UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN

MODELO TRANSFORMER PARA IDENTIFICAR LA NEFROPATÍA DIABÉTICA EN PACIENTES CON DMT2

TESIS PARA OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA:

LUIS RAMÓN TERCERO MARTÍNEZ GONZÁLEZ

BAJO LA DIRECCIÓN DE:

DR. JOSÉ ADÁN HERNÁNDEZ NOLASCO

EN CODIRECCIÓN:

DR. NOEL ZACARÍAS MORALES

CUNDUACÁN, TABASCO, A: ENERO 2026

UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN

MODELO TRANSFORMER PARA IDENTIFICAR LA NEFROPATÍA DIABÉTICA EN PACIENTES CON DMT2

TESIS PARA OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA:

LUIS RAMÓN TERCERO MARTÍNEZ GONZÁLEZ

BAJO LA DIRECCIÓN DE:

DR. JOSÉ ADÁN HERNÁNDEZ NOLASCO

EN CODIRECCIÓN:

DR. NOEL ZACARÍAS MORALES

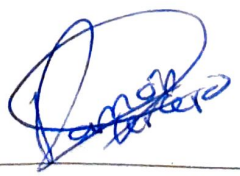
CUNDUACÁN, TABASCO, A: ENERO 2026

Declaración de Autoría y Originalidad

En la Ciudad de Cunduacán el día Doce del mes de Enero del año 2026, el que suscribe **Luis Ramón Tercero Martínez González**, alumno del Programa **Maestro en Ciencias de la Computación** con número de matrícula **232H21004**, adscrito a la **División Académica de Ciencias y Tecnologías de la Información**, de la Universidad Juárez Autónoma de Tabasco, como autor de la Tesis presentada para la obtención de Grado de Maestría y titulada **Modelo Transformer para identificar la nefropatía diabética en pacientes con DMT2**, dirigida por el Dr. José Adán Hernández Nolasco y la Dr. Noel Zacarías Morales.

DECLARO QUE: La Tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la LEY FEDERAL DEL DERECHO DE AUTOR (Decreto por el que se reforman y adicionan diversas disposiciones de la Ley Federal del Derecho de Autor del 01 de Julio de 2020 regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita. Del mismo modo, asumo frente a la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad o contenido de la Tesis presentada de conformidad con el ordenamiento jurídico vigente.

Cunduacán, Tabasco a 12 de Enero de 2026.



Estudiante: Luis Ramón Tercero Martínez González

Cunduacán, Tabasco, a 08 de enero de 2026
Oficio No. 0045/DACYTI/D

Asunto: Autorización de impresión de Tesis

C. Luis Ramon Tercero Martínez González
Egresado de la Maestría en Ciencias de la Computación

En virtud de que cumple satisfactoriamente los requisitos establecidos en el Reglamento General de Estudios de Posgrado vigente en la Universidad, informo a Usted que se autoriza la impresión del trabajo recepcional ***"Modelo Transformer para identificar la nefropatía diabética en pacientes con DMT2"***, para presentar examen y obtener el Grado de Maestro en Ciencias de la Computación.

Sin otro particular, aprovecho la ocasión para enviarle un afectuoso saludo.

Atentamente

Dr. Óscar Alberto González González
Director

Universidad Juárez Autónoma de Tabasco
División Académica de Ciencias y Tecnologías de la Información

C.c.p. Mtra. Yenny Lorena Dussán Rojas. – Encargada del despacho de la Coordinación de Posgrado.
Archivo.
Consecutivo.

DR.*OAGG/YLDR

Carta de Cesión de Derechos

Villahermosa, Tabasco a 12 de Enero de 2026.

Por medio de la presente manifiesto haber colaborado como AUTOR en la producción, creación y/o realización de la obra denominada: **Modelo Transformer para identificar la nefropatía diabética en pacientes con DMT2.**

Con fundamento en el artículo 83 de la Ley Federal del Derecho de Autor y toda vez que, la creación y/o realización de la obra antes mencionada se realizó bajo la comisión de la Universidad Juárez Autónoma de Tabasco; entendemos y aceptamos el alcance del artículo en mención de que tenemos el derecho al reconocimiento como autores de la obra, y a la Universidad Juárez Autónoma de Tabasco mantendrá en un 100% la titularidad de los derechos patrimoniales por un período de 20 años sobre la obra en la que colaboramos, por lo anterior, cedemos el derecho patrimonial exclusivo en favor de la Universidad.

COLABORADOR

Estudiante: Luis Ramón Tercero Martínez González

TESTIGOS

Dr. José Adán Hernández Nolasco

Dr. Noel Zacarías Morales

Dedicatoria

A mi esposa que camina a mi lado siempre, María Teresa Flores Dorantes. Gracias por sostener mi mano en los momentos difíciles y celebrar conmigo las victorias. Gracias por entender mis ausencias y motivarme a ser mejor cada día. Sin ti, este camino no habría tenido el mismo sentido. Te amo.

Este trabajo está dedicado a mis padres. Gracias, por cada sacrificio hecho para que yo pudiera finalizar; espero que este logro sea una pequeña recompensa a todo lo que me han dado.

A mis hermanos, por estar siempre presentes y celebrar cada uno de mis pasos y mis logros.

Finalmente, a mis colegas y amigos de la maestría. Gracias por el apoyo mutuo, las discusiones académicas y la amistad que trascenderá estas aulas.

Índice general

Índice de figuras

Índice de tablas

Resumen

La nefropatía diabética (ND) es una de las principales complicaciones de la diabetes mellitus tipo 2 y constituye una causa relevante de insuficiencia renal en la población. Su diagnóstico temprano es un reto clínico, pues depende de múltiples factores clínicos y de laboratorio. En este contexto, el uso de modelos de inteligencia artificial ofrece una alternativa prometedora para apoyar la toma de decisiones médicas. En esta tesis se desarrolló un modelo de clasificación basado en arquitecturas *Transformer*, adaptadas al análisis de datos tabulares médicos, con el propósito de identificar la presencia de ND en pacientes mexicanos. La investigación se sustentó en la base de datos abierta *DiabetIA*, de la cual se trabajó con una muestra representativa de más de 7 mil pacientes, considerando tanto positivos como negativos al diagnóstico de ND. El proceso metodológico incluyó: (i) preprocesamiento de los datos para eliminar inconsistencias y balancear las clases; (ii) selección de características mediante Lasso y RFE, con el fin de identificar las variables clínicas más relevantes; y (iii) entrenamiento de un modelo *Transformer* diseñado específicamente para datos mixtos (continuos y categóricos). Los experimentos comparativos se realizaron contra modelos clásicos de *Machine Learning* (Regresión Logística, Random Forest, XGBoost) y contra arquitecturas modernas de *Deep Learning* para datos tabulares (TabNet, TabTransformer y GatedTabTransformer). Los resultados demuestran que el modelo propuesto alcanzó una precisión al *99 %*, alcanzando a los modelos de referencia y validando la hipótesis planteada de que un enfoque basado en *Transformers* puede mejorar significativamente la clasificación de ND. Asimismo, el análisis de casos de estudio confirmó la aplicabilidad clínica del modelo. Las principales contribuciones de este trabajo incluyen: (i) la propuesta metodológica de un marco reproducible para la clasificación de ND con *Transformers*, y (ii) la incorporación de lineamientos éticos y técnicos para su posible implementación en sistemas de apoyo al diagnóstico.

Palabras clave: Inteligencia Artificial, Transformers, Aprendizaje profundo, Nefropatía diabética

Abstract

Diabetic nephropathy (DN) is one of the main complications of type 2 diabetes mellitus and constitutes a relevant cause of renal failure in the population. Its early diagnosis represents a clinical challenge, as it depends on multiple clinical and laboratory factors. In this context, the use of artificial intelligence models offers a promising alternative to support medical decision-making. In this thesis, a classification model based on *Transformer* architectures, adapted for the analysis of medical tabular data, was developed with the purpose of identifying the presence of DN in Mexican patients. The research was supported by the open database *DiabetIA*, from which a representative sample of more than 7,000 patients was used, including both positive and negative cases for DN diagnosis. The methodological process included: (i) data preprocessing to eliminate inconsistencies and balance the classes; (ii) feature selection using Lasso and RFE, in order to identify the most relevant clinical variables; and (iii) training of a *Transformer* model specifically designed for mixed data (continuous and categorical). Comparative experiments were conducted against classical *Machine Learning* models (Logistic Regression, Random Forest, XGBoost) and against modern *Deep Learning* architectures for tabular data (TabNet, TabTransformer, and GatedTabTransformer). The results demonstrate that the proposed model achieved an accuracy of *99 %*, matching the reference models and validating the stated hypothesis that a *Transformer*-based approach can significantly improve DN classification. Additionally, the analysis of case studies confirmed the clinical applicability of the model. The main contributions of this work include: (i) the methodological proposal of a reproducible framework for DN classification using *Transformers*, and (ii) the incorporation of ethical and technical guidelines for its potential implementation in diagnostic support systems.

Keywords: Artificial Intelligence, Neural Networks, Diabetic nephropathy, Transformers.

Capítulo 1

Generalidades

1.1. Introducción

En los últimos años, la medicina ha dado la bienvenida a la Inteligencia Artificial (IA) como una herramienta complementaria en el diagnóstico clínico que permitirá elevar el bienestar de los pacientes en términos de salud. Diversas investigaciones se dedican a buscar las posibles aplicaciones de la IA en el campo de la medicina, abordando aspectos que van desde la fase de anticipación y detección hasta la atención y monitoreo de diversas enfermedades (Mayer, 2023).

Por otro lado, la carga global de las enfermedades crónicas como la diabetes mellitus (DM) ha aumentado constantemente. De acuerdo con las proyecciones de la Federación Internacional de Diabetes (FID), se calcula que aumentará a 643 millones en 2030 y a 784 millones en el 2045. Además, entre el 30 % y el 50 % de las personas con diabetes podrían desarrollar enfermedad renal crónica (ERC) en algún momento de sus vidas. Actualmente, la ERC es la principal causa a nivel mundial de enfermedad renal en etapa terminal (Sun et al., 2022).

Asimismo, la evolución de la ERC presenta una considerable variabilidad en su curso natural, lo que dificulta la clasificación y la predicción de su progresión. Algunos pacientes avanzan a través de las cinco etapas clásicas de la nefropatía diabética (ND), mientras que otros, con microalbuminuria, pueden regresar a niveles normales de albúmina en la orina y no experimentar un empeoramiento, sin embargo, algunos casos con niveles normales de albúmina pueden desarrollar una disminución progresiva de la función renal sin pasar por una etapa de microalbuminuria. Por lo tanto, identificar tempranamente a aquellos factores de riesgo para el desarrollo de ERC,

permitirá el inicio oportuno de terapias nefroprotectoras que prevengan o retrasen el progreso de la enfermedad renal (González-Robledo et al., 2020).

Actualmente, existen muchos artículos que utilizan modelos estadísticos clásicos y aprendizaje automático, pero la mayoría no suelen considerar características específicas de profesionales del área en la clasificación y se basan en lo que el algoritmo selecciona. Este trabajo pretende involucrar áreas multidisciplinares que abarquen aspectos relacionados con la patología y el tratamiento farmacológico, ya que estudios recientes reportaron que un estricto control glucémico y de presión arterial elevada, permitió frenar o retrasar la evolución a estadios más avanzados de la ERC.

En la siguiente figura 1.1 se visualiza un diagrama general de una red neuronal artificial:

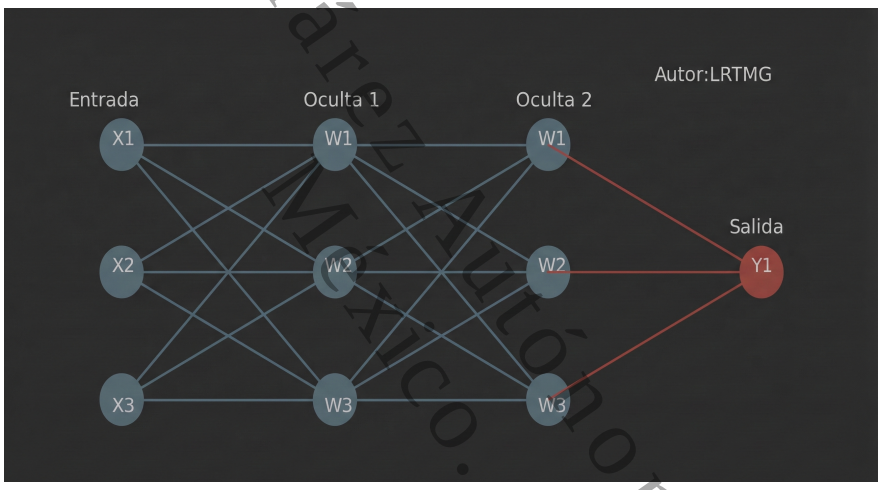


Figura 1.1. Representación gráfica de la Red Neuronal Artificial, Tipo MLP con 2 capas ocultas.

1.2. Planteamiento del problema

Definición del problema

La diabetes mellitus tipo 2 (DMT2) es una enfermedad crónica no transmisible que afecta a millones de personas en todo el mundo y constituye un importante problema de salud pública. Entre sus complicaciones más frecuentes se encuentra la nefropatía diabética (ND), una condición que puede evolucionar hacia enfermedad renal crónica (ERC) y, en etapas avanzadas, derivar en insuficiencia renal terminal, requiriendo tratamientos invasivos como diálisis o trasplante de riñón

(Howlader, 2022).

La detección temprana de la ND resulta crítica para implementar estrategias preventivas y terapéuticas que retrasen o eviten la progresión del daño renal. Sin embargo, los métodos clínicos convencionales presentan limitaciones: suelen ser tardíos, costosos, invasivos y en muchos casos no permiten identificar de forma precisa a los pacientes en riesgo (García-Maset et al., 2022).

Sin embargo, este problema se enfrenta a múltiples desafíos técnicos y clínicos, se deben tomar en cuenta la complejidad de los datos médicos y las diferencias interindividuales de los pacientes; los datos médicos utilizados para la predicción de la ND son inherentemente complejos, multidimensionales y pueden incluir información clínica, resultados de pruebas de laboratorio, registros de pacientes a lo largo del tiempo (García-Maset et al., 2022).

De igual manera, la ND es una enfermedad multifactorial en la que los patrones sutiles pueden ser indicativos de riesgo. La Interpretabilidad y confianza clínica, se refiere a cómo lograr y garantizar que los médicos y profesionales de la salud puedan confiar en los resultados finales del sistema que realiza el diágnostico (Ruiz-Mejía Ramón, 2020).

Por otro lado, el uso y manipulación de la información clínica de pacientes debe seguir estrictos protocolos bioéticos como los señalados en la declaración de Helsinki, además de las leyes de protección de datos. Es por lo anterior que, la base datos utilizada en la investigación que incluye datos reales, está anonimizada (Asociación Médica Mundial, 21 de Marzo de 2017).

Por último, se necesita evaluar parámetros asociados con el desarrollo de ND en pacientes con un diagnóstico reciente de DMT2, ya que los métodos actuales son tardados, evasivos y dolorosos. Es necesario la detección previa para evitar la progresión, ya que la ND es prácticamente irreversible. Por lo tanto, para disminuir la problemática requerirá un enfoque multidisciplinario que involucre a expertos en aprendizaje automático, medicina, farmacéuticos, ética de la salud y regulaciones de privacidad de datos, para lograr una atención y tratamiento personalizada y oportunos.

En el ámbito computacional, se han propuesto modelos basados en estadística y aprendizaje automático para predecir la evolución de la ND. No obstante, al analizar estos enfoques de los autores presentan dos principales limitaciones: (1) no siempre consideran la complejidad multidimensional de los datos clínicos y las diferencias interindividuales de los pacientes, y (2) carecen de interpretabilidad suficiente para generar confianza en el personal médico.

En México, la magnitud del problema es aún más relevante: la prevalencia de diabetes y sus complicaciones continúa en aumento, mientras que el acceso a herramientas diagnósticas innovadoras sigue siendo limitado (Sun et al., 2022). Esto plantea la necesidad de desarrollar modelos de clasificación más eficientes y confiables que apoyen al personal de salud en la toma de decisiones clínicas.

En este contexto, surge la necesidad de aplicar modelos de aprendizaje profundo basados en arquitecturas transformer a datos clínicos tabulares de pacientes con DMT2. Estos modelos han demostrado ventajas frente a otras redes neuronales por su capacidad de procesar relaciones complejas en los datos y optimizar recursos computacionales. No obstante, aún no se cuenta con estudios suficientes que evalúen su eficacia para la predicción temprana de la ND en población mexicana, lo que constituye la brecha de conocimiento principal que aborda la presente investigación.

Delimitación de la investigación

La presente investigación se centra exclusivamente en la complicación de nefropatía diabética (ND) en pacientes con diabetes mellitus tipo 2 (DMT2). El modelo propuesto se entrena y valida con la base de datos DiabetIA, la cual contiene información clínica y de laboratorio anonimizada de pacientes mexicanos. El estudio aborda únicamente el diagnóstico de ND mediante técnicas de aprendizaje automático, sin considerar otros desenlaces clínicos relacionados con la DMT2.

Alcances

- Diseñar, entrenar y evaluar un modelo de clasificación basado en arquitecturas transformer para predecir ND en pacientes con DMT2.
- Probar el desempeño del modelo con atributos clínicos y de laboratorio relevantes, aplicando técnicas de selección de características.
- La investigación se centra solo en la complicación Nefropatía.
- Se utilizó una base de datos del Sistema de Informática de Medicina Familiar (SIMF)-IMSS del estado de Michoacán, México.

Limitaciones

- Se emplea únicamente la base de datos DiabetIA, lo que restringe la generalización de los resultados a otras poblaciones.
- El diseño experimental utiliza hiperparámetros estándar en redes neuronales, lo que podría limitar la optimización del rendimiento.
- La calidad de los datos se ve afectada por problemas comunes en entornos clínicos: desbalance de clases, valores faltantes y registros incompletos.
- El modelo propuesto se enfoca en la clasificación diagnóstica, sin abarcar aún la progresión longitudinal de la enfermedad ni el impacto de tratamientos.

1.3. Pregunta de investigación e hipótesis

- ¿Un modelo transformer aplicado a la base de datos de pacientes con DMT2, obtendrá una precisión de al menos un 90 %, en el diagnóstico de la ND?
- ¿Qué arquitectura debe tener un modelo transformer para obtener un diagnóstico de la ND?

1.4. Objetivos

Objetivo general Construir un modelo transformer para el apoyo al diagnóstico de la ND en pacientes con DMT2.

Objetivos específicos

- Informar de modelos modernos de clasificación binaria.
- Evaluar técnicas de selección de características relevantes.
- Clasificar con un grado de precisión de un 90 % la nefropatía diabética

1.5. Justificación

La nefropatía diabética (ND) constituye una de las complicaciones crónicas más graves de la diabetes mellitus tipo 2 (DMT2). Este padecimiento se origina por lesiones renales progresivas que, en fases avanzadas, pueden derivar en enfermedad renal terminal (ERT), requiriendo terapias sustitutivas como diálisis o trasplante de riñón (Esparragoza et al., 2023).

Pacientes que presentan simultáneamente DMT2 y enfermedad renal crónica (ERC) tienen mayor riesgo de eventos agudos —como hipoglucemia, cetoacidosis diabética o complicaciones cardiovasculares— y de padecer otras comorbilidades a largo plazo, como retinopatía, neuropatía o pie diabético. Esta situación impacta de manera directa en su calidad de vida y representa una carga económica significativa para los sistemas de salud (Chiluisa-Castillo, Bueno-Castro et al., 2023).

A nivel global, la Organización Mundial de la Salud (OMS) estima que 422 millones de personas viven con diabetes mellitus, de los cuales aproximadamente 64 millones corresponden a la región de las Américas. Se prevé que esta cifra aumente un 55 % hacia 2035. En este contexto, México enfrenta una situación crítica debido a la alta prevalencia de diabetes y a la limitada disponibilidad de herramientas diagnósticas oportunas Sun et al., 2022. Ante esta problemática, se vuelve indispensable el desarrollo de soluciones innovadoras que permitan detectar de manera temprana la ND para prevenir la progresión hacia etapas irreversibles. La inteligencia artificial, y en particular los modelos de aprendizaje profundo, ofrecen ventajas para procesar grandes volúmenes de datos clínicos y descubrir patrones complejos no detectables por métodos tradicionales (Navas-Atiaja y Veloz, 2023).

En 2015, los países miembros de las Naciones Unidas (ONU), en colaboración con Organizaciones No Gubernamentales (ONG) y la participación de ciudadanos a nivel global, presentaron una iniciativa para formular 17 Objetivos de Desarrollo Sostenible (ODS), uno de estos es salud y bienestar; agregando que recientemente México da prioridad de investigación en el sector de salud, por lo cual la aplicación de la inteligencia artificial en la medicina ha tenido un gran auge (Objetivos de Desarrollo Sosteniblel, 11 de Aril de 2017).

Los Transformers son una arquitectura de red neuronal inicialmente concebida para transformar datos en el ámbito del procesamiento del lenguaje natural. Con el tiempo, esta estructura

se ha convertido en la herramienta principal para abordar una amplia gama de problemas, que incluyen el procesamiento del lenguaje natural, sonido, imagen, aprendizaje por refuerzo, aprendizaje supervisado y otros desafíos que involucran datos de entrada diversos. Su diseño surgió con el propósito de superar las limitaciones de las redes neuronales recurrentes, convolucionales y de grafos. En cuanto al proceso de entrenamiento, en la práctica, se evidencia que requiere menos épocas, y el tratamiento en paralelo de la computación resulta más fácil de implementar en comparación con otras arquitecturas (de la Torre, 2023).

Se ha comprobado que opera de manera óptima en el procesamiento de otras señales específicas, además de demostrar eficacia en la solución de problemas que involucran la combinación de datos provenientes de diversas fuentes. Esto la posiciona como una herramienta versátil para abordar problemas de índole general. La comprensión detallada de la implementación de la arquitectura de los Transformadores adquiere gran relevancia, dado que puede representar una solución universal para la mayoría de los desafíos en el ámbito de la ciencia de datos (de la Torre, 2023).

En la actualidad, el área de la salud cuenta con pocas herramientas computacional precisa e innovadora que pueda identificar el grado de afectación del paciente con ND, en función de ciertas características específicas. Por esta razón, se plantea la implementación de un modelo computacional transformer para apoyar este problema. Así mismo, se espera que este modelo sea un apoyo y que permita proporcionar a los pacientes con DMT2 la identificación de ND de manera temprana.

1.6. Metodología CRISP-DM

Para guiar el desarrollo de esta investigación se emplea la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), ampliamente utilizada en proyectos de ciencia de datos por su carácter iterativo y flexible. Esta metodología consta de seis fases interconectadas. Por consiguiente, se empleará en esta investigación debido a su adecuación a las necesidades del estudio.

Desarrollo de los pasos de la Metodología CRISP-DM (Cross Industry Standard Process for Data Mining) nació en el seno de dos empresas, DaimlerChrysler y SPSS, que en su día fueron pioneras en la aplicación de técnicas de *data mining* en los procesos de negocio. CRISP-DM es una metodología basada en la práctica y experiencia real de analistas que han contribuido activamente a su desarrollo. Esta metodología se organiza en fases, procesos, documentos entregables y actividades que se interrelacionan de forma cíclica, permitiendo un flujo continuo de mejora y retroalimentación.

En la Figura 1.2 se muestra una representación gráfica de las fases que componen la metodología CRISP-DM. Como puede observarse, el proceso inicia con la comprensión del negocio y los datos, seguido de la preparación y modelado, para finalmente pasar a la evaluación y el despliegue de los resultados. Este ciclo se retroalimenta constantemente, garantizando la mejora continua y la adaptación del modelo a las necesidades cambiantes del negocio.

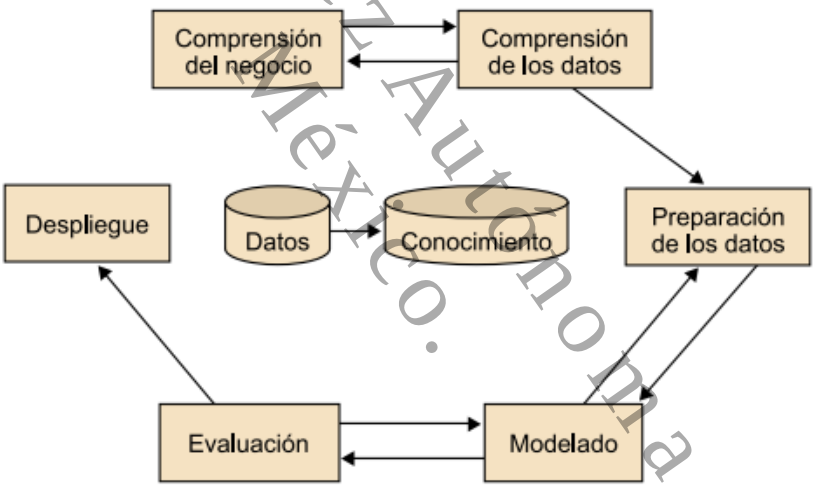


Figura 1.2. Representación gráfica de las fases de CRISP-MD.

Resumen de las fases de la metodología La metodología CRISP-DM estructura el proceso de minería de datos en seis fases principales, que se desarrollan de manera iterativa y no lineal. Cada fase tiene objetivos específicos y actividades que permiten avanzar de la comprensión del problema hasta la implementación del modelo final. La Tabla 1.1 presenta un resumen general de cada una de las fases, junto con sus principales actividades y productos esperados. Este esquema resulta útil para visualizar la secuencia lógica del proceso y la relación entre las distintas etapas

que lo componen.

Tabla 1.1. Fases de la metodología CRISP -DM y actividades principales

Fase	Descripción
Fase 1: Comprensión del problema	▪ Definición de la situación actual ▪ Establecimiento de los objetivos ▪ Selección de criterios de éxito ▪ Definición de restricciones y supuestos
Fase 2: Comprensión de los datos	▪ Búsqueda de fuentes ▪ Descripción de los datos ▪ Verificación de calidad ▪ Exploración de datos
Fase 3: Preparación de los datos	▪ Limpieza ▪ Estructuración ▪ Integración
Fase 4: Modelado	▪ Selección de técnicas de modelado ▪ Generación del plan de prueba ▪ Construcción de modelos
Fase 5: Evaluación	▪ Evaluación de resultados
Fase 6: Implantación	▪ Definición del plan de monitorización e implementación ▪ Revisión del proceso y líneas futuras

Compresión del problema En esta etapa se define el problema que se va a resolver y se identifican los objetivos del proyecto. También se establecen las restricciones y las limitaciones de tiempo y recursos, así como se define el conjunto de datos que se va a utilizar. Esta fase busca comprender el contexto del problema, la naturaleza de los datos disponibles y los requerimientos de los interesados.

La Figura 1.3 muestra las principales actividades que se realizan durante esta fase, las cuales incluyen la identificación de los objetivos del negocio, el análisis de la situación actual, la definición de criterios de éxito y la elaboración del plan de proyecto. Estas actividades permiten establecer una base sólida para orientar los siguientes pasos del proceso de minería de datos.

- Esta fase también incluye la identificación de los requisitos, la definición del alcance y la creación de un plan preliminar.

