

Tamizaje de Riesgo de Diabetes con Deep Learning (HITL)

Maestría en Ciencia de Datos – Universidad de Montevideo

Integrantes: Noelia Blanco, María Eugenia Puchkariov, Gonzalo del Priore

Fecha: Agosto 2025



Introducción: Contexto y Problema

Problema de Salud Pública

La diabetes representa un problema mayor de **salud pública** debido a su alta prevalencia y complicaciones que afectan significativamente la calidad de vida de los pacientes.

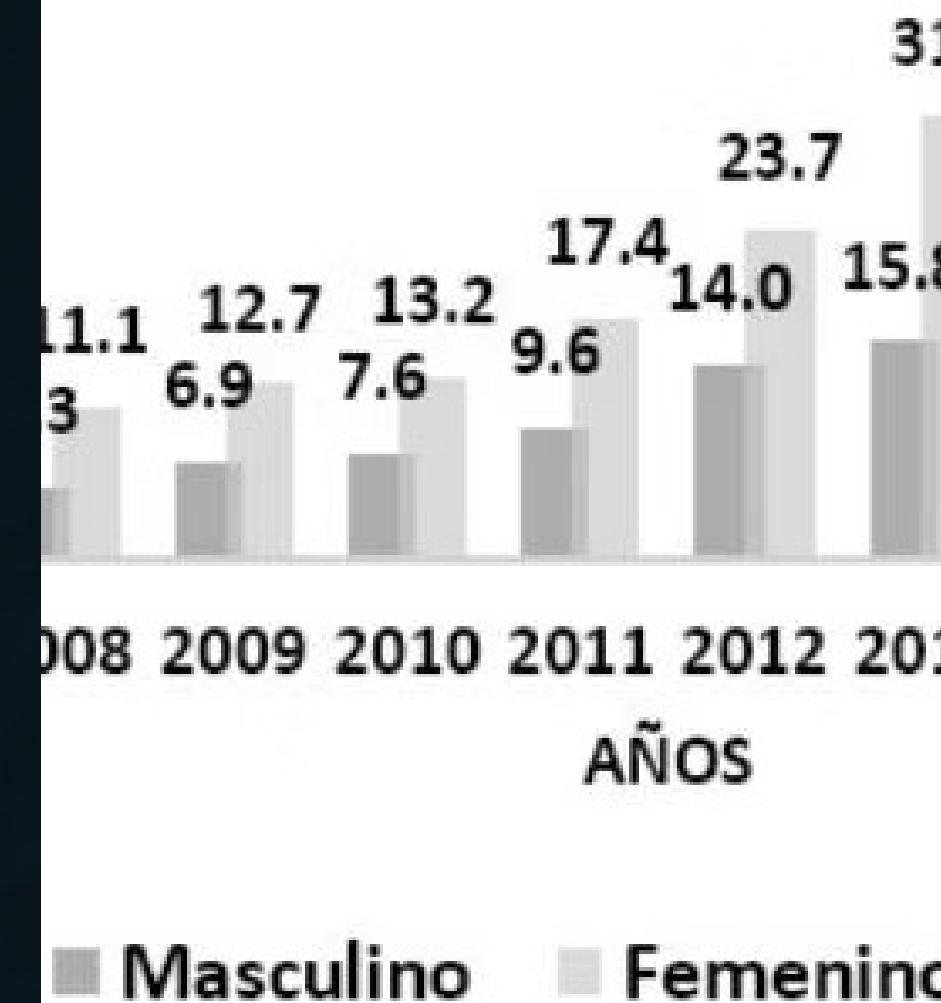
Limitaciones en Atención Primaria

La atención primaria enfrenta **recursos diagnósticos limitados**, lo que lleva a una detección **tardía** de personas en riesgo.

Nuestra Propuesta

Desarrollar un **sistema de tamizaje** con Deep Learning usando variables clínicas simples que permita priorizar estudios confirmatorios y optimizar recursos.

Este proyecto aplica el **ciclo de vida completo** de Deep Learning para crear una solución **viable, calibrada y explicable** que responda a una necesidad real en el sistema de salud.



Marco Teórico

El éxito de un proyecto de Deep Learning trasciende la mera construcción de un modelo. Requiere una integración efectiva en un entorno real, generando valor sostenible alineado con objetivos estratégicos.

Antes de iniciar un proyecto, es fundamental plantearse preguntas críticas:

- ¿Es necesario aplicar Deep Learning?
- ¿Se dispone de datos de calidad?
- ¿Cuál es el impacto esperado?
- ¿Cómo se gestionarán los riesgos asociados?

Arquetipos de proyectos

Software 2.0

Enfoque basado en datos donde el comportamiento del sistema emerge del entrenamiento con ejemplos.

Human-in-the-Loop (HITL)

Combina la capacidad predictiva de los modelos con la revisión humana de casos fronterizos.

En este trabajo adoptamos el enfoque **Software 2.0 complementado con HITL**, equilibrando impacto y factibilidad mientras aseguramos aceptabilidad y confianza en el contexto sanitario.



Objetivo del Proyecto

Diseñar, implementar y validar un **pipeline de Deep Learning** para tamizaje de riesgo de diabetes en atención primaria.

Meta operativa

Minimizar falsos negativos (FN) por su mayor impacto clínico.

