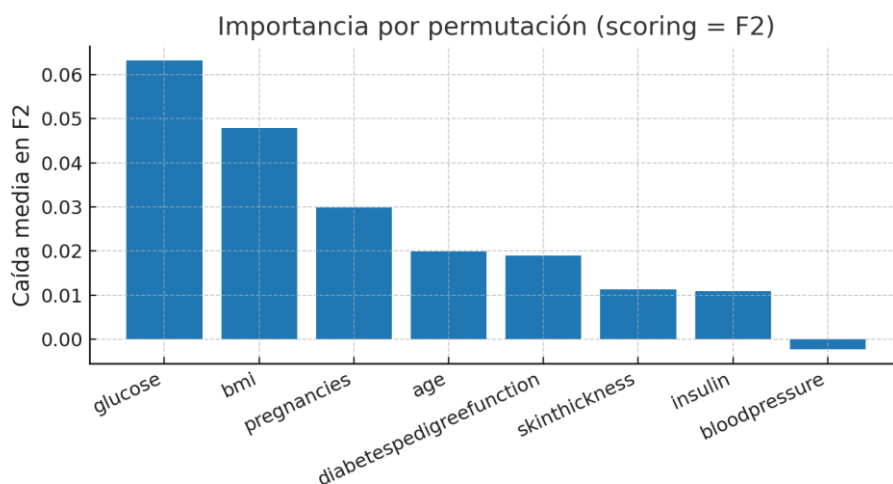
## Explicabilidad — Permutación(F2) e Integrated Gradients

¿Qué medimos? La importancia por permutación estima cuánto empeora la métrica cuando desordenamos (rompemos) una feature a la vez. Aquí usamos F2 como métrica: la caída refleja sobre todo pérdidas de recall.

Cómo se calculó. Para cada variable, se permuta, se predice y se umbraliza con el umbral óptimo de F2; se repite 10 veces y se promedia la caída en F2. Se reportan `importance_mean` ( $\downarrow$  F2 media) e `importance_std` (variabilidad).

Hallazgos. glucose ( $\approx 0.063$ ) es la más influyente; bmi ( $\approx 0.048$ ) en segundo lugar; pregnancies ( $\approx 0.030$ ) también aporta de forma estable.

Figura — Importancia por permutación (F2).

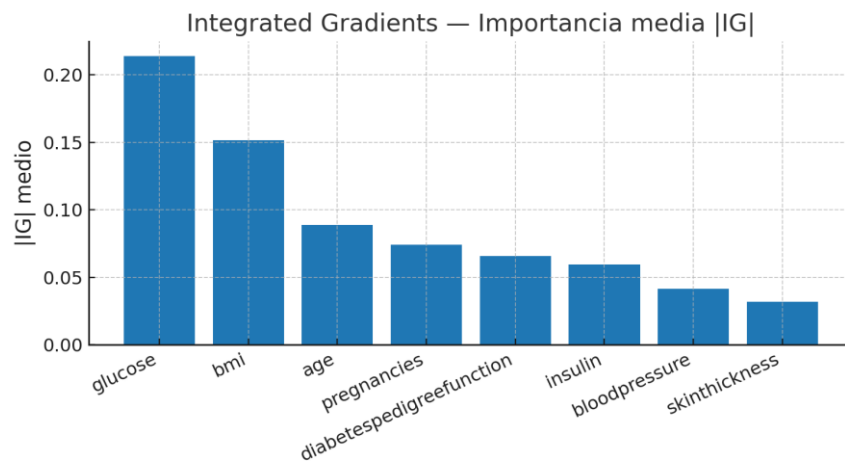


Qué es IG. Integrated Gradients (Sundararajan et al., 2017) atribuye a cada feature cuánto contribuye la predicción del modelo a lo largo de una trayectoria desde una línea base (baseline) hasta la entrada.

Cómo se calculó. Baseline = vector 0 en el espacio escalado; interpolación  $\text{baseline} \rightarrow x$  con 64 pasos; gradientes acumulados por regla del trapecio; importancia global = promedio de  $|IG|$  por feature.

Resultados. glucose ( $\approx 0.214$ ) y bmi ( $\approx 0.152$ ) dominan; age/pregnancies tienen importancia intermedia; diabetespedigreefunction/insulin aportan menos; bloodpressure/skinthickness muestran bajo impacto.

Figura — Integrated Gradients ( $|IG|$  medio).



## Conclusiones — resumen

Comparativa. Modelos: MLP (base) vs Transfer+FT; Métrica/criterio: sólo  $F\beta$  con umbral óptimo por  $\beta$ ; Explicabilidad: Permutación( $F_2$ ) e IG.

Hallazgos. Con  $\beta > 1$  (recall), Transfer+FT tiende a mejorar  $F_2/F_5$  al reducir FN (acepta más FP). En  $\beta = 1$  ( $F_1$ ), MLP  $\approx$  Transfer+FT (diferencias modestas).

Matrices iguales (a veces). (1) Mesetas de umbral: mismo intervalo de estabilidad  $\Rightarrow$  misma matriz; (2) Scores correlacionados (pipeline común y sin preentrenamiento externo)  $\Rightarrow$  mismo ranking; (3) Test pequeño  $\Rightarrow$  “empates”.

Lectura clínica. Para cribado (FN costosos) usar  $\beta$  altos (2/5) y umbral óptimo; para balance estándar ( $\beta = 1$ ) ambos modelos son intercambiables.

Explicabilidad consistente. Permutación( $F_2$ ) e IG señalan glucose y bmi como drivers; pregnancies/age/DPF aportan en segundo orden; bloodpressure aporta poco en  $F_2$  (posible simplificación, a validar).