Comprensión de los datos En esta etapa se realiza una exploración detallada de los datos disponibles para el proyecto. Se busca entender la estructura, la calidad y la consistencia de los datos, identificando posibles problemas como valores faltantes, duplicidad de registros, errores o

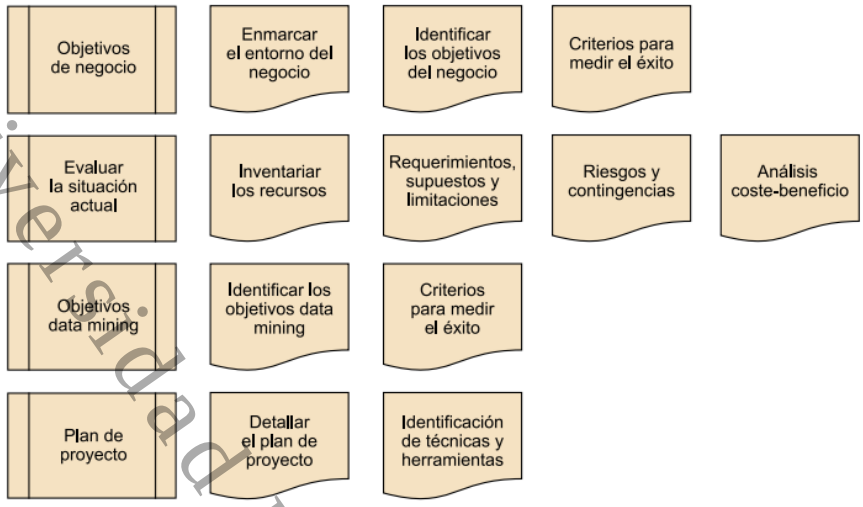


Figura 1.3. Compresión del problema

inconsistencias en los formatos. El objetivo principal es conocer a fondo la naturaleza de los datos para planificar las tareas de preprocesamiento y limpieza que se ejecutarán posteriormente.

La Figura 1.4 presenta una representación de las actividades principales que conforman esta fase. Como puede observarse, el proceso incluye la captura, descripción y exploración de los datos, así como la verificación de su calidad. Además, se elaboran informes de requerimientos y volúmenes de atributos, y se definen hipótesis o propiedades iniciales que guiarán la selección de las técnicas de análisis más adecuadas.

- Se realiza una exploración inicial de los datos disponibles para el proyecto.
- Esta fase también incluye la identificación de problemas de calidad de datos y la selección de las técnicas de ciencias de datos más apropiadas para el proyecto.

Preparación de los datos En esta etapa se realiza el preprocesamiento y la limpieza de los datos, eliminando registros duplicados, valores faltantes y datos inconsistentes. Además, se lleva a cabo la selección de variables, la transformación de los datos y la integración de múltiples fuentes de información. El propósito principal es garantizar que los datos estén en condiciones adecuadas para su posterior análisis y modelado.

La Figura 1.5 ilustra las principales actividades que se ejecutan durante esta fase, tales como la selección y depuración de datos, la generación de conjuntos de entrenamiento y prueba, así co-

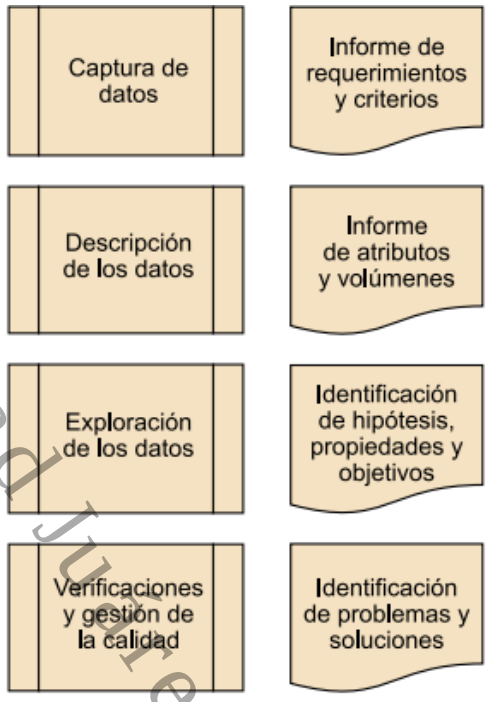


Figura 1.4. Comprensión de los datos.

mo la transformación y el formateo de los datos. Estas acciones permiten asegurar la coherencia y calidad de la información antes de su utilización en las etapas de modelado y evaluación.

- Involucra la elección, depuración y modificación de los datos para prepararlos para su aplicación en el modelo.
- Esta fase también incluye la creación de conjuntos de datos de entrenamiento y prueba.

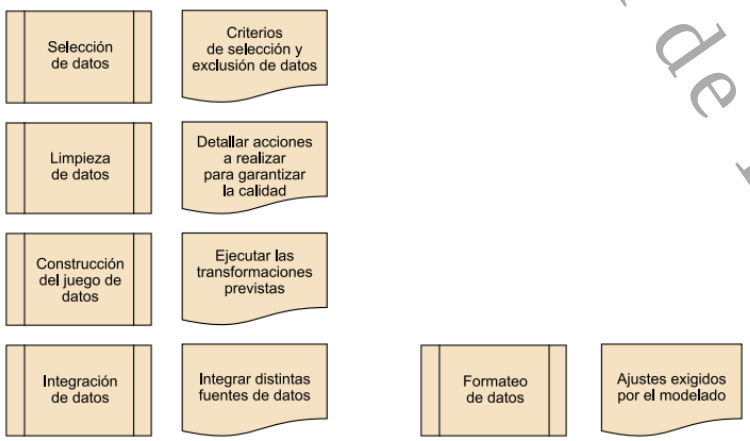


Figura 1.5. Preparación de los datos

Construcción del Modelo En esta etapa se seleccionan los modelos de análisis más adecuados para el problema en cuestión y se construyen los modelos de predicción o clasificación. Se ajustan los parámetros, se entrenan los modelos y se realizan pruebas para evaluar su precisión y efectividad. Esta fase es fundamental, ya que permite comprender el comportamiento de los datos y generar un modelo que aporte valor para la toma de decisiones.

La Figura 1.6 muestra las actividades principales que conforman esta etapa, las cuales incluyen la selección de técnicas y herramientas, la construcción del modelo, el ajuste de parámetros y la evaluación de los resultados. Asimismo, se llevan a cabo pruebas de validación y revisión de los modelos obtenidos para garantizar su desempeño y coherencia con los objetivos del proyecto.

- Se hace la construcción de un modelo utilizando las técnicas de minería de datos seleccionadas. Esta fase también incluye la validación y evaluación del modelo.
- Ajuste del modelo mediante la optimización de hiperparámetros.

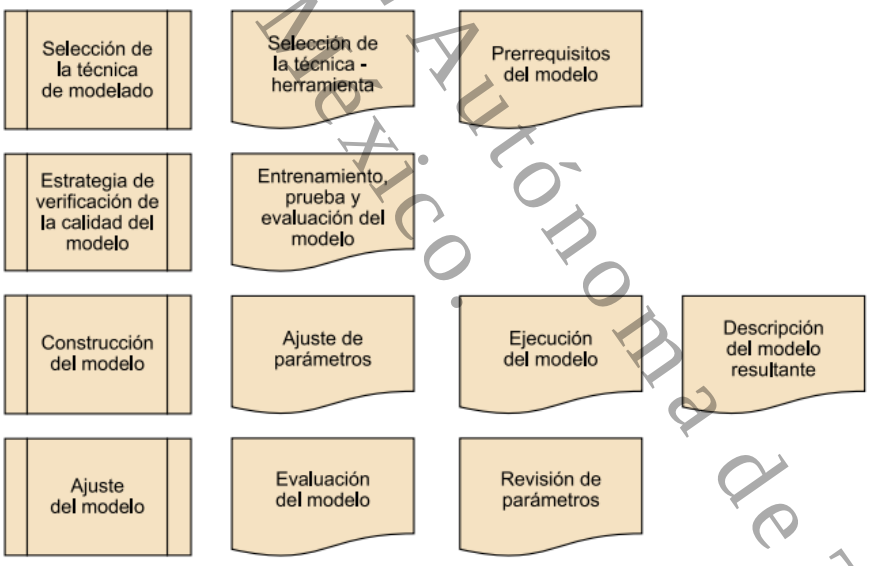


Figura 1.6. Construcción del Modelo

Evaluación En esta etapa se evalúan los modelos de análisis para determinar su eficacia en la resolución del problema planteado. Se analiza la calidad de los resultados obtenidos y se comparan con los objetivos definidos en la fase de comprensión del problema. Además, se realiza una validación cruzada para medir la capacidad de generalización del modelo y asegurar su confiabilidad antes de la implementación final.

La Figura 1.7 muestra las principales actividades de esta fase, que incluyen la verificación de los resultados frente a los criterios de éxito, la revisión del proceso, la identificación de errores o áreas de mejora y la formulación de recomendaciones para las siguientes etapas. El propósito de esta fase es garantizar que el modelo cumpla con los requerimientos establecidos y que sus resultados sean interpretables y útiles para la toma de decisiones.

- Se evalúa la efectividad del modelo y se determina si cumple con los objetivos del proyecto.
- Si el modelo no cumple con los objetivos del proyecto, se repite el proceso de modelado.

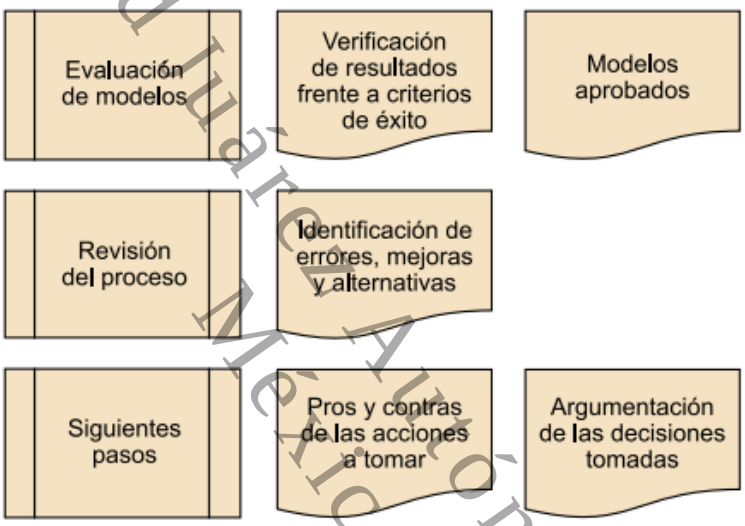


Figura 1.7. Evaluación

Despliegue En esta etapa se implementan los resultados obtenidos durante la fase de modelado, llevando los modelos a la práctica dentro del entorno operativo o productivo. Se elabora un plan de acción para el despliegue y se ejecutan actividades de seguimiento destinadas a verificar el rendimiento y la efectividad del modelo. Asimismo, se documenta todo el proceso para facilitar su replicación, mantenimiento y la correcta interpretación de los resultados.

La Figura 1.8 presenta las actividades más relevantes de esta fase, las cuales incluyen la planificación de la puesta en producción, la elaboración de informes finales, la revisión del proyecto y la recopilación de lecciones aprendidas.

- Se implementa el modelo y se proporcionan las herramientas necesarias para su uso continuo.

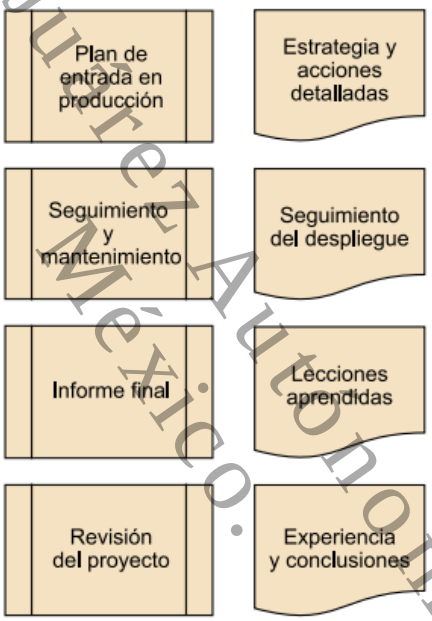


Figura 1.8. Despliegue

Capítulo 2

Marco teórico

2.1. Conceptos y teorías fundamentales de la investigación

Fundamentos médicos relevantes. La diabetes mellitus tipo 2 (DMT2) es una enfermedad crónica caracterizada por resistencia a la insulina e hiperglucemia sostenida. Se estima que entre el 30 % y 50 % de los pacientes con diabetes desarrollarán enfermedad renal crónica (ERC) a lo largo de su vida (Sun et al., 2022).

Entre las complicaciones más graves se encuentra la nefropatía diabética (ND), que afecta hasta al 50 % de los pacientes diabéticos y constituye la principal causa de ERC en etapa terminal (González-Robledo et al., 2020). Su progresión implica alteraciones estructurales y funcionales del riñón, derivando en pérdida irreversible de la función renal.

Factores de Riesgo. Un factor de riesgo es cualquier rasgo, característica o exposición de un individuo que aumente su probabilidad de desarrollar una enfermedad o lesión. Entre los factores de riesgo más relevantes asociados con la nefropatía diabética se incluyen la insuficiencia ponderal, la hipertensión arterial, el consumo de tabaco y alcohol, el agua insalubre, las deficiencias en saneamiento y la falta de higiene *(OMS)*.

La tabla 2.1 resume los principales factores de riesgo identificados en la literatura médica especializada. Como puede observarse, la hiperglucemia prolongada, la hipertensión arterial y la historia familiar de diabetes figuran entre las causas más significativas, seguidas por el tabaquismo, la dislipidemia y la edad avanzada. Estos factores contribuyen al deterioro progresivo de

la función renal, lo que resalta la importancia de su detección temprana y control preventivo en pacientes diabéticos.

No.	Factor de Riesgo
1	Hiperglucemia prolongada
2	Hipertensión arterial
3	Historia familiar de diabetes
4	Obesidad
5	Tabaquismo
6	Dislipidemia
7	Edad avanzada
8	Sexo masculino
9	Presencia de microalbuminuria
10	Duración de la diabetes

Tabla 2.1. Factores de riesgo de la nefropatía diabética

La detección temprana de la ND depende de pruebas como microalbuminuria, proteinuria, creatinina sérica y tasa de filtración glomerular (TFG). Sin embargo, estos métodos suelen ser tardíos, invasivos y poco sensibles a patrones iniciales de riesgo.

Aprendizaje automático (AA). El aprendizaje automático se define como un subcampo de las ciencias de la computación y una rama de la inteligencia artificial que busca desarrollar técnicas que posibiliten que las computadoras adquieran conocimiento y habilidades a través de la experiencia (Suganyadevi et al., 2022).

Aprendizaje profundo (AP). El AP es un método de la inteligencia artificial (IA) que enseña a las computadoras a procesar datos de una manera que se inspira en el cerebro humano. Los modelos de AP son capaces de reconocer patrones complejos en imágenes, textos, sonidos y otros datos, a fin de generar información y predicciones precisas. Es posible utilizar métodos de AP para automatizar tareas que habitualmente requieren inteligencia humana, como la descripción de imágenes o la transcripción a texto de un archivo de sonido (Suganyadevi et al., 2022).

Tipos de Aprendizaje Profundo

1.-Aprendizaje Supervisado. Los desafíos en el aprendizaje supervisado se identifican como sistemas en los cuales los datos de entrenamiento comprenden ejemplos de vectores de entrada y sus correspondientes vectores objetivo. Existen dos categorías principales: la clasificación, que implica la asignación de etiquetas a clases, y la regresión, que implica la predicción de valores numéricos. En problemas de clasificación, el objetivo es prever la etiqueta de clase, mientras que en problemas de regresión, se busca prever una etiqueta numérica significativa(Suganyadevi et al., 2022).

2.-Aprendizaje no supervisado. Este tipo de aprendizaje identifica algunas dificultades relacionadas con el uso de un modelo de relación de datos que describe o elimina relaciones de datos. En comparación con el aprendizaje supervisado, solo se utilizan los datos de entrada, sin resultados ni variables objetivo, no tiene un instructor que corrija el modelo. Hay varias formas de aprendizaje no supervisado, pero tienen dos problemas clave que un profesional encuentra con frecuencia: agrupar los datos y estimar el rango, lo que implica un resumen de la distribución de los datos. La agrupación se representa como un problema de aprendizaje no supervisado que requiere encontrar datos para las clases (Suganyadevi et al., 2022).

3.-Aprendizaje por refuerzo (AR). El AR es un conjunto de desafíos en los que un individuo debe aprender a utilizar la retroalimentación para trabajar en un contexto determinado. Es idéntico al aprendizaje supervisado, aunque la retroalimentación puede retrasarse, y dado que el modelo es sistemáticamente ruidoso, tiene algunas respuestas de las cuales aprender, lo que resulta difícil para la entidad y el modelo vincular la causalidad. El aprendizaje por refuerzo profundo, el Q-learning y el aprendizaje por diferencia temporal son algunos ejemplos comunes de algoritmos de AR (Suganyadevi et al., 2022).

Arquitecturas clásicas y Modernas de Aprendizaje Profundo

1.-Redes neuronales artificiales (RNA). Una RNA es una modelo computacional que imita la forma en que el cerebro humano procesa información, operando mediante la simultaneidad de numerosas unidades de procesamiento interconectadas que actúan como versiones abstractas de neuronas(Suganyadevi et al., 2022).

2.-Algoritmo *backpropagation*. El algoritmo de backpropagation, también conocido como retropropagación, es un método de cálculo que se utiliza para entrenar redes neuronales artificiales. Este algoritmo ajusta los pesos y sesgos de la red neuronal para minimizar el error entre la salida y el resultado esperado.

3.-Redes Neuronales Multicapa. El Perceptrón Multicapa (MLP, por sus siglas en inglés) es una red neuronal artificial que se utiliza para el aprendizaje supervisado en el campo del aprendizaje automático. El MLP está compuesto por múltiples capas de neuronas interconectadas, en las que las salidas de las neuronas de una capa se convierten en entradas para la siguiente capa. La primera capa se llama capa de entrada, la última capa se llama capa de salida y las capas intermedias se llaman capas ocultas.

El MLP es capaz de realizar tareas de clasificación y regresión, y puede ser utilizado para problemas en los que la relación entre las entradas y las salidas no es lineal. El algoritmo de entrenamiento utilizado por el MLP se basa en la propagación hacia atrás (backpropagation), que ajusta los pesos de las conexiones de la red para minimizar el error de predicción entre las salidas producidas por la red y las salidas deseadas.

4.-Red neuronal recurrente (RNN). Los RNN tienen la capacidad de reconocer las secuencias. Los pesos de las neuronas se reparten en todas las medidas. Esto incluye precisiones de última generación en reconocimiento de caracteres, reconocimiento de voz y algunos otros problemas relacionados con el procesamiento del lenguaje natural. El aprendizaje de eventos secuenciales puede modelar condiciones de tiempo. La desventaja es que este método tiene más problemas debido a la desaparición del gradiente y esta arquitectura necesita grandes conjuntos

de datos (Suganyadevi et al., 2022).

5.-Red neuronal convolucionales CNN. Una red neuronal convolucional (CNN,por sus siglas en inglés, también conocida como ConvNet, es un tipo especializado de algoritmo de aprendizaje profundo diseñado principalmente para tareas que requieren reconocimiento de iamgenes, como la clasificación, la detección y la segmentación de imágenes. Las CNN se emplean en diversos casos prácticos, como vehículos autónomos, sistemas de cámaras de seguridad y otros.

6.-Modelo Transformer. Un modelo transformer es una estructura de red neuronal artificial que adquiere contexto y, por ende, comprensión al analizar las relaciones presentes en datos secuenciales, tales como las palabras que conforman la oración. Así mismo, la arquitectura de los transformadores, que recientemente ha experimentado un auge en las aplicaciones en tareas de visión computacional, se han combinado con paradigma convolucional generalizado. Basándose en el proceso de tokenización que divide las entradas en múltiples tokens, los transformers son capaces de extraer sus relaciones por pares mediante la autoatención (Qian et al., 2022).

7.-Modelo Mamba. Es una nueva arquitectura modelo de lenguaje largo (LLM, por sus siglas en inglés) que integra el modelo de secuencia de Espacio de Estado Estructurado (S4) para gestionar secuencias de datos largas. Combinando las mejores características de los modelos recurrentes, convolucionales y de tiempo continuo, el S4 puede simular eficaz y eficientemente las dependencias a largo plazo. Esto le permite manejar datos muestreados irregularmente, tener un contexto ilimitado y mantener la eficiencia computacional durante el entrenamiento y la prueba.

Ampliando el paradigma S4, Mamba aporta varias mejoras dignas de mención, especialmente en el manejo de las operaciones variables en el tiempo. Su arquitectura gira en torno a un mecanismo de selección especial que modifica los parámetros del modelo de espacio de estados estructurado en función de la entrada.

8.-Conjunto de entrenamiento, validación y prueba. Los datos de entrenamiento son los datos que usamos para entrenar un modelo. La calidad de nuestro modelo de aprendizaje automático va a ser directamente proporcional a la calidad de los datos. Por ello las labores de limpieza, depuración consumen un porcentaje importante del tiempo de los científicos de datos.

Los datos de prueba, validación sirven para controlar el entrenamiento y configurar los hiperparámetros. Los datos de prueba sirven para evaluar la capacidad de generalización del modelo entrenado.

9.-Herramientas para la construcción de modelos con marcos de aprendizaje profundo.

1. Tensorflow: TensorFlow es una biblioteca de código abierto desarrollada por Google que permite gestionar e implementar procesos de aprendizaje automático.
2. Torch: PyTorch es un marco de aprendizaje profundo de código abierto basado en software que se emplea para crear redes neuronales, combinando la biblioteca de aprendizaje automático de Torch con una API de alto nivel basada en Python.
3. PyTorch también permite a los científicos de datos ejecutar y probar partes del código en tiempo real, en lugar de esperar a que se implemente todo el código, lo que, para los grandes modelos de aprendizaje profundo, puede llevar mucho tiempo.

2.2. Estado de Arte

En la actualidad, existen investigaciones con soluciones computacionales para el diagnóstico, evolución, tratamiento y prevención de la diabetes mellitus tipo 2 (DMT2). Algunas de ellas contemplan las comorbilidades que las acompañan como es el caso de la nefropatía, a continuación se describen algunos de estos sistemas de clasificación.

2.2.1. Aprendizaje Automático en la clasificación de nefropatía diabética

(Chittora et al., 2021) Destaca la importancia de la enfermedad renal crónica (ERC) como una enfermedad crítica en la actualidad y la necesidad de un diagnóstico adecuado y temprano. Menciona que las técnicas de aprendizaje automático se han vuelto confiables en el tratamiento

médico. Indica que utilizó un conjunto de datos de ERC del repositorio de la University of California, Irvine (UCI). Comparo siete algoritmos de aprendizaje automático distintos. Resalto que los resultados de la investigación, especialmente la precisión, es alcanzada. También menciona las métricas de evaluación utilizadas, como precisión, recuperación, área bajo la curva y coeficiente GINI (el coeficiente de Gini es una medida de la desigualdad). Resalta que también aplico una red neuronal profunda en el estudio y que obtuvo una mayor precisión que los algoritmos de aprendizaje automático.

(Allen et al., 2022) Aplicó algoritmos de aprendizaje automático para predecir la Enfermedad Renal Crónica Diabética (ERCD), utilizando una base de datos que contiene el diagnóstico de DMT2 de 5 años atrás de la investigación. La investigación concluyo que este procedimiento puede proporcionar predicciones oportunas de ERD en pacientes recién diagnosticados con DMT2.

(Dong et al., 2022) Se enfoca en un modelo de predicción de riesgo de ERCD a 3 años en pacientes con DMT2, utilizando aprendizaje automático y registros médicos electrónicos (RME). Menciona que se utilizó un conjunto de datos de 816 pacientes con DMT2 y 3 años de seguimiento en el hospital general. Se utilizaron 46 características médicas disponibles en los registros médicos para desarrollar modelos de predicción utilizando siete algoritmos de aprendizaje automático diferentes. Destaca que el rendimiento de los modelos se midió utilizando el área bajo la curva (ABC) de característica operativa del receptor (COR). Esto muestra una metodología sólida para evaluar la efectividad del modelo. Indica que el modelo LightGBM tuvo el ABC más alto, lo que sugiere que es el mejor modelo de predicción. También señala que el modelo LightGBM podría ser útil para el diagnóstico de la población en el cuidado de DMT2.

(David et al., 2022) Comienza explicando que la ERC es una complicación de la DMT2 que provoca a la pérdida progresiva de la función renal y su importancia en la detección temprana para mejorar los resultados. Menciona que se analizó el rendimiento de nueve técnicas de clasificación diferentes en un conjunto de datos de ERC con 410 instancias y 18 atributos. Se realizó un preprocesamiento de datos y se utilizó una validación cruzada de 10-fold. Por otro lado, destaca las métricas utilizadas para evaluar el rendimiento de las técnicas de clasificación, como el tiempo de ejecución, la precisión, las instancias clasificadas correctamente e incorrectamente, las estadísticas kappa (K), el error absoluto medio, el error cuadrático medio y los valores verdaderos en la matriz de confusión. Resalta que las técnicas de clasificación IBK (Instance-Based K)

y bosque aleatorio obtuvieron una precisión del 93.65 % y un valor K más alto (0.87), lo que las convierte en las técnicas de mejor rendimiento. También menciona que tuvieron la tasa más baja de error cuadrático medio (0.24) y que se identificaron instancias falsas positivas y falsas negativas. Concluye que el estudio identificó las técnicas de clasificación IBK y random tree como las mejores clasificadoras y métodos precisos para predecir ERC.

(Bai et al., 2022) Resume se centra en la evaluación de la viabilidad del uso de técnicas de aprendizaje automático para predecir el riesgo de Enfermedad Renal en Etapa Terminal (ERET) en pacientes con Enfermedad Renal Crónica (ERC). Menciona que se utilizó una cohorte longitudinal de pacientes con ERC y se recopilaron datos que incluían características iniciales de los pacientes y resultados de pruebas de sangre de rutina. Por otro, lado, se imputaron los datos faltantes utilizando la técnica de imputación múltiple. Menciona que se entrenaron y probaron cinco algoritmos de aprendizaje máquina, incluyendo regresión logística, naive Bayes, bosque aleatorio, árbol de decisión y k vecinos más cercanos, utilizando validación cruzada de cinco pliegues. Indica que tres modelos de aprendizaje automático, incluyendo regresión logística, naive Bayes y bosque aleatorio, mostraron una predictibilidad equivalente y una mayor sensibilidad en comparación con la ecuación de riesgo de insuficiencia renal (ERIR). Concluye que el estudio demostró la viabilidad del uso de aprendizaje automático para evaluar el pronóstico de ERC basado en características fácilmente accesibles. Se menciona que futuros estudios incluirán validación externa y mejoramiento de los modelos con variables predictoras adicionales.