Métricas principales

- F_2 y $F_5 \rightarrow$ foco en recall
- AUPRC (Área bajo la curva Precisión-Recall)
- Umbral de Costo (FN más costoso que FP)

Pipeline (calidad y confianza)

- Calibración de probabilidades
- Validación robusta: cross-validation + bootstrap
- Explicabilidad para auditoría clínica

Prototipo funcional

App en Streamlit que muestra la integración práctica en la toma de decisiones clínicas.

Dataset y Preparación

Dataset Pima Indians Diabetes

- N \approx 768 registros de mujeres Pima \geq 21 años
- 8 variables clínicas: embarazos, glucosa, presión arterial, pliegue cutáneo (SkinThickness), insulina, IMC (BMI), pedigrí de diabetes (DPF) y edad
- Objetivo: variable binaria diabetes (1/0)

Preprocesamiento

- **Ceros imposibles** \rightarrow NA en: glucosa, presión, pliegue, insulina, BMI
- Imputación por mediana (variables numéricas)
- Escalado estandarizado para estabilizar el entrenamiento
- Partición estratificada train/test, preservando la prevalencia (~35% positivos)



Datos crudos

768 registros con valores imposibles



Transformación

Imputación y escalado

Limpieza

Identificación de ceros imposibles y conversión a NA

Partición

Train/test estratificado

Modelos Entrenados

Regresión Logística

Línea base interpretable y calibrable; referencia clínica estándar.

MLP (Red Neuronal Básica)

Captura no linealidades e interacciones con complejidad moderada.

DNN Profunda

Varias capas + BatchNorm, Dropout, Gaussian Noise → mejor regularización y generalización.

Transfer Learning

Autoencoder → Fine-Tuning: aprende representaciones latentes y luego ajusta la cabeza de clasificación.

Estrategia de entrenamiento común

Todos los modelos utilizaron **early stopping, tuning de hiperparámetros y validación estratificada**.

Criterio de selección

Evaluamos F_2/F_5 , AUPRC, calibración (Brier/ECE) y costo $FN \gg FP$ para elegir el modelo más robusto en el punto operativo clínico.

Resultados Comparativos

Modelo	AUPRC	F ₂	FN@F ₂	F ₅	Costo
LR	0.673	0.831	1	0.948	55.0
LR+SMOTE	0.676	0.828	2	0.947	56.0
TL+FT	0.698	0.825	2	0.943	57.0
MLP	0.671	0.823	1	0.946	58.0
DNN	0.622	0.765	7	0.848	79.0

Razonamiento clínico

Definimos el punto de decisión que minimiza el impacto clínico/operativo: **Costo = 5×FN + 1×FP** (FN = 5× más grave que FP).

Aunque TL+FT tiene mejor AUPRC, elegimos **Regresión Logística** porque:

- Mantiene alta sensibilidad (menos FN)
- Logra el menor costo total
- Reduce al mínimo los casos no detectados sin sobrecargar al sistema con FP

Calibración y Explicabilidad

Calibración de Probabilidades

Por qué importa: probabilidades bien calibradas = decisiones clínicas confiables (umbral coherente con riesgo real).

Evaluación: Brier Score y ECE (menor = mejor), diagramas de confiabilidad.

Hallazgos: Regresión Logística calibrada (isotónica/sigmoide) → mejor alineación entre probabilidades y frecuencias reales.

Impacto práctico: mayor utilidad clínica al fijar F_2/F_5 o coste 3:1/5:1/7:1 sin sorpresas por mala calibración.

Explicabilidad del Modelo

Objetivo: transparencia y aceptación clínica.

Métodos:

- **Permutation Importance** (global) → ordena variables por impacto
- **LIME y SHAP** (local) → explican cada predicción

Hallazgos clave: mayor influencia de glucosa, IMC (BMI) y edad.

Valor clínico: facilita auditoría, detección de casos atípicos y confianza profesional.



Prototipo en Streamlit

Características del prototipo

- Carga artefactos: preprocessamiento y modelos
- Selector de clasificador (RL, MLP, etc.)
- **Umbral ajustable** (slider) para la decisión
- **Banda HITAL** configurable (zona de duda)

Lógica operativa

- Fuera de la banda → decisión automática
- **Dentro** de la banda → derivación a revisión clínica

El prototipo refleja el enfoque **Human-in-the-Loop** y facilita la integración en la práctica clínica real.



[Enlace al prototipo](#)



Entrada de datos

Variables clínicas simples

Predicción

Modelo calibrado

Decisión

Automática o HITAL

Explicación

LIME/SHAP

Conclusiones



Aplicación integral

Pipeline de **Deep Learning** aplicado a un problema real de **salud pública**, siguiendo el ciclo de vida completo desde la concepción hasta la validación práctica.



Enfoque clínico

Métricas centradas en **reducir FN + calibración** (Brier/ECE) + **explicabilidad** (Permutation, LIME/SHAP) → **validez y confianza**.



Selección del modelo

Aunque los modelos complejos rindieron similar, **Regresión Logística calibrada** destacó por **simplicidad, estabilidad y transparencia** (apta para entorno clínico).



Human-in-the-Loop

Aumenta **aceptabilidad** al combinar predicción automática con **revisión clínica** en casos fronterizos, equilibrando tecnología y juicio humano.

Este proyecto demuestra cómo el **Deep Learning** puede integrarse efectivamente en el ámbito sanitario cuando se diseña con un enfoque centrado en el impacto clínico, la calibración adecuada y la explicabilidad de las decisiones.