(Sabanayagam et al., 2023) Se centra en la mejora de la predicción de la enfermedad renal diabética (ERD) incidente con técnicas de aprendizaje automático para identificar las características más relevantes en datos multidimensionales. El análisis compara la precisión de diferentes algoritmos de ML en la predicción de ERD. Los datos se obtuvieron de un estudio a largo plazo que involucró a 1365 participantes de nacionalidad china, malayo e indio, con edades entre 40 y 80 años, todos diagnosticados con diabetes pero sin ERD al inicio del estudio. Se consideraron 339 características, que incluyen datos personales de los participantes, imágenes de retina, información genética y perfiles de metabolitos sanguíneos como variables predictivas. El estudio comparó el rendimiento de varios modelos de ML con un modelo de regresión logística basado en características comúnmente asociadas con ERD, como la edad, el género, la etnia, la duración de la diabetes, la presión arterial sistólica, la concentración plasmática de HbA1c y el grado de

obesidad. Los resultados indican que el modelo Elastic Net obtuvo la mejor puntuación área bajo la curva, con un aumento del 7.0 % en comparación con el modelo de regresión logística. Las principales variables predictoras incluyeron la edad, la etnia, el uso de medicamentos antidiabéticos, la presión arterial alta, la retinopatía diabética, la presión arterial sistólica, el nivel de HbA1c, la tasa de filtración glomerular estimada y metabolitos relacionados con lípidos, lipoproteínas, ácidos grasos y cuerpos cetónicos. En resumen, este estudio demuestra que la implementación aprendizaje automático, junto con la selección de características relevantes, mejora la precisión en la predicción del riesgo de ERD en una población sin síntomas y estable, al mismo tiempo que identifica factores de riesgo novedosos, como metabolitos.

(Islam et al., 2023) Este estudio investigó técnicas de aprendizaje automático para el apoyo del diagnóstico de ERC y, mediante la modelización predictiva, identificó un conjunto de parámetros efectivos para la detección de la enfermedad. Doce clasificadores basados en aprendizaje automático se evaluaron en un entorno supervisado, con el XgBoost demostrando el mejor rendimiento, con una alta precisión del 0.983, precisión del 0.98, sensibilidad del 0.98 y puntuación F1 de 0.98. Este enfoque, que combina el aprendizaje automático y la modelización predictiva, ofrece una perspectiva interesante para desarrollar soluciones innovadoras en la detección precisa de enfermedades renales, lo que podría tener un impacto significativo en la atención médica y el tratamiento de ERC y más allá.

(Ravindra et al., 2023) Desarrollo un modelo de aprendizaje automático para la detección temprana de la enfermedad renal crónica (ERC). El estudio empleó registros clínicos con múltiples factores de riesgo, combinando técnicas de aprendizaje supervisado y un algoritmo de refuerzo para anticipar la progresión de la enfermedad. Los resultados mostraron que este enfoque superó a los métodos tradicionales —como el árbol aleatorio, KNN y máquinas de vectores de soporte—, al mejorar la precisión en la predicción y permitir una identificación más temprana de pacientes en riesgo. Los autores concluyen que la integración de modelos predictivos basados en inteligencia artificial puede tener un impacto significativo en la atención médica preventiva y en la reducción de complicaciones derivadas de la ERC.

(Swain et al., 2023) Comenta que el diagnóstico temprano de ERC es una preocupación importante en la atención médica, ya que los métodos convencionales no siempre son precisos debido a su dependencia de múltiples atributos biológicos. El aprendizaje automático se presenta como

una herramienta eficaz para mejorar la toma de decisiones clínicas al reducir la aleatoriedad. En este contexto, se desarrolla un modelo de aprendizaje automático que utiliza datos públicos para predecir ERC después de realizar un procesamiento de datos que incluye la imputación de datos faltantes, el equilibrio de datos y la selección de características. La máquina de vectores de soporte y el bosque aleatorio destacan como las técnicas más efectivas, logrando tasas bajas de falsos negativos y una precisión en las pruebas del 99.33 % y el 98.67 %, respectivamente. Estos resultados demuestran la utilidad del aprendizaje automático para la detección temprana de ERD y tienen el potencial de mejorar significativamente la atención médica y la toma de decisiones clínicas en este campo.

(Alsekait et al., 2023) Comenta que la ERC se da por la gradual disminución de la función del riñón a lo largo de un período extendido. Detectar ERC en sus etapas iniciales es esencial, ya que esto influye notablemente en la progresión de la salud del paciente, permitiendo intervenciones como tratamiento farmacológico en casos leves o procedimientos más invasivos como hemodiálisis o trasplante de riñón en situaciones severas. En la actualidad, los modelos de aprendizaje automático y aprendizaje profundo son fundamentales en el ámbito del diagnóstico médico debido a su alta precisión en la predicción de enfermedades. El rendimiento de estos modelos depende en gran medida de la elección de las características apropiadas y los algoritmos adecuados. En este contexto, el artículo se enfoca en presentar un enfoque innovador basado en DL para la detección de ERC. Se utilizaron diversos métodos de selección de características para elegir las más idóneas y se investigó cómo estas características impactan desde el punto de vista médico. El modelo propuesto combina modelos deep learning preentrenados con la máquina de soporte vectorial como modelo metaaprendizaje. Se llevaron a cabo exhaustivos experimentos con un conjunto de 400 pacientes, y los resultados indican que el modelo propuesto supera a otros en la predicción de ERC, especialmente cuando se seleccionan características utilizando el método mutualinfoclassif de la biblioteca de aprendizaje automático Scikit-learn.

(Farjana et al., 2023) Explica que en la era moderna, la conciencia sobre la salud es fundamental, pero debido a las cargas de trabajo y agendas agitadas, las personas a menudo solo prestan atención a su salud cuando aparecen síntomas específicos. Sin embargo, la enfermedad renal crónica (ERC) es una afección que a menudo no presenta síntomas o, en algunos casos, muestra síntomas muy leves, lo que dificulta su predicción, detección y prevención, lo que puede

resultar en daños a largo plazo para la salud. En este contexto, el aprendizaje automático (por sus siglas en inglés, ML) ofrece esperanza, ya que es eficaz en la predicción y el análisis. En este estudio, se proponen nueve enfoques de ML, como K-nearest neighbors (KNN), máquinas de soporte vectorial (SVM), regresión logística (RL), Naïve Bayes, clasificadores de árboles adicionales, AdaBoost, Xgboost y LightGBM. Estos modelos predictivos se construyeron utilizando un conjunto de datos sobre enfermedad renal crónica con 14 atributos y 400 registros para seleccionar el mejor clasificador para predecir la enfermedad renal crónica. El conjunto de datos se recopiló a través de la página web oficial Kaggle. Además, este estudio compara el rendimiento de estos modelos y destaca que el modelo LightGBM permitió predecir la enfermedad renal de manera más precisa que nunca, con un nivel de precisión del 99.00 %.

Estos estudios se centran en la aplicación de diversas técnicas de aprendizaje automático para anticipar la Enfermedad Renal Crónica Diabética (ERCD) o para evaluar el riesgo y progresión en pacientes con DMT2. A pesar de la propuesta prometedora de (Allen et al., 2022), la falta de detalles metodológicos restringe la posibilidad de replicación. (Dong et al., 2022) contribuyeron con información valiosa para la gestión a largo plazo, aunque se plantea la preocupación de la generalización a otras poblaciones. Sin embargo, Kamir realiza una evaluación exhaustiva de diversas técnicas de clasificación, proporcionando una visión completa, pero la aplicabilidad práctica se ve limitada por la falta de detalles clínicos. Por otro lado, (Bai et al., 2022) destacan la viabilidad del aprendizaje automático, aunque la ausencia de la validación externa plantea interrogantes sobre la generalización de los resultados. (Sabanayagam et al., 2023) ampliaron las consideraciones de características para el diagnóstico, pero se requiere más información sobre los modelos de regresión logística para obtener una comprensión completa. En el estudio de (Islam et al., 2023), se resaltó el alto rendimiento de XgBoost; sin embargo, la falta de detalles sobre los parámetros específicos para el diagnóstico limitó la comprensión completa de los resultados. (Ravindra et al., 2023) abordan la detección temprana de la ERC en la India, enfatizando que la precisión del modelo estuvo vinculada a la calidad de los datos clínicos disponibles. (Swain et al., 2023) subrayaron la efectividad de SVM y Random Forest, pero la dependencia de datos públicos limitó la aplicabilidad de los resultados. (Alsekait et al., 2023) propusieron un enfoque innovador basado en aprendizaje profundo, aunque la complejidad del modelo dificultó su implementación en entornos clínicos convencionales. Finalmente, (Farjana et al., 2023) evaluaron nueve enfo-

ques, identificando LightGBM como el más preciso, pero la falta de detalles sobre las variables predictoras para el diagnóstico limitó la comprensión total del rendimiento de los modelos.

2.3. Protección de datos personales

El manejo de datos clínicos requiere medidas estrictas de seguridad y privacidad, especialmente cuando se emplean en investigaciones de salud apoyadas en inteligencia artificial. Para garantizar el cumplimiento normativo y ético, es indispensable aplicar procesos de anonimización y desidentificación de la información. La anonimización es el proceso mediante el cual se eliminan de forma permanente los datos identificadores de los pacientes, de modo que no exista posibilidad de rastrear la información hacia la persona original. Este enfoque es fundamental para cumplir con la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP). Asimismo, responde a los lineamientos éticos establecidos en la Declaración de Helsinki, que orienta el uso responsable de datos médicos en investigación. Por otro lado, la desidentificación consiste en un proceso de pseudoanonimización, en el que se eliminan o sustituyen identificadores explícitos como nombre, CURP, número de seguridad social, teléfono o dirección, sin alterar la validez del registro clínico. A diferencia de la anonimización, la desidentificación puede ser reversible bajo controles estrictos, por lo que requiere protocolos de seguridad adicionales. En el ámbito médico, este procedimiento permite que expedientes electrónicos sean utilizados en proyectos de investigación sin comprometer la privacidad de los pacientes. Algoritmos automatizados de desidentificación resultan clave para procesar grandes volúmenes de expedientes, posibilitando análisis con aprendizaje automático sin exponer información sensible. En esta investigación, la base de datos DiabetIA utilizada se encuentra previamente anonimizada, lo cual asegura la confidencialidad de los pacientes y el cumplimiento de las disposiciones legales y éticas aplicables.

Capítulo 3

Aplicando atención a datos tabulares

3.1. Limitaciones de los modelos tradicionales y motivación del uso de Transformers

Los modelos tradicionales, como bosque aleatorio, XGBoost o incluso las redes neuronales profundas, han demostrado ser eficaces para diagnosticar enfermedades a partir de datos tabulares. Sin embargo, las propuestas basadas en las arquitecturas Transformers como TabTransformer, GatedTabTransformer y TabNet, ofrecen ventajas para problemas médicos complejos como la nefropatía diabética. Estas ventajas no solo apuntan a mejoras en precisión, sino también en la capacidad de capturar relaciones complejas, mejorar la interpretabilidad y adaptarse a diferentes tipos de datos (Badaro et al., 2023)

1. **Captura de interacciones complejas** Los modelos tradicionales capturan relaciones no lineales, pero suelen tratar las interacciones de forma limitada. Los Transformers, mediante su mecanismo de atención, pueden modelar relaciones complejas entre múltiples variables, algo esencial en medicina donde los efectos combinados (p. ej., glucosa, presión arterial y función renal) son determinantes para un diagnóstico (Badaro et al., 2023).

2. **Priorización dinámica de variables** Random Forest o XGBoost asignan una importancia fija a cada variable. Los Transformers, en cambio, ajustan dinámicamente la atención a las características según el contexto de cada paciente, ofreciendo diagnósticos más específicos en enfermedades multifactoriales como la DMT2 (Badaro et al., 2023).

3. **Explicabilidad mejorada** La explicabilidad es crítica en medicina. Modelos como TabNet ofrecen interpretaciones claras sobre qué variables influyen en la predicción, superando la opacidad de algoritmos como XGBoost y aumentando la confianza clínica (Badaro et al., 2023).

4. **Robustez y generalización** Los Transformers generalizan mejor en presencia de datos faltantes o distribuciones heterogéneas, una situación común en entornos médicos donde los registros clínicos suelen ser incompletos o variables (Badaro et al., 2023).

5. **Aprendizaje multimodal y transferencia** Los Transformers permiten integrar simultáneamente distintos tipos de datos (imágenes, texto clínico y registros tabulares). Además, pueden reutilizar conocimiento de otros dominios mediante transferencia de aprendizaje, algo poco factible en modelos clásicos (Badaro et al., 2023).

6. **Escalabilidad** En estudios clínicos con gran número de variables, los Transformers escalan de manera natural y seleccionan automáticamente las características más relevantes, reduciendo la necesidad de ajustes manuales extensivos (Badaro et al., 2023).

Aunque los modelos tradicionales son eficaces, los Transformers ofrecen mejoras sustanciales en interacción de variables, explicabilidad, robustez y escalabilidad, posicionándolos como una herramienta prometedora para el diagnóstico temprano de la nefropatía diabética (Badaro et al., 2023).

3.2. Aprendizaje profundo moderno para datos tabulares

Aunque el aprendizaje profundo ha logrado un éxito sobresaliente en imágenes y lenguaje natural, su aplicación en datos tabulares ha sido más desafiante (Ye y Wang, 2023). Se han explorado diversas adaptaciones:

- **Convoluciones (CNNs):** adaptadas para procesar vectores tabulares, pero con resultados inferiores a los modelos basados en árboles (Ye y Wang, 2023).
- **Redes recurrentes (RNNs y LSTMs):** útiles en series temporales, aunque poco efectivas en datos tabulares clínicos estáticos (Ye y Wang, 2023).

- **Modelos híbridos:** combinan técnicas clásicas como XGBoost con capas profundas para representar mejor variables categóricas y numéricas (Ye y Wang, 2023).

A pesar de estos intentos, los resultados muestran que CNNs y RNNs no superan de forma consistente a los modelos tradicionales. Estas limitaciones motivaron el diseño de arquitecturas específicas como los **Transformers tabulares**, que emplean mecanismos de atención capaces de aprender representaciones de tablas en un espacio vectorial continuo y de mejorar su desempeño en tareas complejas (Ye y Wang, 2023).

3.3. Aplicación de estructuras Transformers a datos tabulares

Los modelos Transformer han revolucionado el análisis de datos tabulares gracias a su capacidad de captar relaciones complejas y patrones en grandes volúmenes de datos estructurados. De hecho, se han consolidado como un grupo de arquitecturas capaces de manejar tanto atributos numéricos como categóricos, e incluso de integrar contexto adicional como texto asociado a tablas.

Los modelos *Transformer* para datos tabulares pueden clasificarse en distintas arquitecturas según su enfoque de atención y tratamiento de variables. La Figura 3.1 muestra una representación general de las arquitecturas más utilizadas en este tipo de datos: TabNet, TabTransformer y GatedTabTransformer.

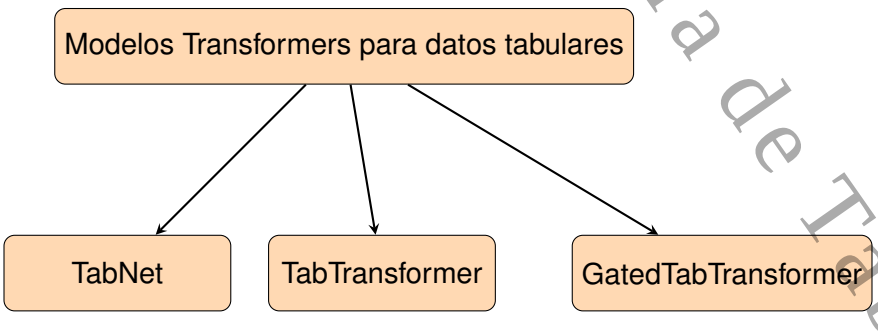


Figura 3.1. Principales arquitecturas *Transformer* aplicadas al análisis de datos tabulares.

3.3.1. TabNet

TabNet, propuesto por (Arik y Pfister, 2020) , es una arquitectura de aprendizaje profundo diseñada específicamente para datos tabulares. Su principal innovación consiste en el uso de un mecanismo de atención secuencial y enmascaramiento esparcido para seleccionar dinámicamente las características más relevantes en cada paso de decisión (Arik y Pfister, 2020). Esto permite que el modelo dedique su capacidad de aprendizaje solo a las variables más informativas, ofreciendo al mismo tiempo interpretabilidad local y global, algo que tradicionalmente ha sido una limitación en redes neuronales aplicadas a datos estructurados.

Aportaciones principales de la arquitectura:

- Selección de características instancia a instancia mediante máscaras aprendibles, lo que reduce la redundancia y aumenta la eficiencia.
- Bloques de decisión secuenciales que permiten procesar diferentes subconjuntos de características en cada paso y combinarlos de forma jerárquica.
- Atención controlada por regularización de esparcida, guiando al modelo a enfocarse en las variables más útiles y evitando el sobreajuste.
- Capacidad de preentrenamiento auto-supervisado (*aprendizaje semi supervisado*) para la reconstrucción de características enmascaradas, mejorando el rendimiento cuando hay datos no etiquetados.
- Interpretabilidad dual:
    - Local: identifica qué variables influyen en la predicción de una muestra específica.
    - Global: cuantifica la importancia agregada de cada variable en el modelo entrenado.

Ventajas:

- Supera a modelos tradicionales como XGBoost, LightGBM o CatBoost.
- Representaciones compactas y eficientes en número de parámetros.

- Capacidad de integrar variables numéricas y categóricas sin preprocesamiento complejo.
- Mejor generalización frente a datos redundantes o con variables irrelevantes.

Desventajas:

- Entrenamiento más lento debido a los pasos secuenciales de decisión.
- Sensible a la elección de hiperparámetros.

La Figura 3.2 muestra la arquitectura propuesta del modelo *TabNet*, la cual combina mecanismos de atención secuenciales con transformaciones de características. En (a) se representa el codificador (*encoder*), encargado de seleccionar dinámicamente las características relevantes mediante el bloque de atención. La parte (b) ilustra el decodificador (*decoder*), que reconstruye las representaciones latentes para el aprendizaje. En (c) se observa la estructura interna de las etapas de decisión, donde se aplican transformaciones compartidas y específicas de cada paso. Finalmente, (d) presenta el flujo completo del modelo, en el que las salidas del codificador se combinan para la predicción final.

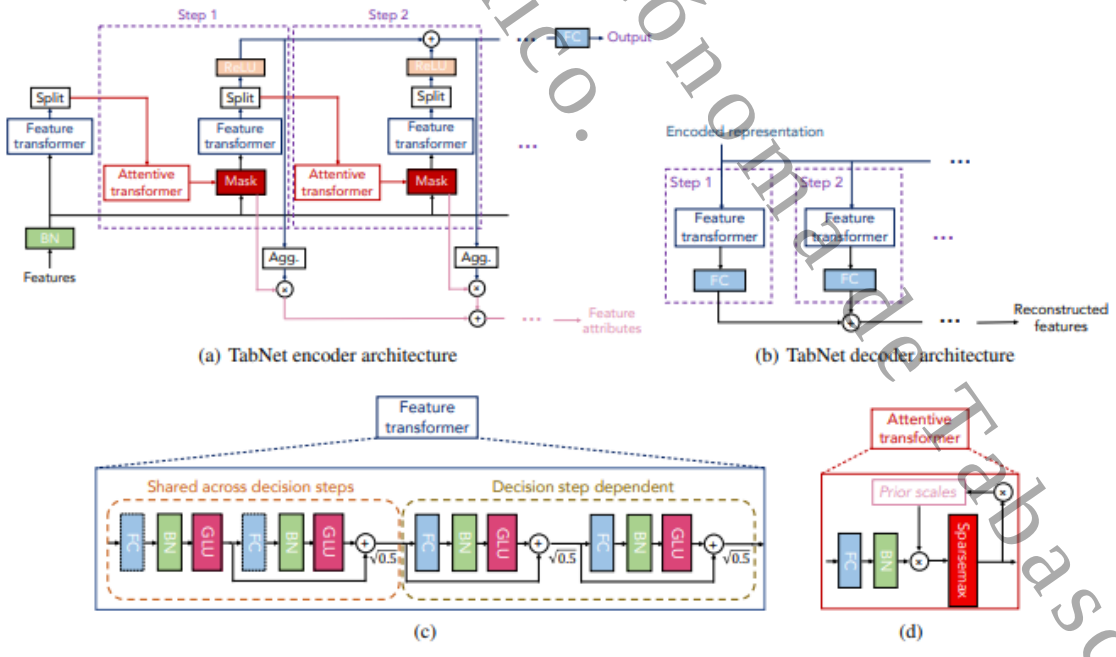


Figura 3.2. Arquitectura del modelo TabNet (Arik y Pfister, 2020).

3.3.2. TabTransformer

El *TabTransformer*, propuesto por Huang, et al (Huang et al., 2020), es un modelo basado en Transformers diseñado para el aprendizaje supervisado y semi-supervisado en datos tabulares. Su principal innovación es el uso de atención multi-cabeza para transformar las representaciones de características categóricas en embeddings contextuales, lo que permite capturar interacciones más ricas y robustas que las obtenidas con embeddings tradicionales (Huang et al., 2020).

Aportaciones principales de la arquitectura:

- Introduce una capa de embeddings de columna que representa de manera diferenciada cada característica categórica, incorporando un identificador único por columna.
- Emplea múltiples capas Transformer para transformar embeddings paramétricos en embeddings contextuales que capturan dependencias entre características.
- Integra las características numéricas directamente con los embeddings contextuales para formar una representación combinada antes de la predicción.
- Desarrolla un procedimiento de preentrenamiento auto-supervisado basado en *Modelado de Lenguaje Enmascarado* y *Detección de Tokens Reemplazados*, que mejora el rendimiento en escenarios con datos no etiquetados.
- Aporta interpretabilidad: los embeddings contextuales permiten visualizar la proximidad semántica entre categorías y entender mejor las relaciones entre variables.

Ventajas:

- Supera a redes neuronales profundas tradicionales (MLP) en múltiples benchmarks de clasificación binaria, con una ganancia promedio de 1 % en AUC.
- Robusto ante ruido y valores faltantes en variables categóricas gracias a la contextualización de embeddings.
- Permite aprendizaje semi-supervisado mediante preentrenamiento en grandes volúmenes de datos sin etiqueta.

- Mejora la interpretabilidad frente a MLP al generar agrupaciones semánticamente coherentes en el espacio de embeddings.

Desventajas:

- Mayor complejidad computacional que los modelos basados en árboles, especialmente en escenarios con muchas categorías.
- Alto costo de entrenamiento debido a la pila de capas Transformer.
- Sensible a la cantidad de datos disponibles: requiere un volumen considerable de datos para superar de forma consistente a modelos tradicionales.

La Figura 3.3 muestra la arquitectura del modelo *TabTransformer*, propuesta por Huang et al (Huang et al., 2020). El modelo combina un bloque de normalización de capas y una capa de *embedding* para representar tanto variables categóricas como continuas. Posteriormente, las características categóricas son procesadas por múltiples capas *Transformer* con atención multicabeza, que aprenden dependencias entre variables. Las salidas de las capas de atención se concatenan y se introducen en una red neuronal *feed-forward* (MLP) que genera la predicción final. Esta arquitectura mejora la representación contextual de las variables tabulares y ofrece un mejor desempeño en tareas supervisadas y semi-supervisadas.

3.3.3. SAINT

El modelo *Self-Attention and Intersample Attention Transformer (SAINT)*, propuesto por Somepalli et al (Somepalli et al., 2021), introduce una arquitectura híbrida diseñada para superar las limitaciones de los modelos clásicos y de otros Transformers aplicados a datos tabulares. SAINT combina auto-atención a nivel de características con un novedoso mecanismo de atención entre muestras, lo que permite a cada fila de la tabla interactuar con otras filas en el mismo lote de datos (Somepalli et al., 2021).

Aportaciones principales de la arquitectura:

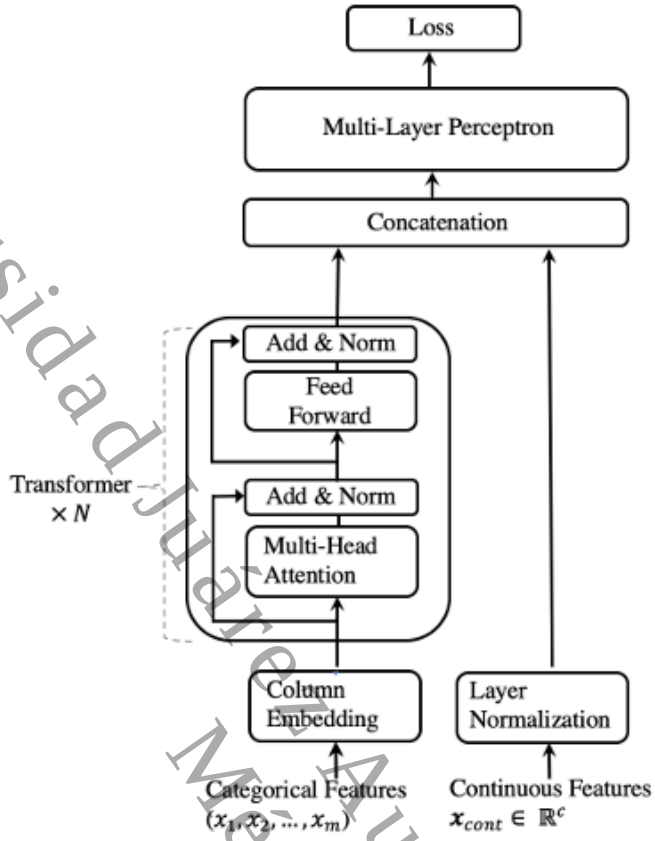


Figura 3.3. Arquitectura del modelo TabTransformer (Huang et al., 2020).

- Proyecta tanto variables categóricas como continuas en un espacio d-dimensional unificado antes de pasarlas por el encoder Transformer.
- Introduce un bloque de *atención entre muestras* que permite a cada muestra mejorar su representación utilizando información de otras filas de la tabla, similar a un esquema de vecinos más cercanos aprendido de manera fin a fin.
- Combina auto-atención y atención entre muestras en cada etapa de la arquitectura, capturando tanto dependencias internas de una fila como patrones comunes entre filas.
- Incorpora un procedimiento de preentrenamiento auto-supervisado basado en aprendizaje contrastivo y técnicas de aumento de datos tabulares (CutMix y Mixup), lo que mejora el rendimiento en escenarios semi-supervisados.
- Ofrece interpretabilidad mediante la visualización de los mapas de atención, permitiendo

analizar qué variables y qué instancias influyen más en la predicción.

Ventajas:

- En promedio, SAINT supera a modelos clásicos de *gradient boosting* (XGBoost, LightGBM y CatBoost) y a arquitecturas profundas como TabNet y TabTransformer en tareas supervisadas y semi-supervisadas.
- Robusto frente a ruido y valores faltantes, gracias al mecanismo de intersample attention que permite "tomar prestada" información de filas similares.
- Mejora notable en escenarios con pocos datos etiquetados mediante preentrenamiento contrastivo.
- Adaptable a datos con alto número de características, mostrando mejor desempeño en dominios como biomedicina y finanzas.

Desventajas:

- El entrenamiento es más costoso en cómputo que en modelos basados en árboles y que en arquitecturas más simples como MLPs.
- Sensibilidad al tamaño del lote: el rendimiento del bloque de intersample attention depende del número de instancias presentes en el batch.
- Aunque ofrece interpretabilidad, su complejidad hace que la explicación sea menos directa que en TabNet.
- El preentrenamiento contrastivo aporta beneficios principalmente en entornos semi-supervisados; en escenarios totalmente supervisados la mejora es menor.

La Figura 3.4 muestra la arquitectura del modelo *Self-Attention and Intersample Attention Transformer* (SAINT). En la parte izquierda se observa la fase de preentrenamiento auto-supervisado, donde se utilizan técnicas de enmascaramiento y aprendizaje contrastivo para mejorar la representación de los datos. La parte derecha ilustra la etapa de ajuste fino (*fine-tuning*), en la que las características preentrenadas se optimizan para tareas supervisadas específicas. Este diseño

híbrido permite que SAINT combine la atención entre características e instancias, capturando tanto relaciones dentro de cada fila como dependencias entre muestras del conjunto de datos.

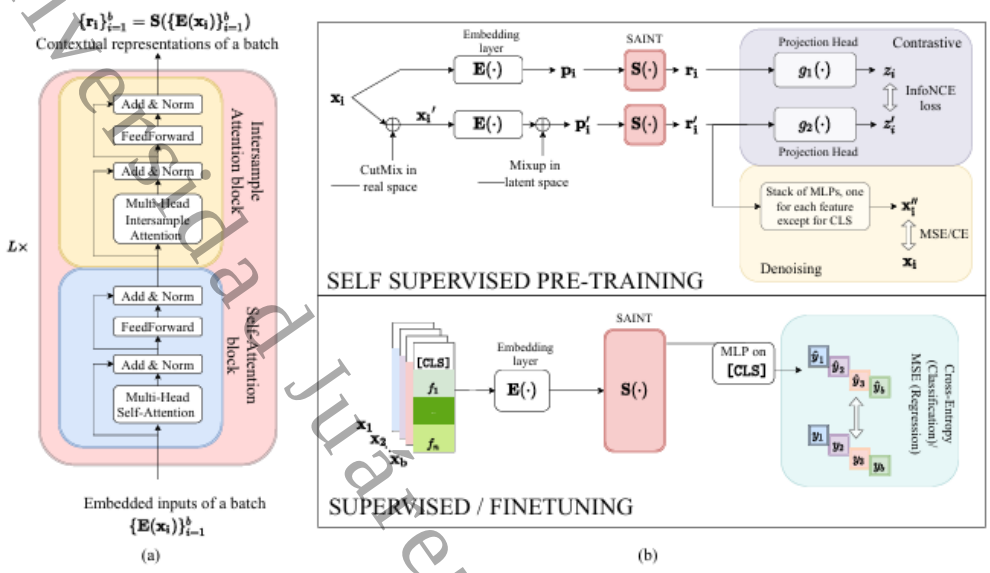


Figura 3.4. Arquitectura del modelo SAINT (Somepalli et al., 2021).

3.3.4. ARM-Net

El *Adaptive Relation Modeling Network (ARM-Net)*, propuesto por Cai et al (Cai et al., 2021), está diseñado específicamente para modelar datos estructurados, como los provenientes de bases de datos relacionales. Su principal innovación es la incorporación de neuronas exponenciales y un mecanismo de atención multi-cabeza con compuertas que permiten modelar interacciones de características de manera selectiva y dinámica (Cai et al., 2021).

Aportaciones principales de la arquitectura:

- Modela explícitamente interacciones entre características (*cross features*) en el espacio exponencial, superando las limitaciones de modelos previos como AFN que operaban en el espacio logarítmico.
- Introduce neuronas exponenciales que pueden capturar interacciones de orden arbitrario entre variables de entrada.

- Utiliza atención multi-cabeza con compuertas para asignar pesos dinámicos a las interacciones, filtrando características irrelevantes y enfocándose en las más informativas.
- Ofrece interpretabilidad dual:
  - Global: importancia de características agregada a nivel de todo el modelo.
  - Local: identifica las interacciones específicas que afectan a cada predicción.
- Se integra dentro de un marco ligero llamado ARMOR, que facilita el análisis de datos relacionales con interpretabilidad y eficiencia.

Ventajas:

- Ofrece interpretabilidad transparente, al especificar explícitamente qué combinaciones de variables influyen en cada predicción.
- Escalable y eficiente: su complejidad es lineal respecto al número de atributos, y logra aceleraciones significativas en GPU.
- Robusto en aplicaciones críticas, como predicción de tasas de reingreso hospitalario o sistemas de recomendación, donde la explicabilidad es esencial.

Desventajas:

- Mayor complejidad computacional que métodos clásicos basados en árboles o factorización, especialmente en grandes volúmenes de datos.
- Requiere una cuidadosa selección de hiperparámetros como el número de cabezas de atención, neuronas exponenciales y el nivel de sparsidad (α).
- Interpretabilidad más técnica que en TabNet: aunque ARM-Net ofrece explicaciones explícitas, estas se basan en pesos de atención y términos de cruce, lo que puede resultar menos intuitivo para usuarios finales.
- En escenarios con pocos datos, la ganancia sobre modelos más simples puede no justificar el costo computacional adicional.

La Figura 3.5 ilustra la arquitectura del modelo. El modelo está compuesto por un bloque de transformación exponencial que proyecta las características de entrada en un espacio relacional, seguido por un mecanismo de atención multi-cabeza con compuertas que ajusta dinámicamente los pesos de interacción entre variables. Finalmente, el módulo de alineación bilineal permite combinar las representaciones aprendidas para generar términos de interacción más interpretables y eficientes en tareas tabulares complejas.

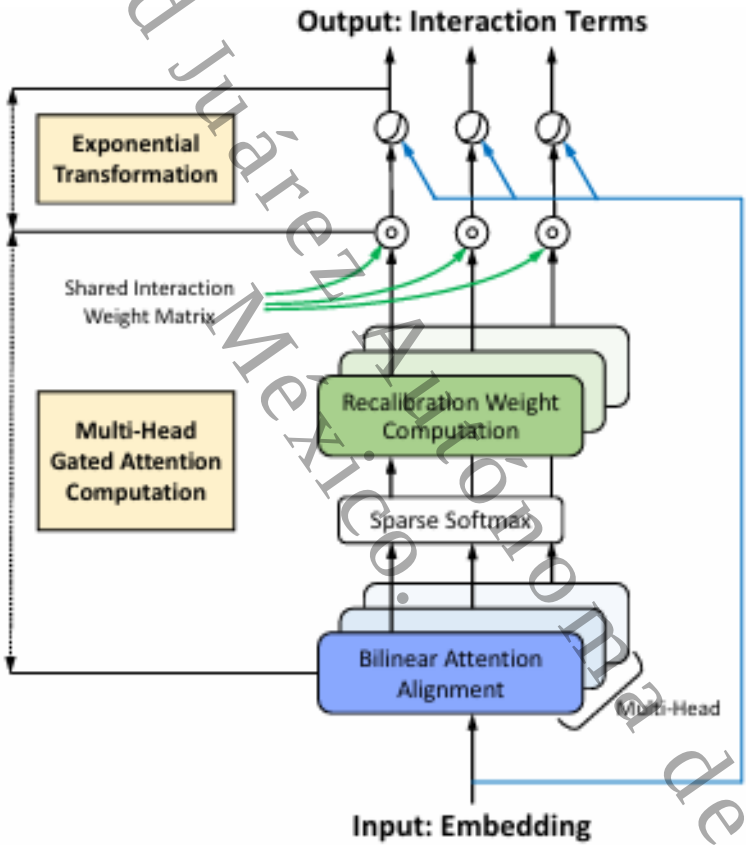


Figura 3.5. Arquitectura del modelo ARM-Net (Cai et al., 2021).

3.3.5. GatedTabTransformer

El *GatedTabTransformer*, propuesto por Cholakov et al (Cholakov y Kolev, 2022), es una extensión del TabTransformer que incorpora un bloque gated multilayer perceptron (gMLP) al final de la arquitectura. Mientras que TabTransformer transforma variables categóricas en **embeddings**

contextuales mediante autoatención multi-cabeza, GatedTabTransformer reemplaza la capa MLP estándar por un **gMLP con unidades de compuerta espaciales**, lo que mejora la capacidad del modelo para capturar interacciones complejas entre características (Cholakov y Kolev, 2022).

Aportaciones principales de la arquitectura:

- Introduce el bloque gMLP en lugar del MLP final del TabTransformer, añadiendo proyecciones lineales y una unidad de compuerta espacial que capturan interacciones entre tokens de entrada.
- Mejora la capacidad de generalización en tareas de clasificación binaria, con incrementos de entre 0.5 % y 1.1 % en AUROC sobre el TabTransformer original.
- Permite ajustar funciones de activación (ReLU, GELU, SELU, LeakyReLU) y estructuras de capas mediante optimización de hiperparámetros, adaptando la arquitectura a distintos conjuntos de datos.
- Abre la posibilidad de incorporar técnicas modernas de *AutoML* y *Neural Architecture Search* para explorar variantes más avanzadas del modelo.

Ventajas:

- Supera al TabTransformer, TabNet y MLPs en tareas de clasificación binaria.
- Beneficios claros en AUROC, con mejoras reportadas de hasta 3.1 % sobre TabNet y 1 % sobre TabTransformer.
- Flexibilidad para experimentar con diferentes activaciones y profundidades de red, lo que ofrece mayor adaptabilidad.
- Conserva la capacidad del TabTransformer para manejar datos categóricos y numéricos de manera conjunta.

Desventajas:

- Mayor complejidad computacional que TabTransformer debido a la incorporación del gMLP.

- Requiere procesos de optimización de hiperparámetros más exhaustivos (tuning de profundidad, dimensiones ocultas y tasas de aprendizaje).
- En datasets pequeños, el incremento de rendimiento puede ser marginal frente al costo adicional de cómputo.
- A pesar de su mejora en interpretabilidad frente a MLPs, sigue siendo menos intuitivo que TabNet en términos de explicación directa de características.

La Figura 3.6 presenta la arquitectura del modelo *GatedTabTransformer*. Este modelo extiende el *TabTransformer* al incorporar un bloque *gated multilayer perceptron* (gMLP) al final de la arquitectura, lo que permite combinar la atención multi-cabeza con mecanismos de compuerta espacial (*spatial gating units*). El bloque superior de la figura muestra la estructura interna del gMLP, donde las proyecciones lineales y las compuertas ajustan dinámicamente las interacciones entre canales. Las capas inferiores ilustran cómo estas representaciones se integran con los *embeddings* categóricos y las variables continuas, logrando una mayor capacidad de generalización en tareas de clasificación tabular.

Análisis objetivo de las arquitecturas mencionadas

El análisis comparativo de las cinco arquitecturas basadas en Transformers aplicadas a datos tabulares evidencia que cada una aporta innovaciones específicas para superar las limitaciones de los modelos tradicionales y de las redes neuronales profundas convencionales.

En conjunto, estas arquitecturas representan la evolución de los Transformers en el ámbito tabular: desde la búsqueda de interpretabilidad y eficiencia (TabNet), pasando por la contextualización de variables categóricas (TabTransformer), la integración de relaciones inter-muestra (SAINT), el modelado explícito de interacciones complejas (ARM-Net), hasta la hibridación con perceptrones avanzados (GatedTabTransformer).

Su análisis demuestra que no existe un modelo universalmente superior, sino que la elección depende del equilibrio entre precisión, interpretabilidad y recursos computacionales disponibles. En el contexto de aplicaciones médicas, estas arquitecturas constituyen herramientas prometedoras para el desarrollo de sistemas de diagnóstico más precisos, explicables y adaptativos.

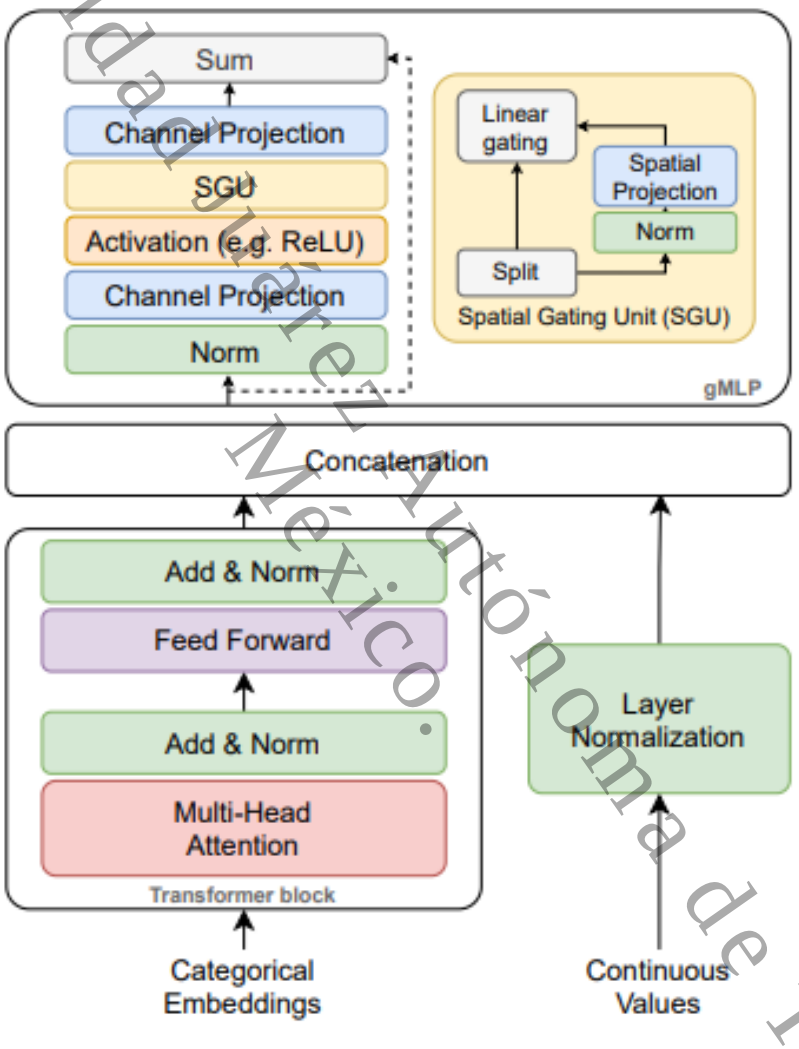


Figura 3.6. Arquitectura del modelo GatedTabTransformer (Cholakov y Kolev, 2022).

Capítulo 4

Descripción y evaluación del modelo de clasificación transformer

4.1. Introducción

En este capítulo se describe el diseño, implementación y evaluación de un modelo basado en la arquitectura *Transformer* aplicado a datos tabulares. El objetivo principal es apoyar el diagnóstico de la Nefropatía Diabética (ND) en pacientes con Diabetes Mellitus Tipo 2 (DMT2).

La metodología seguida corresponde al proceso CRISP-DM, que abarca desde la comprensión del problema hasta la evaluación del modelo. El modelo desarrollado se centra en la clasificación binaria (paciente con ND o sin ND), integrando variables clínicas tanto continuas como categóricas.

4.2. Comprensión del Problema

El objetivo del proyecto es aplicar mecanismos de atención en redes neuronales para la clasificación binaria de la nefropatía diabética.

Este modelo busca:

- Detectar de manera temprana pacientes con riesgo de ND.
- Minimizar falsos negativos, dado el impacto clínico.

- Evaluar la viabilidad de *Transformers* en datos médicos tabulares.

La base de datos utilizada fue DiabetIA, un repositorio abierto con registros clínicos anonimizados de pacientes mexicanos. El conjunto incluye variables de tipo demográfico, bioquímico y clínico.

4.3. Comprensión de los Datos

Visión general del conjunto de datos

La base de datos original estuvo conformada por 477,000 registros y más de 500 variables clínicas y demográficas. Tras un proceso de depuración y estandarización, el conjunto se redujo a 22,246 observaciones y 237 variables, lo que permitió trabajar con datos de mayor calidad y consistencia para los análisis posteriores.

Variable objetivo (ND) y distribución inicial

La variable objetivo, referida como ND, se codificó de la siguiente forma: ND $= 0$ indica paciente *sin diagnóstico de ND* y ND $= 1$ indica *con diagnóstico de ND*. En el conjunto depurado se observó la siguiente distribución por clases:

- ND $= 0$ (pacientes sin ND): 18,689 casos.
- ND $= 1$ (pacientes con ND): 3,500 casos.

Criterios de calidad y selección de observaciones

Para garantizar comparabilidad entre modelos y reducir el impacto de valores faltantes, se priorizó la completitud de las observaciones:

1. **Depuración inicial** eliminación de duplicados, estandarización de formatos y descarte de variables con baja variabilidad o alta proporción de faltantes.
2. **Conjunto altamente completo**: se retuvo un subconjunto de 5,000 observaciones con cobertura completa en las 237 variables (criterio de máxima completitud) para análisis exploratorios y pruebas de consistencia.

3. **Muestra de trabajo con balanceo aproximado**: para el modelado principal se construyó una muestra *aproximadamente balanceada* de 8,500 pacientes, compuesta por 5,000 casos ND $= 0$ y 3,500 casos ND $= 1$.

Estructura final de variables

El conjunto final empleado para el modelado contiene **237 variables**, de las cuales **25** son **numéricas** y **212** son **categóricas**.

- Matriz completa: $(8{,}500 \times 237)$.
- Submatriz numérica: $(8{,}500 \times 25)$.
- Submatriz categórica: $(8{,}500 \times 212)$.

4.3.1. Resumen cuantitativo

Tabla 4.1. Evolución del conjunto de datos a lo largo del proceso de depuración y selección

Etapa	Registros (n)	Variables (p)
Conjunto original	477,000	> 500
Depurado (calidad/estandarización)	22,246	237
Con etiqueta ND disponible	22,189	237
Subconjunto altamente completo	5,000	237
Muestra de trabajo (aprox. balanceada)	8,500	237

La Tabla 4.1 resume el proceso de depuración y selección del conjunto de datos, mostrando la evolución en el número de registros y variables a lo largo de las distintas etapas. Se observa que el conjunto original contenía más de 477 000 registros, los cuales fueron reducidos tras aplicar controles de calidad, estandarización y filtrado de etiquetas válidas. Finalmente, la muestra de trabajo balanceada quedó conformada por 8 500 registros y 237 variables, lo que permitió mantener la representatividad y estabilidad estadística en el modelado.

Tabla 4.2. Distribución por clase (ND) en las principales etapas

Etapa	ND $= 0$	ND $= 1$	Total
Depurado (con etiqueta ND)	18,689	3,500	22,189
Muestra de trabajo	5,000	3,500	8,500

La Tabla 4.2 presenta la distribución de instancias por clase, correspondiente al diagnóstico de nefropatía diabética (ND). Se aprecia un ligero desbalance entre las clases, con una mayor proporción de casos negativos (ND = 0) en comparación con positivos (ND = 1). Este equilibrio moderado (59 % frente a 41 %) fue considerado adecuado para el entrenamiento de modelos sin necesidad de aplicar estrategias de sobremuestreo o ponderación de clases.

4.3.2. Tipos de variables en el conjunto final

Tabla 4.3. Desglose de tipos de variables en el conjunto final

Tipo	n. de variables	Dimensión de la submatriz
Numéricas	25	$8{,}500 \times 25$
Categóricas	212	$8{,}500 \times 212$
Total	**237**	$8{,}500 \times 237$

La Tabla 4.3 detalla la composición del conjunto de datos final, desglosando el número de variables por tipo y la dimensión correspondiente de cada sub-matriz. Esta estructura híbrida combina mediciones clínicas continuas (por ejemplo, niveles de glucosa o edad) con variables discretas y cualitativas (como sexo o presencia de hipertensión), lo cual resulta idóneo para enfoques de aprendizaje profundo sobre datos tabulares.

Tabla 4.4. Ejemplo de variables relevantes en el dataset.

Variable	Tipo	Descripción
`Glucosa`	Continua	Niveles de glucosa en sangre (mg/dL)
`HbA1c`	Continua	Hemoglobina glicosilada (%)
`Creatinina` sérica	Continua	mg/dL
`Microalbuminuria`	Continua	mg/24h en orina
`Sexo`	Categórica	Masculino / Femenino
`Edad`	Continua	Años
`Hipertensión`	Categórica	Sí / No

La Tabla 4.4 muestra ejemplos de variables representativas incluidas en el conjunto de datos, junto con su tipo y una breve descripción. Se incluyen parámetros clínicos clave para la detección de nefropatía diabética, tales como glucosa, hemoglobina glicosilada (HbA1c), creatinina sérica y microalbuminuria, además de factores demográficos y de riesgo como edad, sexo e hipertensión. Estas variables se seleccionaron por su relevancia médica y su capacidad predictiva en el contexto de la enfermedad renal asociada a diabetes mellitus tipo 2.

4.4. Preparación de los datos

En esta etapa, se realiza el preprocesamiento y la limpieza de los datos. Se eliminan los registros duplicados, los valores faltantes y los datos inconsistentes. Además, se realiza la selección de variables, la transformación de datos. Esta etapa es crucial para garantizar que los datos sean adecuados para el análisis.

Preprocesamiento de los datos

Extraer Observaciones de pacientes negativos potenciales.

En DiabetIA se encuentra información de varias comorbilidades de la DMT2: nefropatía diabética (ND), retinopatía, neuropatía, pie diabético; además de la edad del paciente, su diagnostico, sexo, IMC e hipertensión arterial sistémica (HAS) registrado durante 3 años. Se encontró que hay columnas en su totalidad 0 en todos los registros de los pacientes con ND y también hay columnas relevantes para el diagnóstico, pero su información no representa información relevante, así que se procede a su eliminación total. De igual modo se hizo la extracción de columnas en totalmente en ceros, erróneas y posteriormente su eliminación.

Visualización y Análisis de las distribuciones de las variables numéricas y categóricas

La Figura 4.1 muestra las distribuciones de las variables numéricas incluidas en el conjunto de datos. Cada histograma representa la densidad de probabilidad de una variable cuantitativa, normalizada para facilitar la comparación entre escalas diferentes. Se observa que la mayoría de las variables presentan distribuciones asimétricas y concentraciones hacia valores bajos, lo cual es característico de datos clínicos reales. Estas observaciones justificaron la aplicación posterior de técnicas de normalización y escalado, con el fin de estabilizar la varianza y mejorar el desempeño de los modelos de aprendizaje profundo.

La Figura 4.2 ilustra la distribución de frecuencias de algunas de las variables categóricas más representativas del conjunto de datos. Se incluyen atributos demográficos y clínicos discretos, como el sexo, la presencia de hipertensión y el grupo etario. Se aprecia una ligera predominancia de ciertas categorías, lo que refleja la composición poblacional de la muestra analizada. Esta infor-

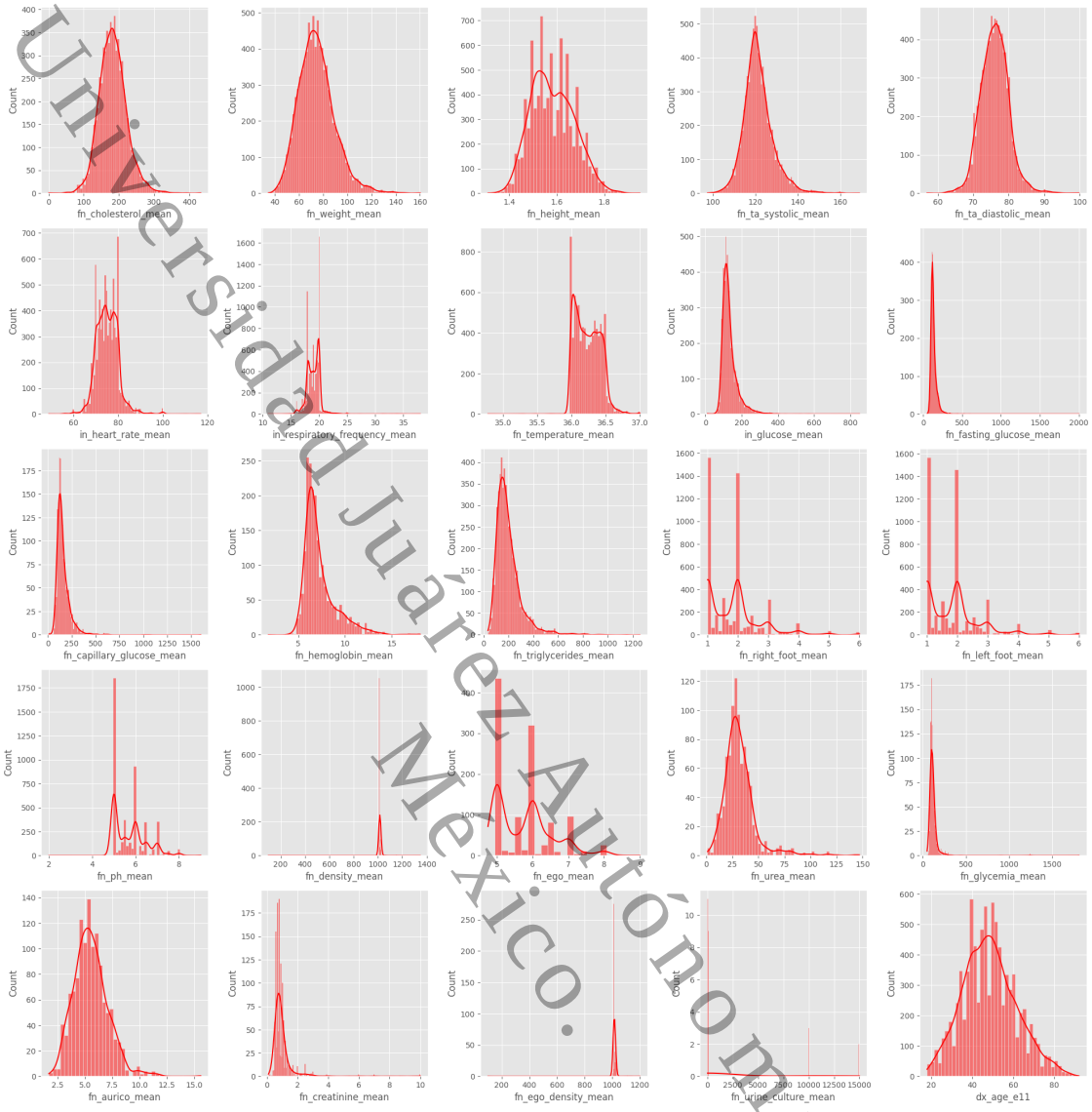


Figura 4.1. Distribución de columnas numéricas

mación fue utilizada para detectar posibles sesgos y definir estrategias de codificación mediante *one-hot encoding* o embeddings aprendibles antes del modelado con arquitecturas Transformer.

Análisis Exploratorio de Datos (EDA)

Eliminar Valores Nulos o *NaN*

De acuerdo con la literatura revisada, se recomienda eliminar aquellas columnas que presenten más del 75 % de valores nulos dentro del *dataset*. Sin embargo, al aplicar este criterio se observó que también se eliminarían algunas variables clínicamente relevantes, por lo que se es-

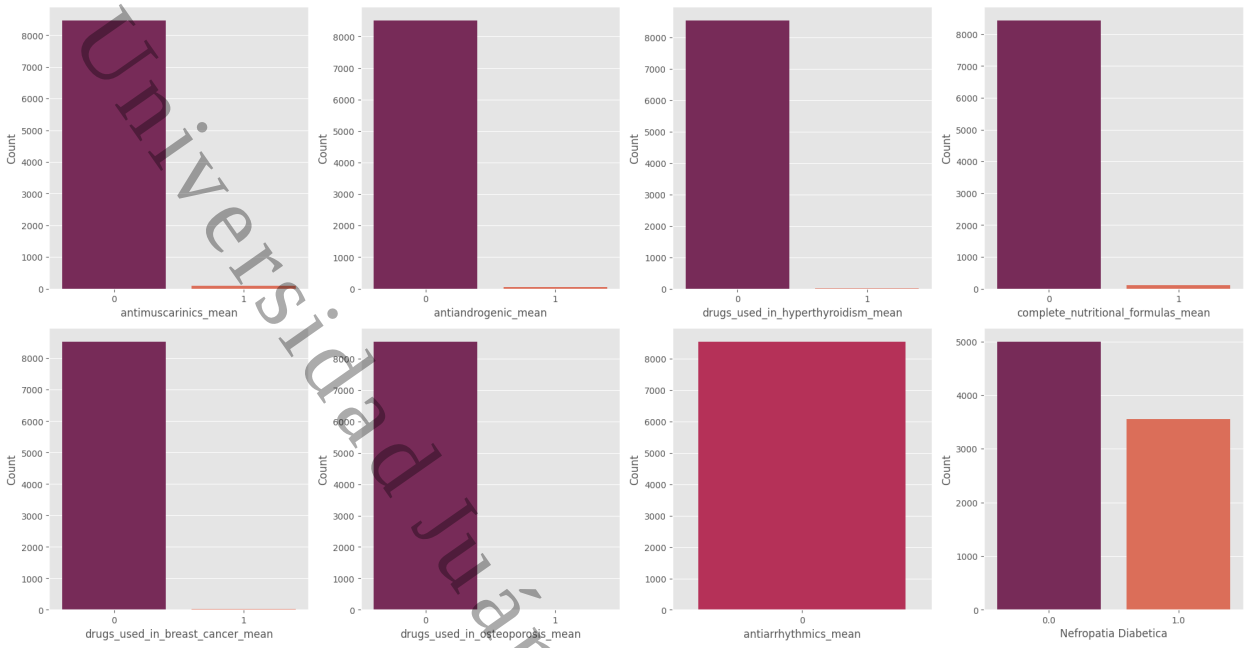


Figura 4.2. Distribución de algunas columnas categóricas

tableció un umbral del 0.75 % de datos faltantes como punto de equilibrio, permitiendo conservar información útil sin comprometer la calidad del conjunto de datos.

Para este propósito, se desarrolló una función que identifica y elimina automáticamente las columnas con un porcentaje de valores nulos superior al umbral especificado (por defecto, 75 %). El algoritmo calcula el porcentaje de valores nulos por columna, selecciona aquellas que superan el límite y las excluye del *DataFrame*. Como resultado, se eliminaron 189 columnas categóricas con más del 30 % de valores faltantes.

La revisión de las columnas eliminadas reveló que muchas correspondían a comorbilidades, trastornos orgánicos generales o uso de fármacos. Aunque algunas podrían tener relevancia clínica en el contexto de la nefropatía diabética, la evidencia bibliográfica indica que existen factores predictivos más consistentes, por lo que su exclusión no afecta negativamente el análisis.

Después del proceso de depuración, el conjunto de datos final quedó conformado por un total de **70 columnas**, de las cuales **25 son numéricas** y **45 son categóricas**.

Correlación de Datos

Calcular la matriz de correlación columnas numéricas

En el caso de las variables numéricas, calcular la matriz de correlación resulta fundamental, ya que permite identificar la fuerza y dirección de la relación lineal entre cada par de variables. Valores cercanos a +1 o −1 indican asociaciones fuertes (positivas o negativas, respectivamente), mientras que valores próximos a 0 sugieren ausencia de relación lineal. Este análisis facilita comprender la estructura interna del conjunto de datos, así como detectar patrones relevantes que podrían influir en la variable objetivo o en el comportamiento general de los pacientes estudiados.

Además, la matriz de correlación es útil para detectar multicolinealidad, es decir, la presencia de variables altamente correlacionadas que aportan información redundante y pueden afectar el rendimiento de los modelos de aprendizaje automático. Al identificarlas, es posible decidir si conviene eliminarlas, transformarlas o reducir su impacto mediante técnicas avanzadas como la reducción de dimensionalidad (PCA). De esta forma, se garantiza una mejor selección de variables, se previene el sesgo en el modelado y se construyen representaciones más robustas para el entrenamiento de los algoritmos.

La Figura 4.3 presenta el mapa de calor de correlaciones entre las variables numéricas del conjunto de datos. Cada celda muestra el coeficiente de correlación de Pearson, el cual cuantifica la relación lineal entre pares de variables. Los valores cercanos a $+1$ o -1 indican una asociación fuerte positiva o negativa, respectivamente, mientras que valores próximos a 0 reflejan ausencia de relación lineal. Se observa que algunas variables clínicas, como los niveles de glucosa, colesterol y presión arterial, presentan correlaciones moderadas, lo que sugiere posibles interdependencias fisiológicas entre los indicadores metabólicos. Este análisis permitió detectar redundancias, reducir la multicolinealidad y seleccionar las características más informativas para el modelado predictivo.

Calcular la matriz de correlación en columnas categóricas

Para analizar la relación entre variables categóricas se utilizó la medida de asociación conocida como **Cramér's V**. Esta técnica se basa en la prueba de independencia χ^2, donde a partir de una

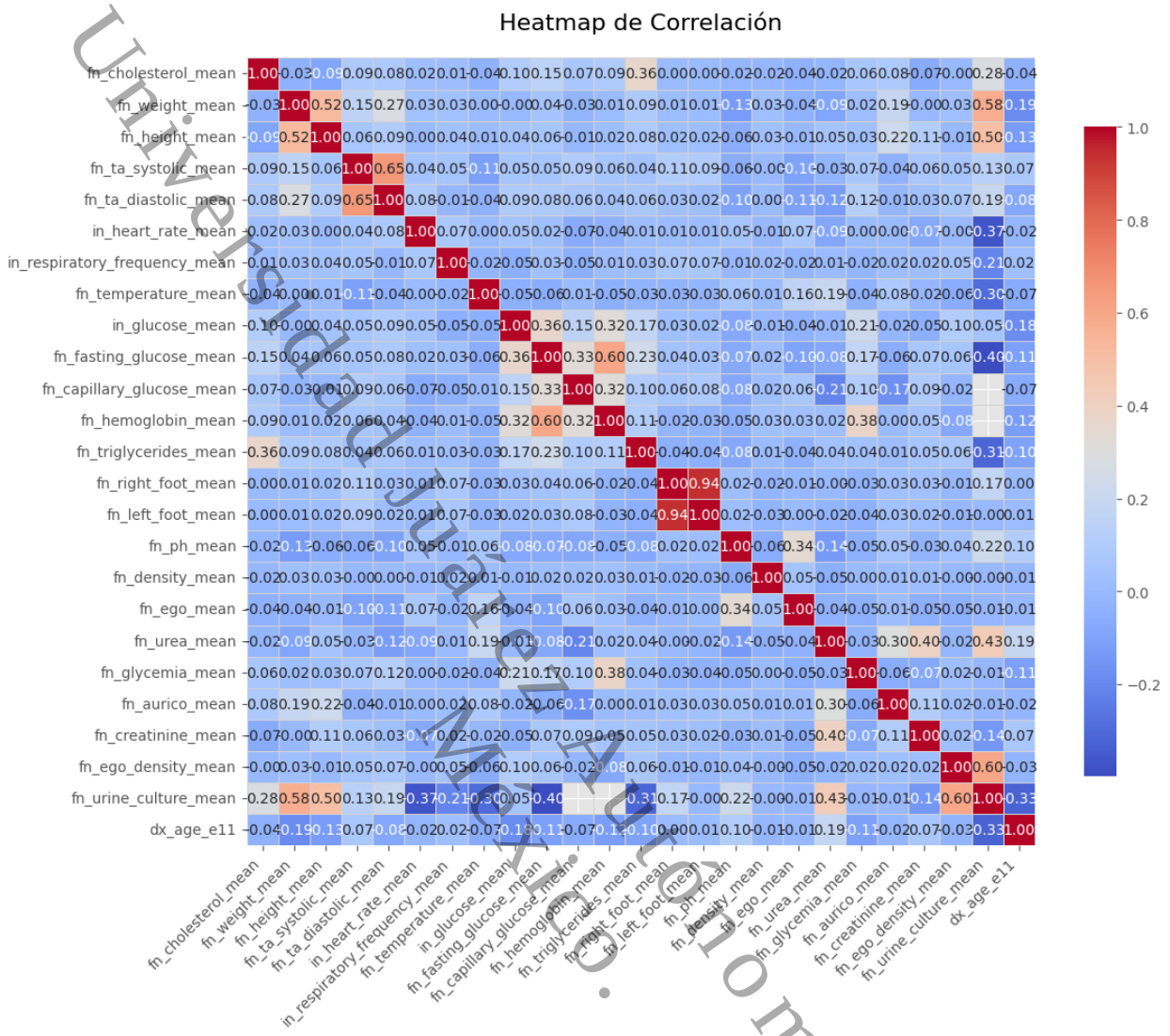


Figura 4.3. correlación columnas numéricas

tabla de contingencia se contrasta la hipótesis nula de que las dos variables son independientes. El estadístico χ^2 se normaliza considerando el tamaño de la muestra y la dimensión mínima de la tabla, de manera que el coeficiente resultante se calcula con la siguiente expresión:

$$V = \sqrt{\frac{\chi^2}{n \cdot (k-1)}}$$

donde n representa el número total de observaciones y k corresponde al número de categorías de la variable con menor cardinalidad. El coeficiente V toma valores entre 0 y 1: valores cercanos a 0 indican ausencia de asociación, mientras que valores próximos a 1 reflejan una

fuerte dependencia entre las categorías.

La matriz de correlación de Cramér's V se construye calculando dicho coeficiente para cada par de variables categóricas del conjunto de datos. De este modo, es posible identificar patrones de dependencia entre atributos cualitativos, detectar redundancias y guiar la selección de variables más representativas para el modelado. Al igual que en el caso de las variables numéricas, este análisis permite reducir la complejidad del conjunto de datos y prevenir sesgos derivados de información redundante o altamente asociada.

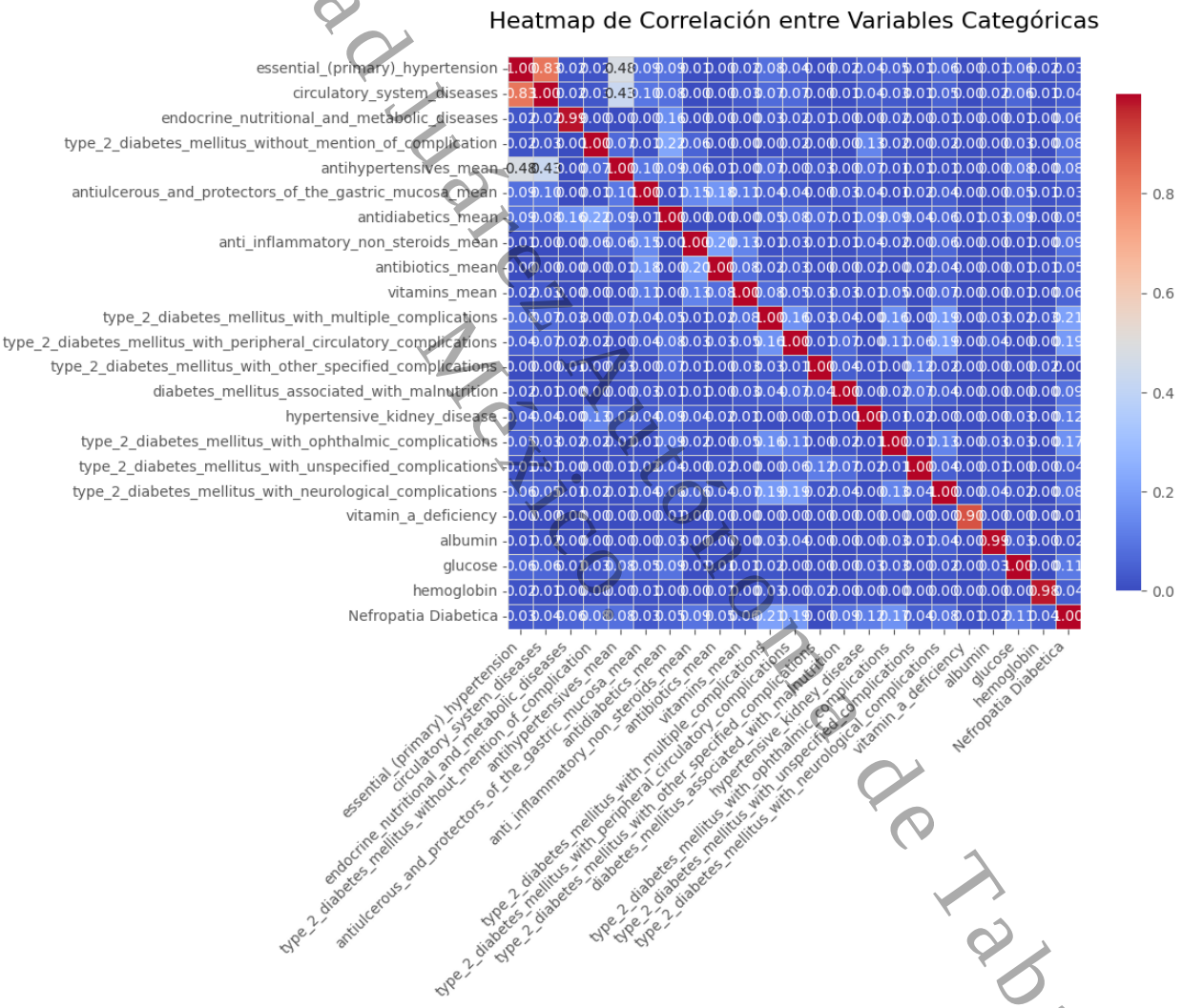


Figura 4.4. correlación columnas categóricas

La Figura 4.4 muestra la matriz de correlaciones entre variables categóricas, calculada mediante el coeficiente de asociación de Cramér's V. Este estadístico mide la intensidad de la relación entre dos variables cualitativas, con valores que oscilan entre 0 (sin asociación) y 1 (asociación

perfecta). El análisis permitió identificar relaciones relevantes entre variables diagnósticas y factores de riesgo, como la coexistencia de hipertensión y diabetes, así como la influencia del sexo y la edad sobre las comorbilidades observadas. Estos resultados orientaron la depuración de variables redundantes y contribuyeron a una representación más compacta del conjunto de datos para el entrenamiento de los modelos basados en *Transformers*.

Imputación de datos para variables continuas

En el conjunto de datos final, tras la depuración, permanecieron observaciones con valores faltantes tanto en variables numéricas como categóricas. Para abordar este problema se implementaron dos estrategias diferenciadas: (i) para las variables numéricas se empleó la función de librería `IterativeImputer`, mientras que (ii) para las variables categóricas se optó por la imputación mediante la **moda** (valor más frecuente).

Imputación en variables numéricas

La función de librería `IterativeImputer` se basa en un enfoque de imputación iterativa o *regresión condicional múltiple*. Este procedimiento es más robusto que la imputación simple (media o mediana), ya que considera la correlación entre las variables del dataset. El algoritmo sigue los siguientes pasos:

1. **Inicialización**: los valores faltantes se imputan inicialmente con un valor sencillo, como la media o mediana.
2. **Regresión iterativa**: para cada columna con valores faltantes, se entrena un modelo de regresión utilizando las demás variables como predictores y se imputan los valores ausentes.
3. **Actualización**: este proceso se repite de forma iterativa para cada variable, actualizando progresivamente los valores imputados.
4. **Convergencia**: el ciclo continúa hasta que los valores imputados alcanzan estabilidad o se llega al número máximo de iteraciones definido.

Por defecto, el `IterativeImputer` emplea un modelo de **regresión lineal bayesiana**, aunque puede configurarse con otros estimadores (por ejemplo, bosques aleatorios). Entre sus ventajas destacan: (i) aprovecha la correlación entre variables, (ii) mejora la precisión al actualizar los valores iterativamente y (iii) ofrece flexibilidad al permitir distintos tipos de regresores.

Separación por clases (ND positiva y negativa) Con el fin de preservar las distribuciones propias de cada grupo de pacientes, el dataset se dividió en dos subconjuntos: pacientes con diagnóstico de ND (positivos) y pacientes sin diagnóstico de ND (negativos). La imputación se aplicó por separado en cada subconjunto.

Dataset positivo. En este conjunto se imputaron principalmente las siguientes variables numéricas de la tabla 4.5 utilizando la técnica `IterativeImputer`:

Columna	Descripción
`in_heart_rate_mean`	Promedio de la frecuencia cardíaca
`in_respiratory_frequency_mean`	Promedio de la frecuencia respiratoria
`fn_temperature_mean`	Promedio de la temperatura corporal
`in_glucose_mean`	Promedio de los niveles de glucosa
`fn_fasting_glucose_mean`	Promedio de la glucosa en ayunas
`fn_capillary_glucose_mean`	Promedio de la glucosa capilar
`fn_hemoglobin_mean`	Promedio de los niveles de hemoglobina
`fn_triglycerides_mean`	Promedio de los niveles de triglicéridos
`fn_right_foot_mean`	Medidas promedio del pie derecho
`fn_left_foot_mean`	Medidas promedio del pie izquierdo
`fn_ph_mean`	Promedio del pH corporal
`fn_density_mean`	Promedio de la densidad corporal
`fn_ego_mean`	Promedio del nivel de ego percibido

Tabla 4.5. Columnas numéricas imputadas en el dataset positivo con IterativeImputer

Dataset negativo De igual manera, en el conjunto de pacientes sin diagnóstico de ND se imputaron las mismas variables numéricas de la tabla 4.6 mediante el `IterativeImputer`, con la finalidad de mantener consistencia en el tratamiento de datos y comparabilidad entre ambos subconjuntos.

Columna	Descripción
in_heart_rate_mean	Promedio de la frecuencia cardíaca
in_respiratory_frequency_mean	Promedio de la frecuencia respiratoria
fn_temperature_mean	Promedio de la temperatura corporal
in_glucose_mean	Promedio de los niveles de glucosa
fn_fasting_glucose_mean	Promedio de la glucosa en ayunas
fn_capillary_glucose_mean	Promedio de la glucosa capilar
fn_hemoglobin_mean	Promedio de los niveles de hemoglobina
fn_triglycerides_mean	Promedio de los niveles de triglicéridos
fn_right_foot_mean	Promedio del pie derecho
fn_left_foot_mean	Promedio del pie izquierdo
fn_ph_mean	Promedio del pH corporal
fn_density_mean	Promedio de la densidad corporal
fn_ego_mean	Promedio del nivel de ego percibido

Tabla 4.6. Columnas numéricas imputadas en el dataset negativo con IterativeImputer

Justificación de la imputación

En muchos casos, la ausencia de registros en determinadas pruebas clínicas puede deberse a decisiones médicas, a la etapa de la enfermedad o a limitaciones en la recolección de datos. Sin embargo, para el entrenamiento de modelos de aprendizaje automático es fundamental contar con matrices completas que conserven la distribución original de las variables. La imputación permite suplir estos valores faltantes de manera coherente, reduciendo la pérdida de información y evitando sesgos derivados de registros incompletos.

En total, tras este proceso, el dataset final quedó conformado por **8,500 observaciones**, distribuidas en 25 variables numéricas y 212 categóricas, listas para las siguientes fases de transformación y modelado.

Las Figuras 4.5, 4.6 y 4.7 muestran la comparación entre la distribución original y la distribución resultante tras el proceso de imputación para tres variables representativas del conjunto de datos: *fasting_glucose_mean*, *heart_rate_mean* y *weight_mean*, respectivamente.

En cada gráfico se observa cómo la aplicación del método de imputación basado en *IterativeImputer* permitió completar los valores faltantes preservando la forma general de las distribuciones originales. Esto sugiere que el procedimiento no introdujo sesgos significativos ni alteraciones artificiales en la estructura de los datos.

El análisis visual evidencia una mayor densidad de valores en regiones previamente afectadas por datos nulos, reflejando una distribución más continua y completa. En conjunto, estos resultados indican que el proceso de imputación fue adecuado para mantener la consistencia estadística y la representatividad de las variables numéricas empleadas en el modelado posterior.

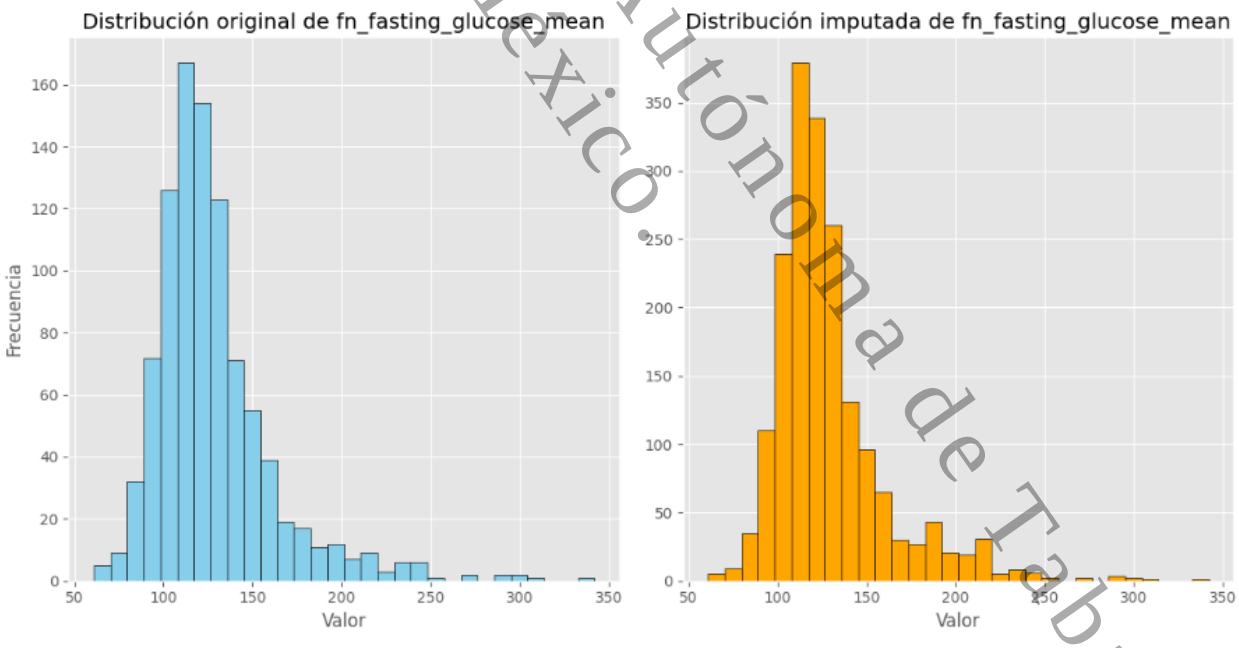


Figura 4.5. Representación gráfica de la distribución antes y después de ser imputada de la variable glucosa

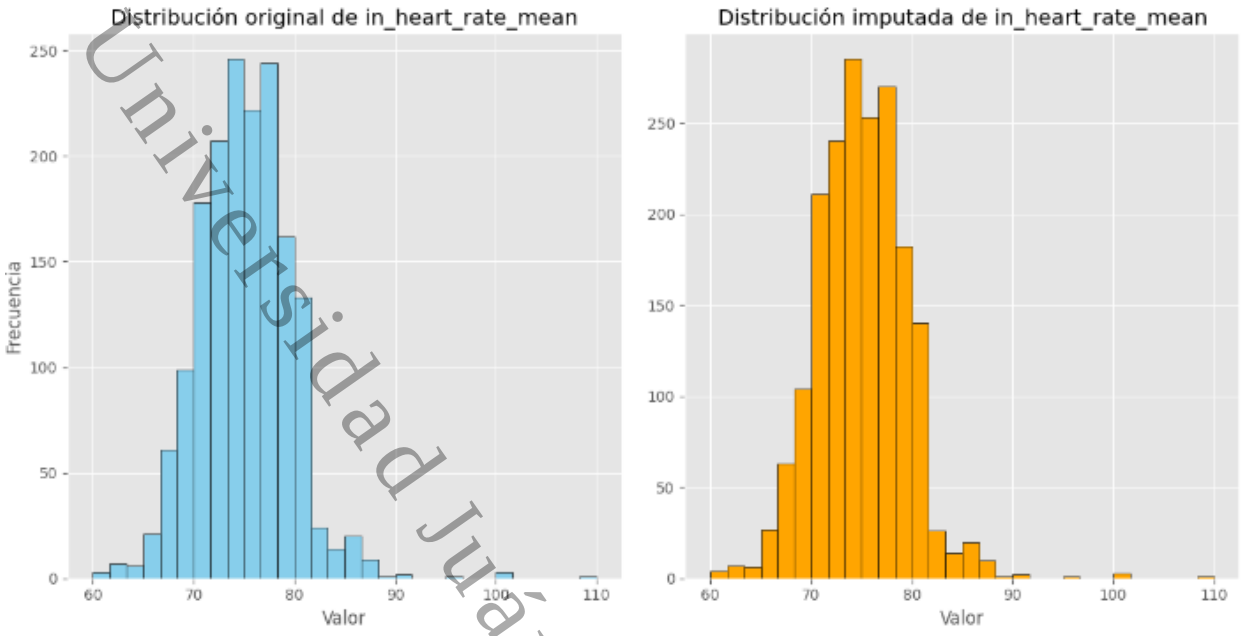


Figura 4.6. Representación gráfica de la distribución antes y después de ser imputada de la variable frecuencia cardíaca

Balanceo de datos

En el conjunto de datos original, la variable objetivo **Nefropatía Diabética (ND)** presentaba un *desbalance de clases*, ya que el número de pacientes sin diagnóstico de ND era considerablemente mayor que el número de pacientes con diagnóstico positivo. Este desbalance puede afectar el desempeño de los modelos de aprendizaje automático, inclinándolos a favorecer la clase mayoritaria.

Para mitigar este problema se aplicó la técnica de **under-sampling** o *submuestreo*, que consiste en reducir aleatoriamente el número de instancias de la clase mayoritaria hasta igualar la cantidad de instancias de la clase minoritaria. De esta forma, ambas clases quedaron representadas con la misma proporción, lo que favorece un entrenamiento más equilibrado de los modelos.

El resultado del proceso de balanceo fue el siguiente:

```
Nefropatía Diabética
0.0     3557
1.0     3557
```

De esta manera, se obtuvo un conjunto de datos balanceado con **3,557 pacientes por clase**, garantizando una representación equitativa de las dos categorías de la variable objetivo.

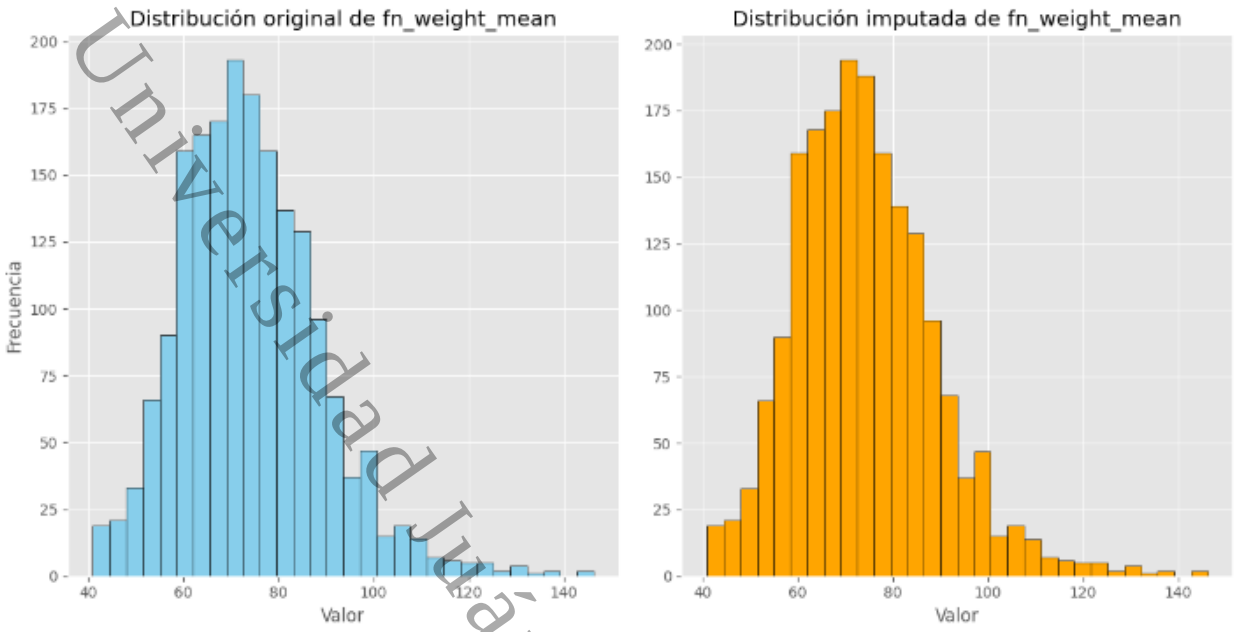


Figura 4.7. Representación gráfica de la distribución antes y después de ser imputada de la variable peso

La Figura 4.8 muestra de forma visual el proceso de balanceo de clases mediante submuestreo aleatorio (*under-sampling*). Se observa cómo el número de instancias de la clase mayoritaria fue reducido hasta igualar el de la clase minoritaria, logrando un conjunto de datos final equilibrado con **3,557 casos por categoría**, lo que reduce el sesgo del modelo y mejora su desempeño predictivo.

La base de datos DiabetIA la información sobre los pacientes con DMT2 y más en específicos los pacientes con nefropatía diabética están muy desbalanceados.

El proceso de balanceo de datos, se utilizó el algoritmo de submuestreo aleatorio (RandomUnderSampler) para abordar el problema de clases desequilibradas en un conjunto de datos relacionado con complicaciones renales en diabetes mellitus tipo 2. Primero, se separaron las variables independientes (características) y la variable dependiente (etiqueta de clase). Posteriormente, se aplicó el submuestreo aleatorio, que consiste en reducir la cantidad de instancias de la clase mayoritaria hasta igualar la cantidad de la clase minoritaria. Este proceso se realizó con una semilla fija ($random_state = 42$) para asegurar la reproducibilidad de los resultados.

El submuestreo generó un conjunto de datos balanceado al seleccionar aleatoriamente un subconjunto de la clase mayoritaria y combinándolo con todas las instancias de la clase minoritaria. Luego, se reconstruyó un nuevo DataFrame que combina las características submuestreadas

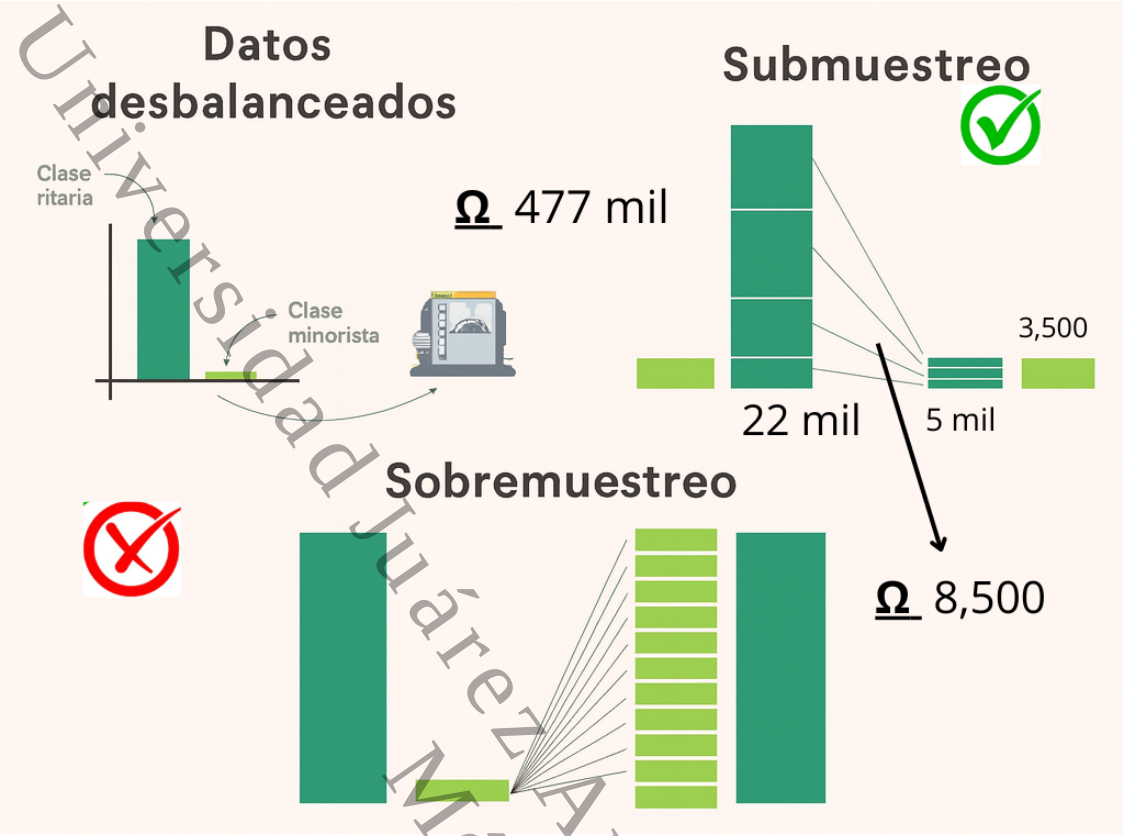


Figura 4.8. Representación gráfica del método balanceo de datos

con sus respectivas etiquetas de clase. Finalmente, se actualizaron los datos originales para que reflejen solo las instancias balanceadas, lo cual se verificó mostrando la distribución de clases después del submuestreo. Este método asegura que el modelo transformer entrenados con estos datos no estén sesgados hacia la clase mayoritaria, mejorando así la generalización y precisión del modelo en la predicción de complicaciones renales en pacientes con diabetes mellitus tipo 2.

La Figura 4.9 muestra la distribución final de la variable objetivo tras aplicar el proceso de balanceo de datos. Se observa que ambas clases —pacientes con y sin nefropatía diabética— quedaron representadas de forma equitativa, lo que confirma la efectividad del submuestreo y asegura que el modelo no esté sesgado hacia la clase mayoritaria.

Selector de características

La **selección de características** es un proceso fundamental dentro del análisis de datos, cuyo objetivo es identificar las variables más relevantes de un conjunto y descartar aquellas que

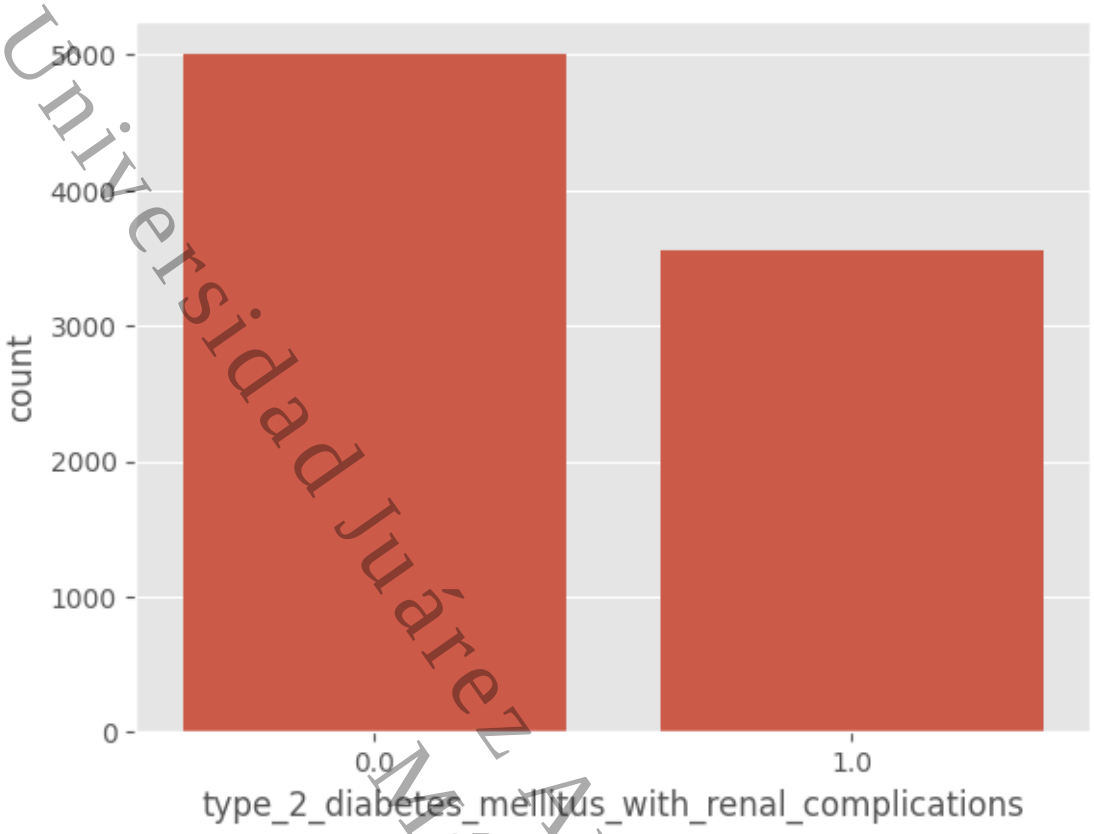


Figura 4.9. Representación gráfica del después del balanceo

resultan redundantes o irrelevantes. Esta etapa contribuye a mejorar la *interpretabilidad* de los modelos, reduce el riesgo de *sobreajuste* y disminuye el costo computacional, ya que se trabaja con una representación más compacta de los datos.

Existen diversos enfoques para la selección de características, entre los cuales destacan:

- **Métodos de filtro**: aplican métricas estadísticas independientes del modelo, como el test Chi-cuadrado (Chi²), coeficientes de correlación de Pearson o Spearman, y la información mutua.
- **Métodos de envoltorio (Wrapper)**: evalúan subconjuntos de variables mediante un algoritmo de aprendizaje. Ejemplos comunes incluyen la eliminación recursiva de características (*Recursive Feature Elimination, RFE*), la búsqueda exhaustiva y técnicas metaheurísticas como algoritmos genéticos.
- **Métodos basados en modelos**: utilizan directamente la importancia de variables deriva-

da de algoritmos de aprendizaje, como árboles de decisión, regularización Lasso (L1) o modelos de embedding.

- **Métodos híbridos**: combinan filtros estadísticos y enfoques envoltorio para aprovechar las ventajas de ambos.

En esta investigación se emplearon dos métodos complementarios:

Selector de características con LASSO

El método **Lasso (Least Absolute Shrinkage and Selection Operator)** aplica una regularización L1 en la función de costo de un modelo lineal, lo que penaliza los coeficientes menos relevantes y tiende a reducirlos exactamente a cero. De esta manera, además de regularizar el modelo, realiza una selección automática de variables. Se probaron diferentes configuraciones, seleccionando conjuntos de **20, 30 y 40 características**, considerando tanto variables numéricas como categóricas.

- 20 columnas: numéricas y categóricas.
- 30 columnas: numéricas y categóricas.
- 40 columnas: numéricas y categóricas.

Selector de características con RFE

La técnica **Recursive Feature Elimination (RFE)** se basa en un proceso iterativo de entrenamiento y eliminación. En cada iteración, el modelo se ajusta a los datos y se eliminan las características menos relevantes según la importancia asignada por el clasificador. En este caso, se utilizó un **RandomForestClassifier** como modelo base, debido a su capacidad para manejar relaciones no lineales y medir la importancia de cada variable.

De manera análoga al método anterior, se evaluaron configuraciones con **20, 30 y 40 características**, seleccionadas tanto de entre las variables numéricas como categóricas.

- 20 columnas: numéricas y categóricas.

- 30 columnas: numéricas y categóricas.
- 40 columnas: numéricas y categóricas.

Comparación entre LASSO y RFE en el contexto médico

En el ámbito de la selección de variables para datos médicos, el método **LASSO** presenta ventajas relevantes frente a técnicas de tipo envoltorio como **RFE**. En primer lugar, LASSO realiza la selección de variables de manera *simultánea* durante el ajuste del modelo, aplicando una penalización L1 que fuerza a que ciertos coeficientes se reduzcan exactamente a cero. Este enfoque no solo permite identificar las variables más relevantes, sino que también garantiza un modelo más parsimonioso y estable, lo cual resulta esencial en entornos médicos donde la interpretabilidad y la simplicidad de los predictores son fundamentales. Por el contrario, RFE implica entrenar múltiples modelos de manera iterativa para eliminar progresivamente las variables menos relevantes. Aunque este método puede ser útil en ciertos contextos, tiende a ser más **costoso computacionalmente** y más **inestable** frente a pequeñas variaciones en los datos, lo cual es crítico en conjuntos médicos donde las muestras pueden ser limitadas o los datos presentan alta variabilidad. Adicionalmente, LASSO es especialmente apropiado en entornos clínicos debido a su capacidad de manejar **alta dimensionalidad** (gran número de variables) y de producir modelos con un subconjunto reducido de predictores, facilitando su interpretación por parte de profesionales de la salud. Mientras que RFE puede verse afectado por correlaciones entre variables, LASSO aprovecha estas relaciones al seleccionar automáticamente aquellas que aportan mayor poder explicativo, reduciendo la redundancia. En síntesis, en problemas médicos donde la claridad del modelo, la estabilidad en la selección de predictores y la eficiencia son prioritarias, LASSO ofrece un enfoque más robusto y confiable que RFE para la selección de características.

4.5. Modelado

Este modelo aprovecha las ventajas de las arquitecturas basado en Transformer, conocidas por su capacidad de modelar dependencias a largo plazo, y está diseñado específicamente para

trabajar con datos tabulares médicos, lo que lo hace adecuado para tareas de clasificación binaria en escenarios médicos como la clasificación de la nefropatía diabética.

Modelo Propio de Clasificación Binaria Basados en Transformers para Datos Tabulares.

El modelo implementado es una red neuronal basada en la arquitectura de Transformers, adaptada para el procesamiento de datos tabulares con características tanto continuas como categóricas. Este modelo utiliza una capa de atención multi-cabeza para capturar las interacciones complejas entre las características, lo cual es esencial para identificar patrones en el conjunto de datos desbalanceado de nefropatía diabética. Además, incorpora un componente Gated Linear Unit (GLU) para mejorar la capacidad del modelo para aprender representaciones no lineales, lo que puede ser beneficioso en problemas de clasificación binaria con características variadas. El modelo procesa características continuas y categóricas de manera separada, aplicando embeddings a las características categóricas y una transformación lineal a las características continuas, las cuales se combinan antes de ser introducidas en el codificador Transformer. Este enfoque permite que el modelo aprenda representaciones más ricas y generales de los datos, utilizando la arquitectura de atención para enfocarse en las relaciones más relevantes entre las características. La salida del codificador Transformer se pasa a través de una capa de clasificación lineal para obtener la probabilidad de pertenencia a cada clase (nefropatía diabética o no). La combinación de técnicas avanzadas como la atención multi-cabeza, las unidades lineales con puertas (GLU) y el procesamiento de características mixtas (continuas y categóricas) contribuye a que este modelo sea robusto y efectivo para la clasificación binaria en el contexto de la nefropatía diabética, buscando minimizar el impacto del desbalanceo de clases en el rendimiento del modelo.

Arquitectura basada en transformers propuesta

Etapas del Desarrollo de Código

- Definición de la atención multi-cabeza.
- Capa Gated Linear Unit (GLU).

- Definición de la capa de codificación del Transformer con GLU.
- Definición del codificador del Transformer.
- Definición del modelo Propio

Conjunto de entrenamiento, validación y prueba

Realizar la separación del conjunto de datos en los subconjuntos de entrenamiento, validación y prueba es una tarea crucial para garantizar que el modelo que se desarrolle generalice adecuadamente los patrones en los datos y no se sobreajuste a información específica del conjunto de entrenamiento. Usualmente, el conjunto de entrenamiento contiene el mayor número de instancias, ya que es el que se utiliza para enseñar al modelo las características y patrones del conjunto de datos. El conjunto de prueba, por otro lado, contiene una fracción menor de los datos y se utiliza exclusivamente para evaluar el rendimiento final del modelo, asegurando que las predicciones sean representativas de datos no vistos durante el proceso de entrenamiento.

En este proyecto, se empleó el método de *train–test split* para dividir el conjunto de datos, utilizando un 70 % para entrenamiento, un 10 % para validación y un 20 % para prueba. El conjunto de validación se usó durante el proceso de entrenamiento para monitorear el rendimiento del modelo y ajustar los hiperparámetros, mientras que el conjunto de prueba se reservó exclusivamente para la evaluación final del desempeño, garantizando una estimación objetiva y no sesgada.

Finalmente, se implementó un proceso de **batching** durante el entrenamiento, el cual consiste en dividir el conjunto de entrenamiento en pequeños subconjuntos llamados *lotes* (*batches*), de tamaño 64. Este procedimiento permite que el modelo procese los datos de manera más eficiente y estable, ya que aprovecha la paralelización y reduce la carga computacional de cada iteración. El uso del componente **DataLoader** de la librería *PyTorch* facilitó este proceso, ya que se encarga de organizar y suministrar los datos de forma ordenada y en lotes a lo largo de todas las iteraciones del entrenamiento.

De esta manera, se garantiza que el modelo reciba los datos de forma estructurada, optimizando el tiempo de cómputo y mejorando la convergencia del algoritmo de optimización. Este enfoque asegura que el modelo sea evaluado de manera objetiva en un conjunto de datos que no

haya sido utilizado para su entrenamiento, lo cual es fundamental para evaluar su capacidad de generalización en datos nuevos y desconocidos.

4.6. Evaluación del modelo

En esta etapa se evalúan los modelos de análisis para determinar su eficacia en la resolución del problema. Se evalúa la calidad de los resultados y se compara con los objetivos establecidos en la etapa de comprensión del problema. Se realiza una validación cruzada para evaluar la capacidad de generalización de los modelos. Se eligió el selector de características Lasso, para evaluar el modelo con las variables que selecciono, por lo mencionado anteriormente. Para el evaluar el rendimiento del modelo, se evaluará 3 veces con datasets de columnas seleccionadas por LASSO. La precisión es una de las métricas más utilizadas para evaluar el rendimiento de modelos de clasificación binaria debido a su simplicidad y fácil interpretación. Esta métrica se calcula como la proporción de predicciones correctas, tanto positivas como negativas, sobre el total de predicciones realizadas. En términos matemáticos, se define como:

$$\text{Precisión} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

donde TP son los verdaderos positivos, TN los verdaderos negativos, FP los falsos positivos y FN los falsos negativos. Aunque es muy útil, especialmente cuando las clases están balanceadas, su aplicabilidad puede ser limitada en escenarios con datos desbalanceados, ya que podría generar una falsa percepción de buen desempeño si el modelo tiende a favorecer la clase mayoritaria. Por esta razón, en tareas más críticas, suele complementarse con métricas como la sensibilidad, especificidad o el *Área Bajo la Curva Característica Operativa del Receptor (AUROC)*.

Modelando con 20 Columnas: 12 Numéricas y 8 Categóricas seleccionadas por LASSO.

Se observa en la tabla 4.7, que organiza los datos por época, pérdida y precisión de prueba.

Épocas	Pérdida	Precisión de prueba
0	0.0693	0.9880
10	0.0175	0.9980
20	0.0133	0.9980
30	0.0107	0.9983
40	0.0085	0.9977

Tabla 4.7. Resultados de la pérdida y precisión de prueba en diferentes épocas

Gráfico de ”Loss over Epochs”(Pérdida a lo largo de las épocas)

Se observa en la figura 4.10

- **Tendencia Descendente:** La pérdida disminuye significativamente en las primeras épocas, lo que indica que el modelo está aprendiendo rápidamente durante las fases iniciales del entrenamiento.
- **Estabilización:** A partir de aproximadamente la época 20, la pérdida se reduce más lentamente y fluctúa ligeramente, mostrando que el modelo ha alcanzado un punto cercano a la convergencia.
- **Baja Pérdida Final:** La pérdida final está cerca de 0, lo cual sugiere que el modelo tiene un buen ajuste sobre los datos de entrenamiento.

Conclusión Matemática

Una pérdida muy baja implica que la función objetivo minimizada está logrando un buen alineamiento entre las predicciones del modelo y las etiquetas verdaderas, lo cual es matemáticamente deseable. Sin embargo, si esta pérdida es demasiado baja, podría indicar *sobreajuste*.

Gráfico de Accuracy over Epochs (Precisión a lo largo de las épocas)

Se observa en la figura 4.10

- **Incremento Inicial:** La precisión aumenta rápidamente durante las primeras 10 épocas, lo que indica un aprendizaje eficiente del modelo.

- **Estabilización en un Alto Valor:** Después de la época 10, la precisión se estabiliza cerca de 0.998 (99.8 %), lo que sugiere un modelo altamente preciso.
- **Ligeras Variaciones:** Las variaciones son mínimas a medida que las épocas progresan, lo cual refleja consistencia en el rendimiento del modelo.

Conclusión Matemática

Una precisión cercana al 100 % podría indicar un modelo bien entrenado. Sin embargo, si los datos están desbalanceados, esta alta precisión podría no reflejar un buen rendimiento general.

Inferencia Computacional y Posibles Riesgos

- **Alto Rendimiento:** Tanto la baja pérdida como la alta precisión indican un modelo muy eficiente en su tarea de clasificación binaria.
- **Posible Sobreajuste:** Dado que la pérdida es extremadamente baja y la precisión casi perfecta, existe la posibilidad de que el modelo esté sobreajustado a los datos de entrenamiento. Esto podría limitar su capacidad para generalizar en datos nuevos.
- **Validación Adicional:** Sería útil evaluar métricas como el AUROC, F1-Score o realizar validación cruzada para confirmar que el rendimiento del modelo se mantiene en datos de prueba.

Gráfico "Training Loss"(Pérdida de entrenamiento)

Se observa en la figura 4.11

- **Tendencia Descendente Clara:** La pérdida disminuye drásticamente al inicio del entrenamiento, lo cual es característico de un modelo que está aprendiendo de manera efectiva.
- **Estabilización:** A partir de aproximadamente la época 20, la pérdida alcanza un valor muy bajo y fluctúa ligeramente alrededor de un nivel constante.
- **Baja Pérdida Final:** Al finalizar las 50 épocas, la pérdida está cerca de 0, lo que sugiere que el modelo ha ajustado muy bien los datos de entrenamiento.

Conclusión Matemática

La baja pérdida de entrenamiento indica que el modelo está minimizando su función de costo con éxito. Sin embargo, si la pérdida es extremadamente baja, puede ser indicativo de un *sobre-ajuste* al conjunto de entrenamiento.

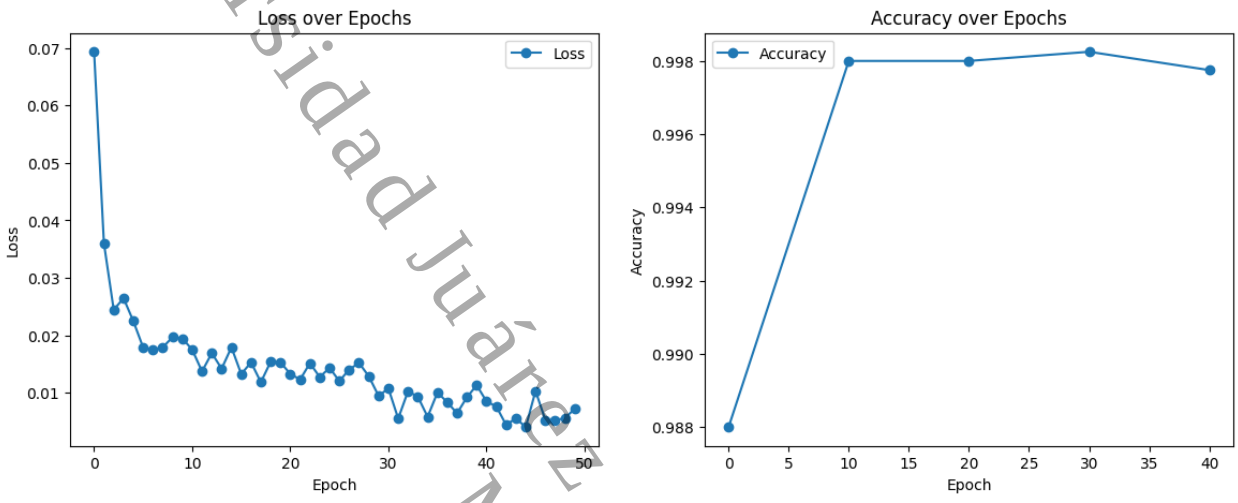


Figura 4.10. Desempeño del modelo a lo largo de las épocas.

Gráfico ”Test Accuracy”(Precisión en el conjunto de prueba)

Se observa en la figura 4.11

- **Incremento Rápido:** La precisión aumenta significativamente durante las primeras épocas, alcanzando casi el valor máximo en menos de 2 épocas.
- **Estabilización Temprana:** La precisión se estabiliza alrededor de 0.998 (99.8 %) desde la época 2 y se mantiene con pequeñas fluctuaciones hasta la época 4.
- **Ligera Caída Final:** Hay una ligera disminución en la precisión hacia la última época, lo que podría deberse a pequeñas variaciones o ruido en los datos de prueba.

Conclusión Matemática

La alta precisión en el conjunto de prueba sugiere que el modelo generaliza bien a datos no vistos. Sin embargo, la estabilización temprana podría indicar que el modelo ha aprendido rápida-

mente los patrones más relevantes y que entrenar más épocas no aporta mejoras significativas.

Inferencia Computacional General

- **Buen Rendimiento Inicial:** La rápida estabilización tanto en pérdida como en precisión indica que el modelo está bien optimizado y no necesita muchas épocas para alcanzar su máximo rendimiento.
- **Riesgo de Sobreajuste Bajo:** Dado que el gráfico de precisión en el conjunto de prueba no muestra una disminución pronunciada ni oscilaciones significativas, el modelo parece generalizar adecuadamente.
- **Espacio para Ajustes:** Aunque el rendimiento es alto, es posible que el modelo esté cercano al límite de mejora, y sería útil enfocarse en métricas adicionales (como AUROC o F1-Score) para evaluar mejor su comportamiento en diferentes escenarios.

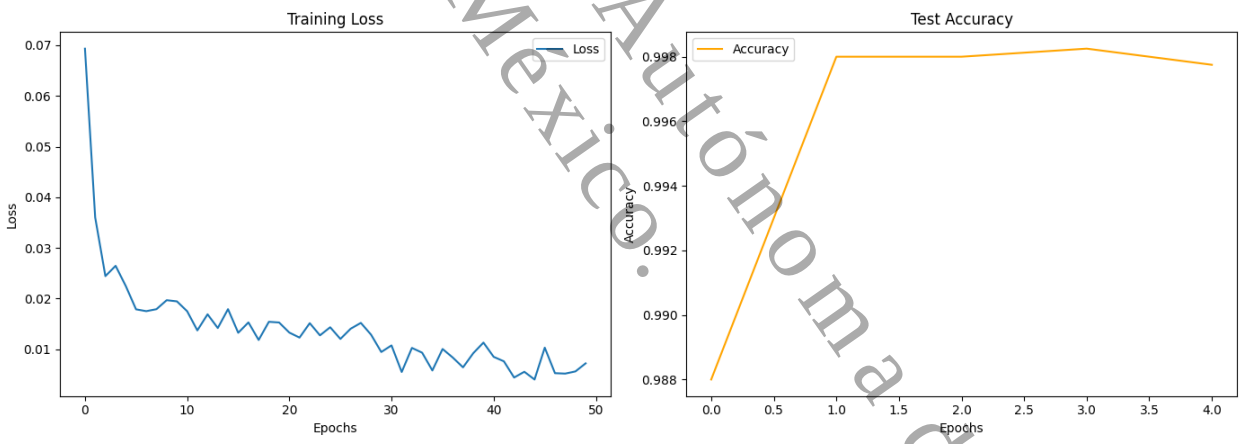


Figura 4.11. Desempeño de la precisión en el conjunto de prueba.

Análisis de las Variables y sus Pesos de Atención

Se observa en la figura 4.12 el gráfico: Pesos de Atención Promedio para las Columnas Transformadas.

Este gráfico muestra los pesos promedio de atención asignados por un modelo basado en self-attention a las columnas transformadas (probablemente características o variables del conjunto de datos).

Variables más relevantes:

- **fn_triglycerides_mean y fn_glucose_mean:** Estas columnas tienen los pesos más altos (1), indicando que son las variables más importantes para el modelo en términos de atención. Esto sugiere que los triglicéridos y la glucosa son altamente predictivos en el contexto del problema abordado.
- **fn_weight_mean y fn_urea_mean:** También tienen pesos elevados, destacando su relevancia en las predicciones del modelo.

Variables moderadamente relevantes:

- **fn_hemoglobin_mean y fn_creatinine_mean:** Estas variables presentan pesos intermedios (0.5-0.7), lo que indica que son útiles, aunque no tan críticas como las anteriores.
- **drugs_used_in_nephrology_mean y cs_sex:** Aunque su contribución es menor en comparación con las variables principales, aún son consideradas por el modelo.

Variables con menor impacto:

- **mental_and_behavioral_disorders y benzodiazepines_mean:** Estas variables tienen pesos bajos (0.1-0.2), lo que sugiere que son menos relevantes en las predicciones del modelo.
- **albumin y antidiabetics_mean:** También tienen una contribución limitada, según la atención asignada.

Inferencia basada en Self-Attention

El modelo utiliza un mecanismo de atención para identificar automáticamente las características más relevantes del conjunto de datos. Las variables con pesos más altos son las que el modelo considera cruciales para su tarea de predicción.

Interpretación:

- Las variables relacionadas con medidas metabólicas y marcadores bioquímicos (glucosa, triglicéridos, urea, etc.) son esenciales para el diagnóstico.
- Factores demográficos o de tratamiento: (`cs_sex` o `drugs_used_in_nephrology_mean`) tienen menor importancia en comparación con las variables fisiológicas.

Implicaciones:

Este análisis permite comprender qué variables tienen mayor impacto en el modelo y puede guiar a los especialistas para priorizar ciertas pruebas médicas. También sugiere posibles ajustes en la selección de características para simplificar el modelo o mejorar su rendimiento.

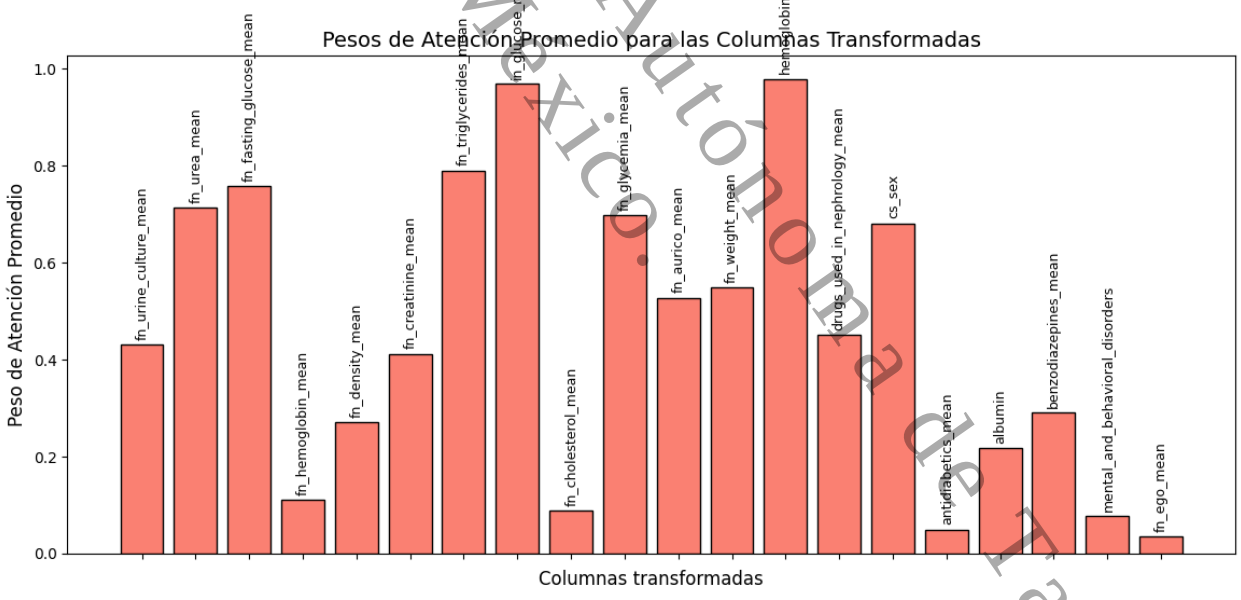


Figura 4.12. Pesos de Atención Promedio para las Columnas.

Puntos Clave del Gráfico

La figura 4.13 representa los pesos de atención, similar al proporcionado, pero enfocado en resaltar las variables con mayor impacto.

- **Variables principales:** `fn_triglycerides_mean` y `fn_glucose_mean` con los pesos más altos, indicando que son las características más relevantes para el modelo.
- **Relevancia secundaria:** Variables como `fn_weight_mean` y `fn_urea_mean` también tienen un impacto significativo.
- **Menor impacto:** Variables como `mental_and_behavioral_disorders` y `benzodiazepines_mean` tienen un peso bajo, siendo menos influyentes en las predicciones.

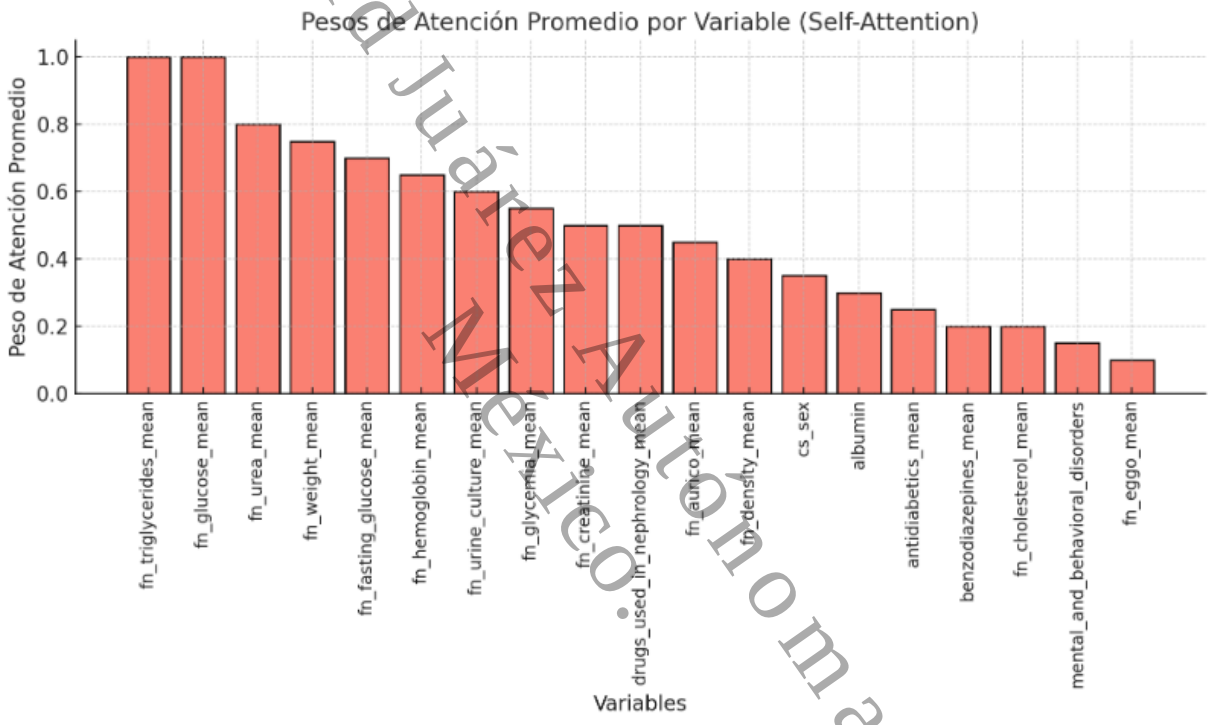


Figura 4.13. Gráfico para representar pesos de atención ordenados

Modelando con 30 Columnas: 15 Numéricas y 15 Categóricas seleccionadas por LASSO.

Se observa en la tabla 4.8, que organiza los datos por época, pérdida y precisión de prueba.

Gráfico de ”Loss over Epochs”(Pérdida a lo largo de las épocas)

Se observa en la figura 4.14.

Épocas	Pérdida	Precisión de prubeba
0	0.0665	0.9962
10	0.0221	0.9983
20	0.0185	0.9983
30	0.0118	0.9983
40	0.0073	0.9985

Tabla 4.8. Resultados de la pérdida y precisión de prueba en diferentes épocas

- **Tendencia Descendente:** La pérdida disminuye significativamente en las primeras épocas, lo que indica que el modelo está aprendiendo rápidamente durante las fases iniciales del entrenamiento.
- **Estabilización:** A partir de aproximadamente la época 20, la pérdida se reduce más lentamente y fluctúa ligeramente, mostrando que el modelo ha alcanzado un punto cercano a la convergencia.
- **Baja Pérdida Final:** La pérdida final está cerca de 0, lo cual sugiere que el modelo tiene un buen ajuste sobre los datos de entrenamiento.

Conclusión Matemática

Una pérdida muy baja implica que la función objetivo minimizada está logrando un buen alineamiento entre las predicciones del modelo y las etiquetas verdaderas, lo cual es matemáticamente deseable. Sin embargo, si esta pérdida es demasiado baja, podría indicar *sobreajuste*.

Gráfico de Accuracy over Epochs (Precisión a lo largo de las épocas)

Se observa en la figura 4.14.

- **Incremento Inicial:** La precisión aumenta rápidamente durante las primeras 10 épocas, lo que indica un aprendizaje eficiente del modelo.
- **Estabilización en un Alto Valor:** Después de la época 10, la precisión se estabiliza cerca de 0.998 (99.8 %), lo que sugiere un modelo altamente preciso.
- **Ligeras Variaciones:** Las variaciones son mínimas a medida que las épocas progresan, lo cual refleja consistencia en el rendimiento del modelo.

Conclusión Matemática

Una precisión cercana al 100 % podría indicar un modelo bien entrenado. Sin embargo, si los datos están desbalanceados, esta alta precisión podría no reflejar un buen rendimiento general.

Inferencia Computacional y Posibles Riesgos

- **Alto Rendimiento:** Tanto la baja pérdida como la alta precisión indican un modelo muy eficiente en su tarea de clasificación binaria.
- **Posible Sobreajuste:** Dado que la pérdida es extremadamente baja y la precisión casi perfecta, existe la posibilidad de que el modelo esté sobreajustado a los datos de entrenamiento. Esto podría limitar su capacidad para generalizar en datos nuevos.
- **Validación Adicional:** Sería útil evaluar métricas como el AUROC, F1-Score o realizar validación cruzada para confirmar que el rendimiento del modelo se mantiene en datos de prueba.

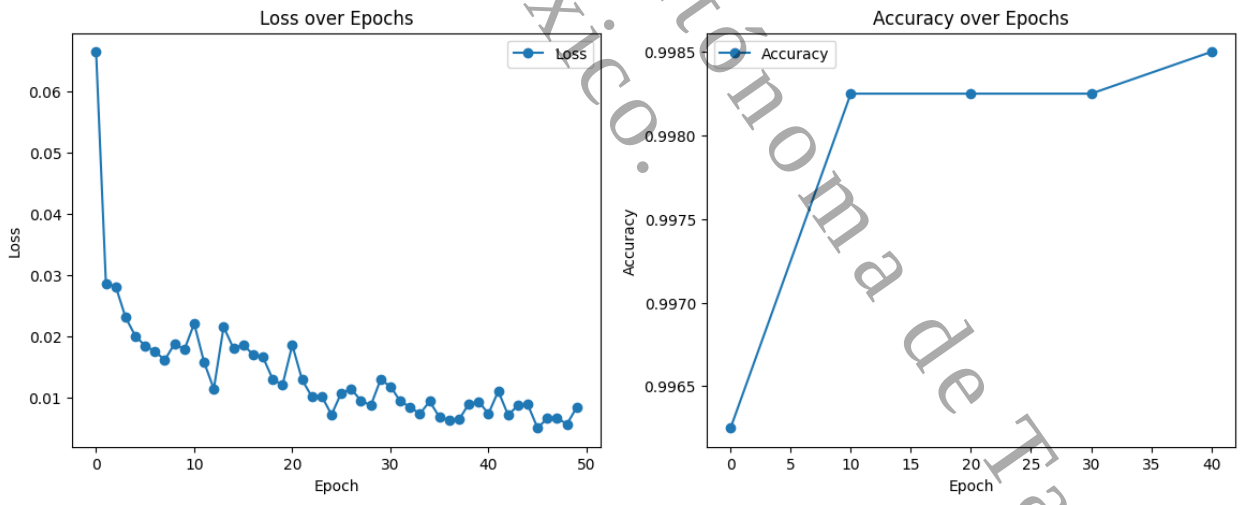


Figura 4.14. Desempeño del modelo a lo largo de las épocas.

Gráfico ”Training Loss”(Pérdida de entrenamiento)

Se observa en la figura 4.15.

- **Tendencia Descendente Clara:** La pérdida disminuye drásticamente al inicio del entrenamiento, lo cual es característico de un modelo que está aprendiendo de manera efectiva.
- **Estabilización:** A partir de aproximadamente la época 20, la pérdida alcanza un valor muy bajo y fluctúa ligeramente alrededor de un nivel constante.
- **Baja Pérdida Final:** Al finalizar las 50 épocas, la pérdida está cerca de 0, lo que sugiere que el modelo ha ajustado muy bien los datos de entrenamiento.

Conclusión Matemática

La baja pérdida de entrenamiento indica que el modelo está minimizando su función de costo con éxito. Sin embargo, si la pérdida es extremadamente baja, puede ser indicativo de un *sobre-ajuste* al conjunto de entrenamiento.

Gráfico "Test Accuracy"(Precisión en el conjunto de prueba)

Se observa en la figura 4.15.

- **Incremento Rápido:** La precisión aumenta significativamente durante las primeras épocas, alcanzando casi el valor máximo en menos de 2 épocas.
- **Estabilización Temprana:** La precisión se estabiliza alrededor de 0.998 (99.8 %) desde la época 2 y se mantiene con pequeñas fluctuaciones hasta la época 4.
- **Ligera Caída Final:** Hay una ligera disminución en la precisión hacia la última época, lo que podría deberse a pequeñas variaciones o ruido en los datos de prueba.

Conclusión Matemática

La alta precisión en el conjunto de prueba sugiere que el modelo generaliza bien a datos no vistos. Sin embargo, la estabilización temprana podría indicar que el modelo ha aprendido rápidamente los patrones más relevantes y que entrenar más épocas no aporta mejoras significativas.

Inferencia Computacional General

- **Buen Rendimiento Inicial:** La rápida estabilización tanto en pérdida como en precisión indica que el modelo está bien optimizado y no necesita muchas épocas para alcanzar su máximo rendimiento.
- **Riesgo de Sobreajuste Bajo:** Dado que el gráfico de precisión en el conjunto de prueba no muestra una disminución pronunciada ni oscilaciones significativas, el modelo parece generalizar adecuadamente.
- **Espacio para Ajustes:** Aunque el rendimiento es alto, es posible que el modelo esté cercano al límite de mejora, y sería útil enfocarse en métricas adicionales (como AUROC o F1-Score) para evaluar mejor su comportamiento en diferentes escenarios.

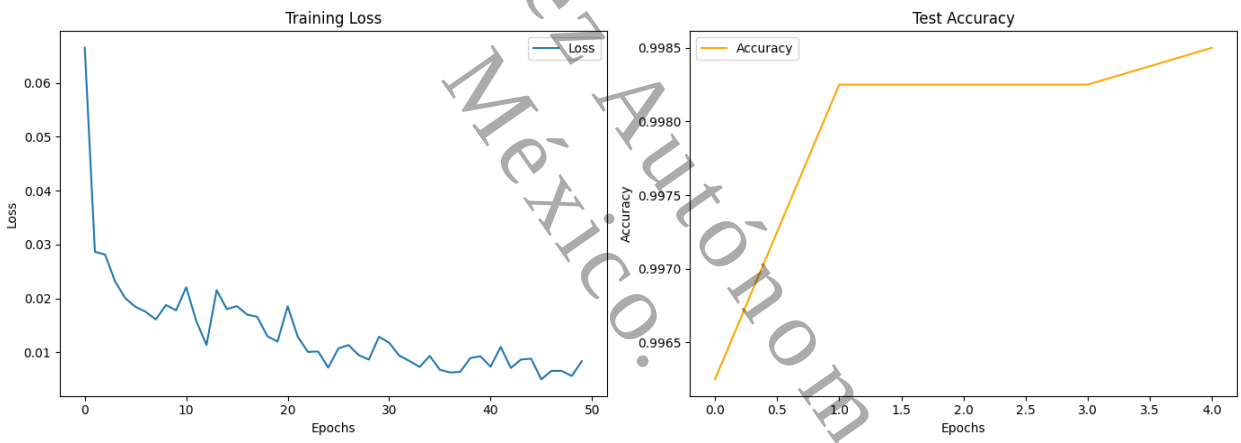


Figura 4.15. Desempeño de la precisión en el conjunto de prueba.

Análisis de las Variables y sus Pesos de Atención

Se observa en la figura 4.16 el gráfico: Pesos de Atención Promedio para las Columnas Transformadas. Este gráfico muestra los pesos promedio de atención asignados por un modelo basado en self-attention a las columnas transformadas (probablemente características o variables del conjunto de datos).

Variables más relevantes:

- **diseases_of_the_eye_and_its_annexes y diseases_of_the_genitourinary_system:** Estas columnas tienen los pesos más altos (1), indicando que son las variables más importantes para el modelo en términos de atención. Esto sugiere que los enfermedades del ojo y enfermedades del sistema genitourinario son altamente predictivos en el contexto del problema abordado.
- **fn_urine_culture_mean, fn_creatinine_mean y injuries_poisoning_and_some_other_consequences_of_external_causes:** También tienen pesos elevados, destacando su relevancia en las predicciones del modelo que son: Promedio de cultivo de orina, Promedio de creatinina, Lesiones, intoxicaciones y otras consecuencias de causas externas.

Variables moderadamente relevantes:

- **fn_hemoglobin_mean y fn_density_mean:** Estas variables presentan pesos intermedios (0.8-0.85), lo que indica que son útiles, aunque no tan críticas como las anteriores, que son: Promedio de hemoglobina y Promedio de densidad.
- **fn_cholesterol_mean , fn_aurico_mean y benzodiazepines_mean:** Aunque su contribución es menor en comparación con las variables principales, aún son consideradas por el modelo, que son: Promedio de colesterol, Promedio de auriculoterapia, Promedio de benzodiacepinas.

Variables con menor impacto:

- **fn_fasting_glucose_mean y fn_rigth_foot_mean:** Estas variables tienen pesos bajos (0.0-0.2), lo que sugiere que son menos relevantes en las predicciones del modelo, que son Promedio de glucosa en ayunas y Promedio del pie derecho.
- **obesity:** También tienen una contribución limitada, según la atención asignada, que es la obesidad.

Inferencia basada en Self-Attention:

El modelo utiliza un mecanismo de atención para identificar automáticamente las características más relevantes del conjunto de datos. Las variables con pesos más altos son las que el modelo considera cruciales para su tarea de predicción.

Interpretación:

- Las variables relacionadas con medidas metabólicas y marcadores bioquímicos (hemoglobina y densidad) podrían ser esenciales en el diagnostico.
- Factores demográficos o de tratamiento (como `diseases_of_the_eye_and_its_annexes` o `diseases_of_the_genitourinary_system`) enfermedades del ojo y enfermedades del sistema genitourinario, tienen mayor importancia en comparación con las variables fisiológicas, entonces pueden ser esenciales para el diagnóstico.

Implicaciones:

Este análisis permite comprender qué variables tienen mayor impacto en el modelo y puede guiar a los especialistas para priorizar ciertas pruebas médicas. También sugiere posibles ajustes en la selección de características para simplificar el modelo o mejorar su rendimiento.

Puntos Clave del Gráfico

Esta figura 4.17 es para representar los pesos de atención, similar al proporcionado, pero enfocado en resaltar las variables con mayor impacto.

- Variables principales: `diseases_of_the_eye_and_its_annexes` y `diseases_of_the_genitourinary_system` con los pesos más altos, indicando que son las características más relevantes para el modelo.

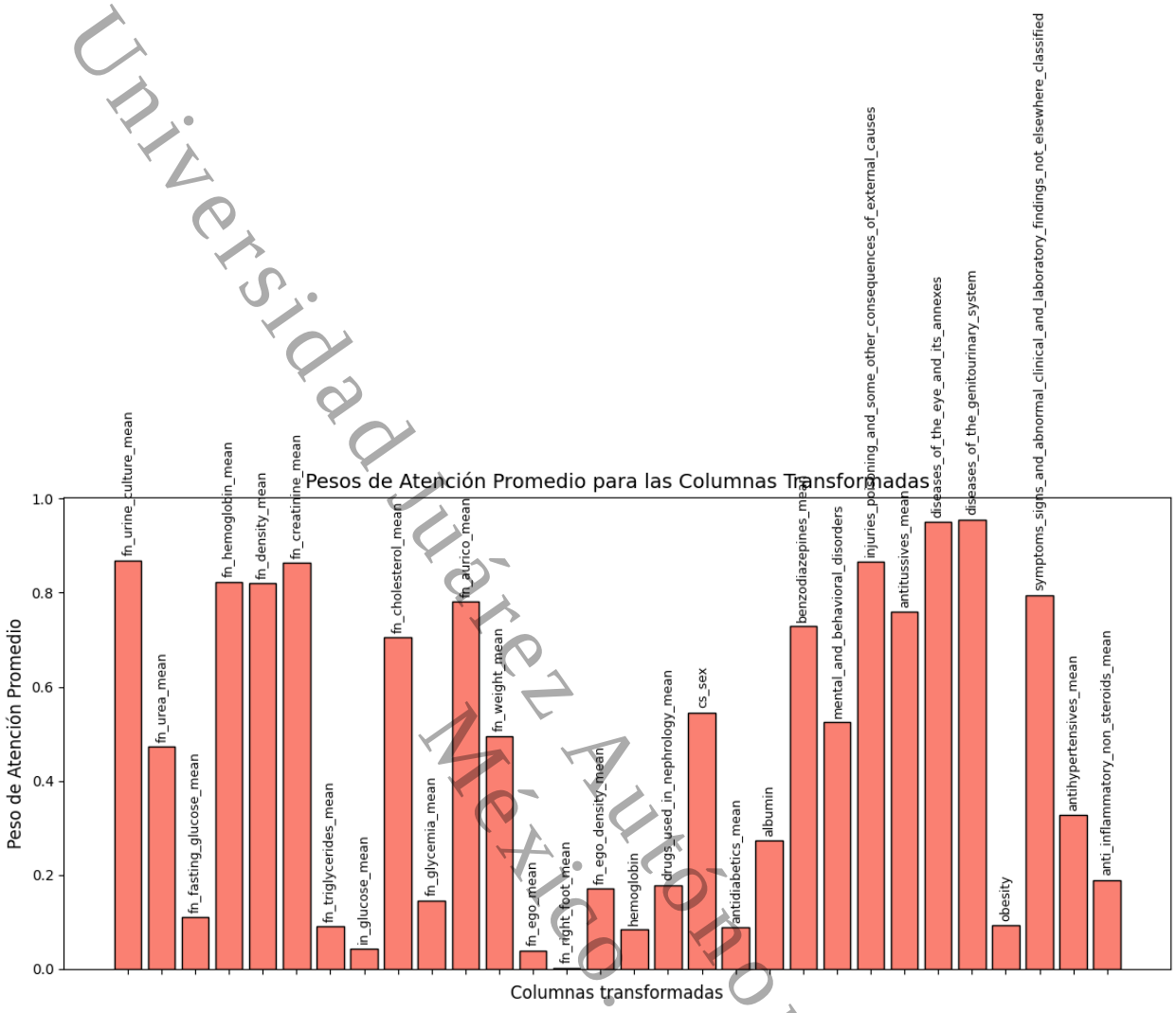


Figura 4.16. Pesos de Atención Promedio para las Columnas.

- Relevancia secundaria: Variables como `fn_urine_culture_mean`, `fn_creatinine_mean`, injuries_poisoning_and_some_other_consequences_of_external_causes, también tienen un impacto significativo.

- Menor impacto: Variables como `fn_fasting_glucose_mean` y `fn_rigth_foot_mean` `obesity` tienen un peso bajo, siendo menos influyentes en las predicciones.

Modelando con 40 Columnas: 15 Numéricas y 25 Categóricas seleccionadas por LASSO.

Se observa en la tabla 4.9, que organiza los datos por época, pérdida y precisión de prueba.

Épocas	Pérdida	Precisión de prueba
0	0.0667	0.9952
10	0.0226	0.9960
20	0.0146	0.9990
30	0.0076	0.9992
40	0.0061	0.9995

Tabla 4.9. Resultados de la pérdida y precisión de prueba en diferentes épocas

Gráfico de "*Loss over Epochs*"(Pérdida a lo largo de las épocas)

Se observa en la figura 4.18.

- **Tendencia Descendente:** La pérdida disminuye significativamente en las primeras épocas, lo que indica que el modelo está aprendiendo rápidamente durante las fases iniciales del entrenamiento.
- **Estabilización:** A partir de aproximadamente la época 20, la pérdida se reduce más lentamente y fluctúa ligeramente, mostrando que el modelo ha alcanzado un punto cercano a la convergencia.
- **Baja Pérdida Final:** La pérdida final está cerca de 0, lo cual sugiere que el modelo tiene un buen ajuste sobre los datos de entrenamiento.

Conclusión Matemática

Una pérdida muy baja implica que la función objetivo minimizada está logrando un buen alineamiento entre las predicciones del modelo y las etiquetas verdaderas, lo cual es matemáticamente deseable. Sin embargo, si esta pérdida es demasiado baja, podría indicar *sobreajuste*.

Gráfico de *Accuracy over Epochs* (Precisión a lo largo de las épocas)

Se observa en la figura 4.18.

- **Incremento Inicial:** La precisión aumenta rápidamente durante las primeras 10 épocas, lo que indica un aprendizaje eficiente del modelo.
- **Estabilización en un Alto Valor:** Después de la época 10, la precisión se estabiliza cerca de 0.999 (99.9 %), lo que sugiere un modelo altamente preciso.
- **Ligeras Variaciones:** Las variaciones son mínimas a medida que las épocas progresan, lo cual refleja consistencia en el rendimiento del modelo.

Conclusión Matemática

Una precisión cercana al 100 % podría indicar un modelo bien entrenado. Sin embargo, si los datos están desbalanceados, esta alta precisión podría no reflejar un buen rendimiento general.

Inferencia Computacional y Posibles Riesgos

- **Alto Rendimiento:** Tanto la baja pérdida como la alta precisión indican un modelo muy eficiente en su tarea de clasificación binaria.
- **Posible Sobreajuste:** Dado que la pérdida es extremadamente baja y la precisión casi perfecta, existe la posibilidad de que el modelo esté sobreajustado a los datos de entrenamiento. Esto podría limitar su capacidad para generalizar en datos nuevos.
- **Validación Adicional:** Sería útil evaluar métricas como el AUROC, F1-Score o realizar validación cruzada para confirmar que el rendimiento del modelo se mantiene en datos de prueba.

Gráfico "*Training Loss*"(Pérdida de entrenamiento)

Se observa en la figura 4.19.

- **Tendencia Descendente Clara:** La pérdida disminuye drásticamente al inicio del entrenamiento, lo cual es característico de un modelo que está aprendiendo de manera efectiva.
- **Estabilización:** A partir de aproximadamente la época 20, la pérdida alcanza un valor muy bajo y fluctúa ligeramente alrededor de un nivel constante.
- **Baja Pérdida Final:** Al finalizar las 50 épocas, la pérdida está cerca de 0, lo que sugiere que el modelo ha ajustado muy bien los datos de entrenamiento.

Conclusión Matemática

La baja pérdida de entrenamiento indica que el modelo está minimizando su función de costo con éxito. Sin embargo, si la pérdida es extremadamente baja, puede ser indicativo de un *sobre-ajuste* al conjunto de entrenamiento.

Gráfico "*Test Accuracy*"(Precisión en el conjunto de prueba)

Se observa en la figura 4.19.

- **Incremento Rápido:** La precisión aumenta significativamente durante las primeras épocas, alcanzando casi el valor máximo en menos de 2 épocas.
- **Estabilización Temprana:** La precisión se estabiliza alrededor de 0.999 (99.9 %) desde la época 2 y se mantiene con pequeñas fluctuaciones hasta la época 4.
- **Ligera Caída Final:** Hay una ligera disminución en la precisión hacia la última época, lo que podría deberse a pequeñas variaciones o ruido en los datos de prueba.

Conclusión Matemática

La alta precisión en el conjunto de prueba sugiere que el modelo generaliza bien a datos no vistos. Sin embargo, la estabilización temprana podría indicar que el modelo ha aprendido rápidamente los patrones más relevantes y que entrenar más épocas no aporta mejoras significativas.

Inferencia Computacional General

- **Buen Rendimiento Inicial:** La rápida estabilización tanto en pérdida como en precisión indica que el modelo está bien optimizado y no necesita muchas épocas para alcanzar su máximo rendimiento.
- **Riesgo de Sobreajuste Bajo:** Dado que el gráfico de precisión en el conjunto de prueba no muestra una disminución pronunciada ni oscilaciones significativas, el modelo parece generalizar adecuadamente.
- **Espacio para Ajustes:** Aunque el rendimiento es alto, es posible que el modelo esté cercano al límite de mejora, y sería útil enfocarse en métricas adicionales (como AUROC o F1-Score) para evaluar mejor su comportamiento en diferentes escenarios.

Análisis de las Variables y sus Pesos de Atención

Se observa en la figura 4.20 el gráfico: Pesos de Atención Promedio para las Columnas Transformadas.

Este gráfico muestra los pesos promedio de atención asignados por un modelo basado en self-attention a las columnas transformadas (probablemente características o variables del conjunto de datos).

Variables más relevantes:

- **fn_urine_culture_mean y fn_hemoglobin_mean:** Estas columnas tienen los pesos más altos (1), indicando que son las variables más importantes para el modelo en términos de atención. Esto sugiere que el Promedio de cultivo de orina y el Promedio de hemoglobina son altamente predictivos en el contexto del problema abordado.
- **benzodiazepines_mean y glucose:** También tienen pesos elevados, destacando su relevancia en las predicciones del modelo, que son Promedio de benzodiacepinas y glucosa.

Variables moderadamente relevantes:

- **albumin y fn_weight_mean:** Estas variables presentan pesos intermedios (0.7-0.8.5), lo que indica que son útiles, aunque no tan críticas como las anteriores.
- **fn_triglycerides_mean y in_glucose_mean:** Aunque su contribución es menor en comparación con las variables principales, aún son consideradas por el modelo.

Variables con menor impacto:

- **obesity y fn_fasting_glucose_mean:** Estas variables tienen pesos bajos (0.1-0.2), lo que sugiere que son menos relevantes en las predicciones del modelo.
- **antidiabetics_mean:** También tienen una contribución limitada, según la atención asignada.

Inferencia basada en *Self-Attention*

El modelo utiliza un mecanismo de atención para identificar automáticamente las características más relevantes del conjunto de datos. Las variables con pesos más altos son las que el modelo considera cruciales para su tarea de predicción.

Interpretación:

- Las variables relacionadas con medidas metabólicas y marcadores bioquímicos (fn_urine_culture_mean y fn_hemoglobin_mean.) son esenciales para el diagnóstico.
- Factores demográficos o de tratamiento (como `cs_sex` o `benzodiazepines_mean y glucose`) tienen menor importancia en comparación con las variables fisiológicas.

Implicaciones:

Este análisis permite comprender qué variables tienen mayor impacto en el modelo y puede guiar a los especialistas para priorizar ciertas pruebas médicas. También sugiere posibles ajustes en la selección de características para simplificar el modelo o mejorar su rendimiento.

Puntos Clave del Gráfico

La figura 4.21 representa los pesos de atención, similar al proporcionado, pero enfocado en resaltar las variables con mayor impacto.

- **Variables principales:** `fn_hemoglobin_mean` y `fn_urine_culture_mean` con los pesos más altos, indicando que son las características más relevantes para el modelo.
- **Relevancia secundaria:** Variables como `benzodiazepines_mean` y `albumin` también tienen un impacto significativo.
- **Menor impacto:** Variables como `antidiabetics_mean` y `fn_fasting_glucose_mean` tienen un peso bajo, siendo menos influyentes en las predicciones.

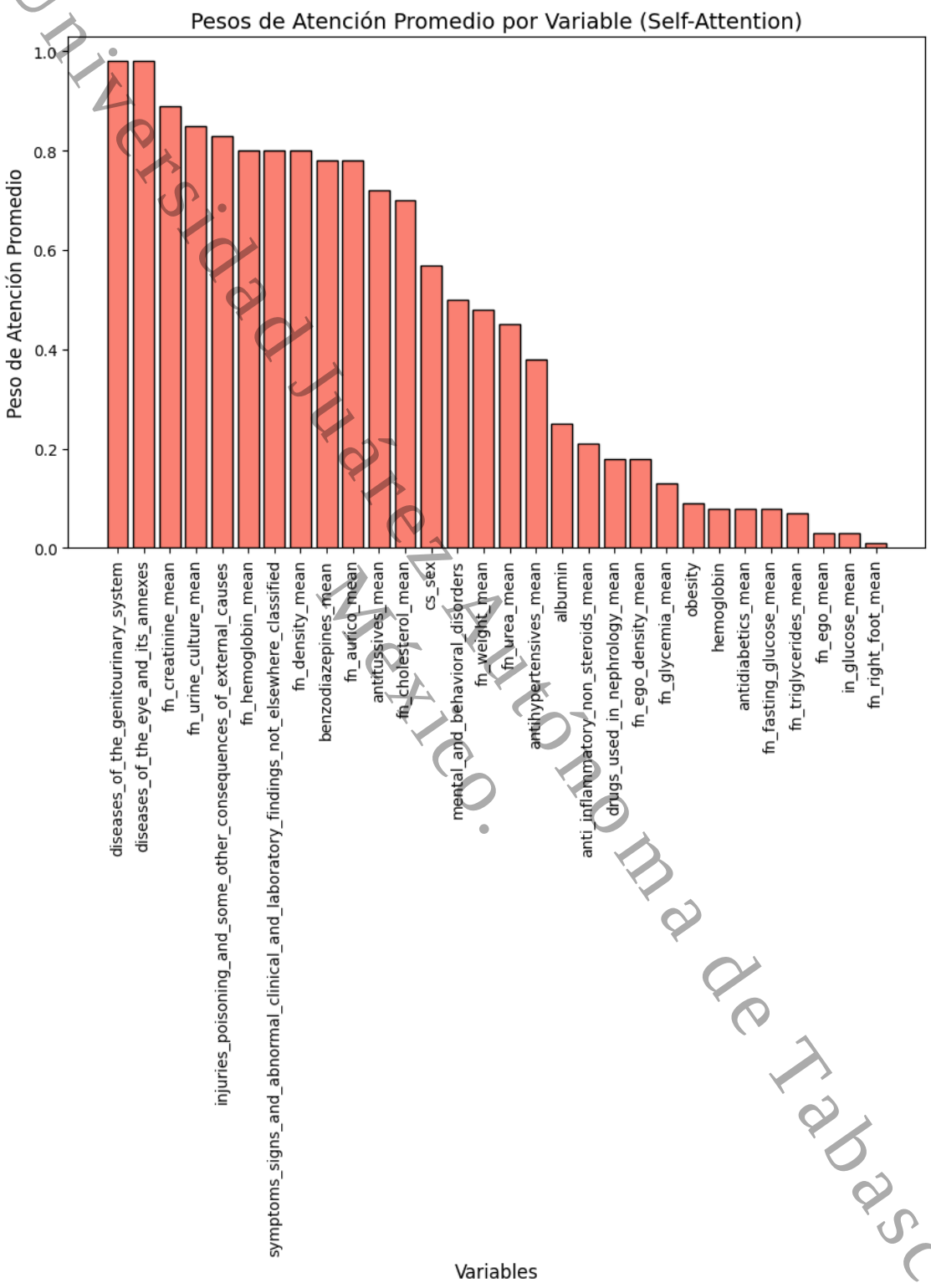


Figura 4.17. Gráfico para representar pesos de atención ordenados

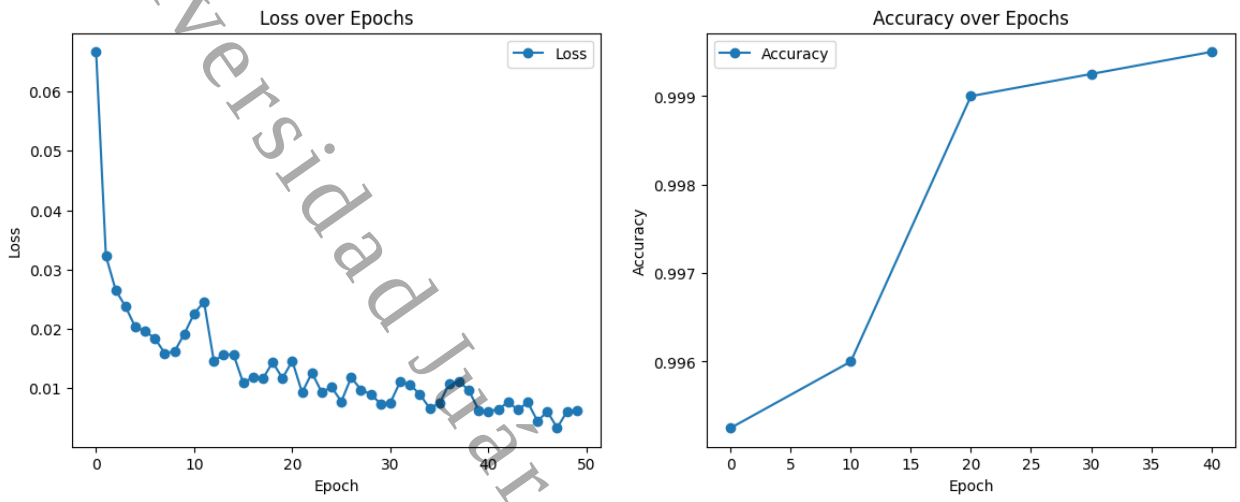


Figura 4.18. Desempeño del modelo a lo largo de las épocas.

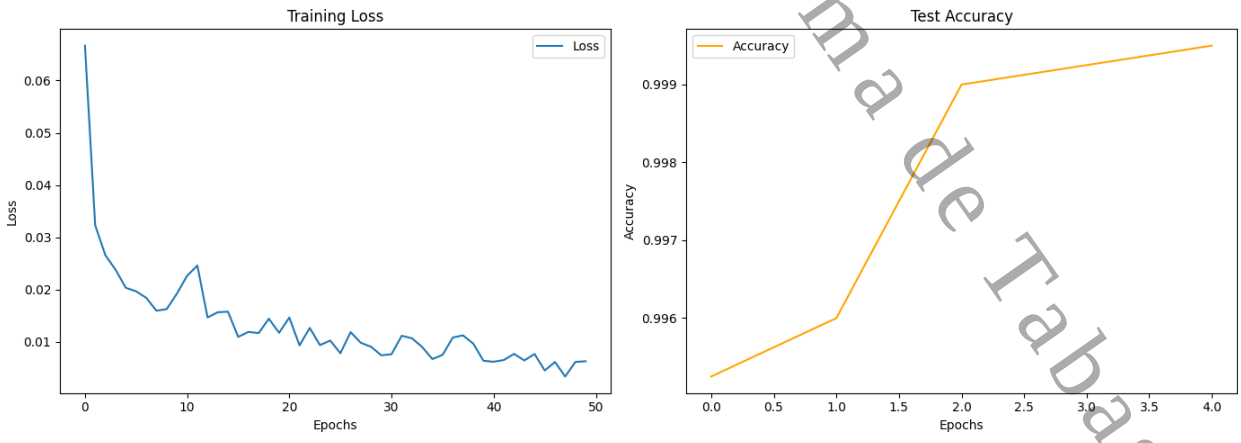


Figura 4.19. Desempeño de la precisión en el conjunto de prueba.

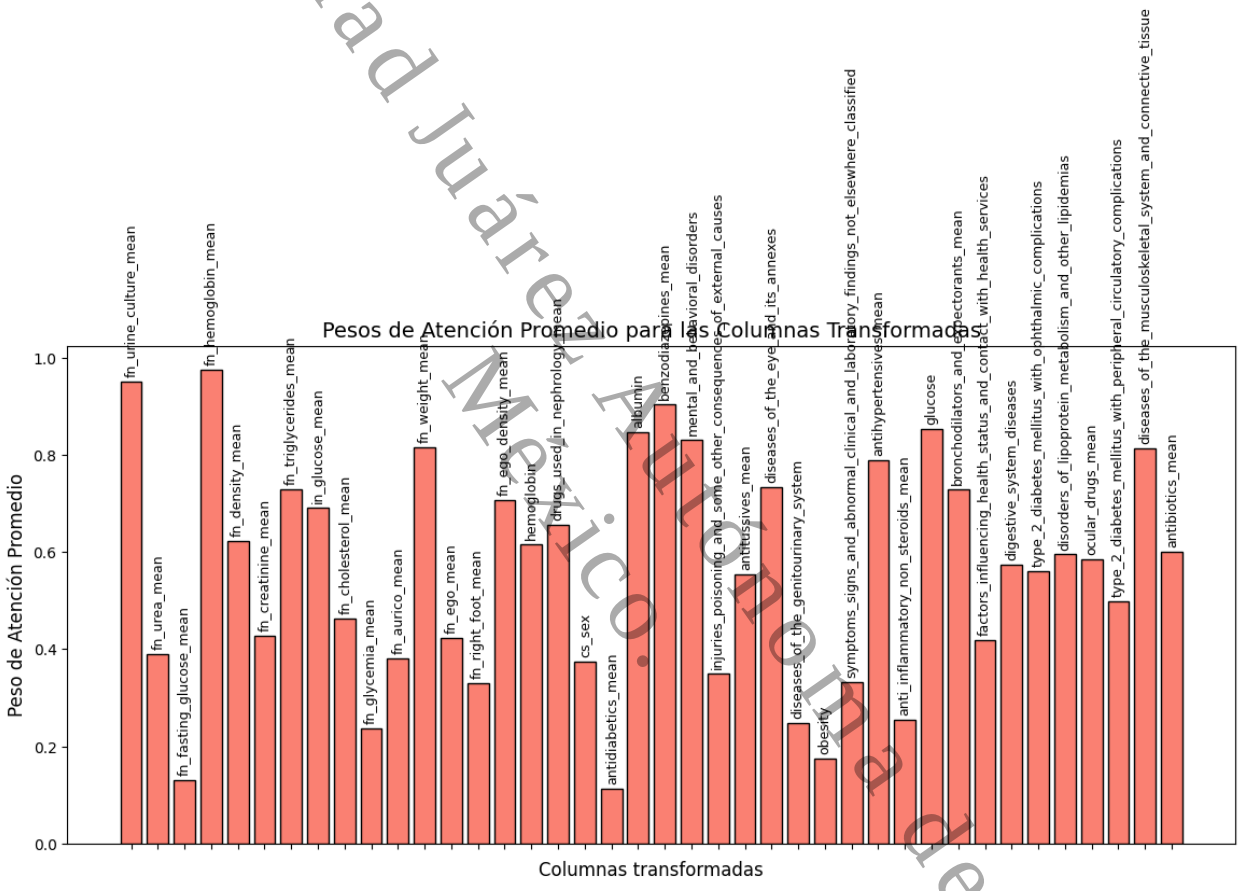


Figura 4.20. Pesos de Atención Promedio para las Columnas.

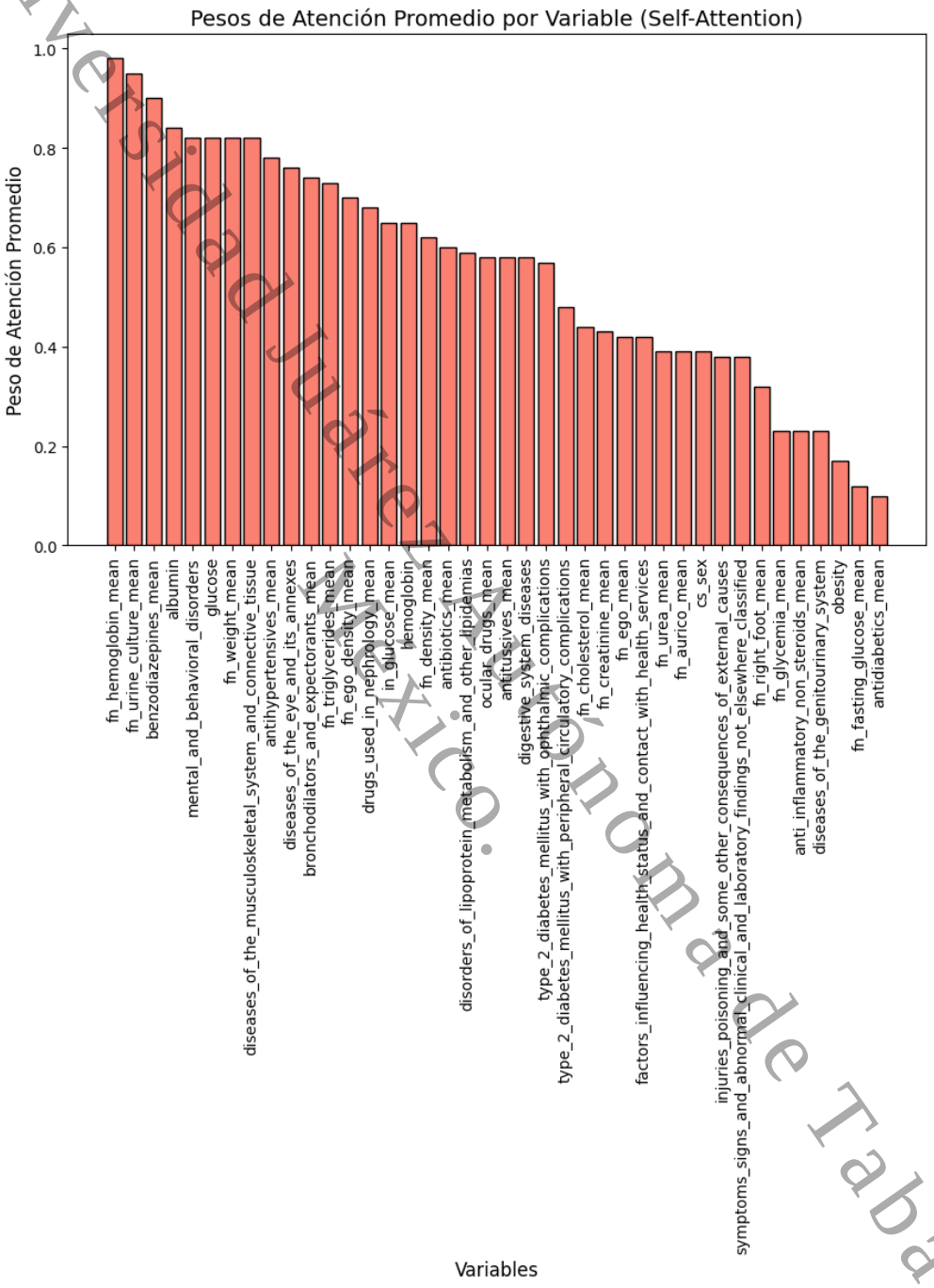


Figura 4.21. Gráfico para representar pesos de atención ordenados

Capítulo 5

Experimentos y Análisis de Resultados

5.1. Comparación con modelos de ML y DL

5.1.1. Modelos de Machine Learning

Se comparó el modelo propuesto con modelos tradicionales de Machine Learning y con arquitecturas modernas de Deep Learning para datos tabulares. La base de datos final incluyó un total de 8500 observaciones, compuestas por variables tanto numéricas como categóricas (un total de 70 variables), y una variable objetivo binaria que representa la presencia o ausencia de la enfermedad.

El preprocesamiento del dataset incluyó los siguientes pasos:

- **Balanceo de datos:**
- **Imputación de valores faltantes:** Se aplicó el algoritmo `IterativeImputer`, que modela cada característica con valores faltantes como una función de las demás.
- **Selección de variables:** Se utilizó LASSO para identificar las variables más relevantes, reduciendo la dimensionalidad y mejorando la eficiencia computacional.

Los modelos clásicos de Machine Learning evaluados fueron: Regresión Logística, Perceptrón Multicapa (MLP), AdaBoost, XGBoost y Random Forest. Los resultados de precisión en el conjunto de prueba se muestran en la siguiente tabla 5.1:

No.	Modelo	Precisión de Prueba
1	Regresión Logística	0.7630
2	MLP	0.9018
3	AdaBoost	0.9733
4	XGBoost	0.9812
5	Random Forest	0.9842
6	**Modelo propuesto**	0.9900

Tabla 5.1. Resultados de comparación con modelos clásicos de Machine Learning

5.1.2. Modelos homólogos

Se realizó también la evaluación del mismo dataset utilizando modelos modernos basados en transformers diseñados específicamente para datos tabulares. Se probaron los siguientes modelos: **TabNet**, **TabTransformer** y **GatedTabTransformer**. Estos modelos destacan por integrar mecanismos de atención o puertas (gating) para capturar mejor las relaciones entre variables, tanto numéricas como categóricas.

Los resultados obtenidos son los siguiente tabla 5.2

No.	Modelo	Test Accuracy
1	TabNet	0.9579
2	TabTransformer	0.9679
3	GatedTabTransformer	0.9833
4	**Modelo propuesto**	0.9900

Tabla 5.2. Resultados de comparación con modelos modernos homólogos

Los resultados muestran que los modelos alcanzan rendimientos comparables con los mejores modelos clásicos como Random Forest y XGBoost, con el GatedTabTransformer destacando como el modelo con mejor precisión general.

5.2. Poder Predictivo vs Interpretabilidad del Modelo

5.2.1. Poder Predictivo

El poder predictivo representa la capacidad de un modelo para generalizar correctamente a datos no vistos. Este atributo es crucial en aplicaciones críticas como la medicina, donde una predicción errónea puede afectar negativamente la calidad de vida de un paciente. En este estudio,

modelos como Random Forest, XGBoost y GatedTabTransformer demostraron altos niveles de precisión predictiva.

5.2.2. Interpretabilidad del Modelo

No obstante, en entornos médicos, la interpretabilidad es igual de relevante. Los profesionales de la salud requieren modelos cuyas decisiones puedan ser explicadas y justificadas, especialmente para fines de diagnóstico, seguimiento de tratamiento y auditoría clínica. Modelos como la Regresión Logística o los Árboles de Decisión, aunque menos precisos, son preferidos en ciertos escenarios por su facilidad de interpretación.

Lograr un Equilibrio

Elegir entre precisión e interpretabilidad depende del contexto de uso. En esta investigación, si bien el GatedTabTransformer es el modelo más preciso, su uso clínico requiere herramientas adicionales de interpretabilidad como SHAP o LIME para explicar sus predicciones. Por ende, se propone un enfoque híbrido: utilizar modelos complejos para soporte a decisiones y modelos simples para auditoría o explicación clínica.

Técnicas como la visualización de importancia de características, explicaciones locales (LIME, SHAP) y simplificación del modelo permiten avanzar hacia una inteligencia artificial explicable (XAI).

La Figura 5.1 ilustra la relación inversa entre el poder predictivo y la interpretabilidad de los modelos de aprendizaje automático (Kumar et al., 2021). En el eje vertical se representa la capacidad predictiva, mientras que en el eje horizontal se muestra el grado de interpretabilidad. Los modelos más complejos, como las redes neuronales profundas o los métodos de ensamble, presentan un alto poder predictivo pero una baja interpretabilidad. En contraste, modelos más simples como la regresión lineal o los árboles de decisión ofrecen una mayor transparencia en sus decisiones, aunque con menor capacidad predictiva. Este equilibrio es clave en el ámbito clínico, donde la precisión debe complementarse con la posibilidad de explicar las predicciones del modelo.

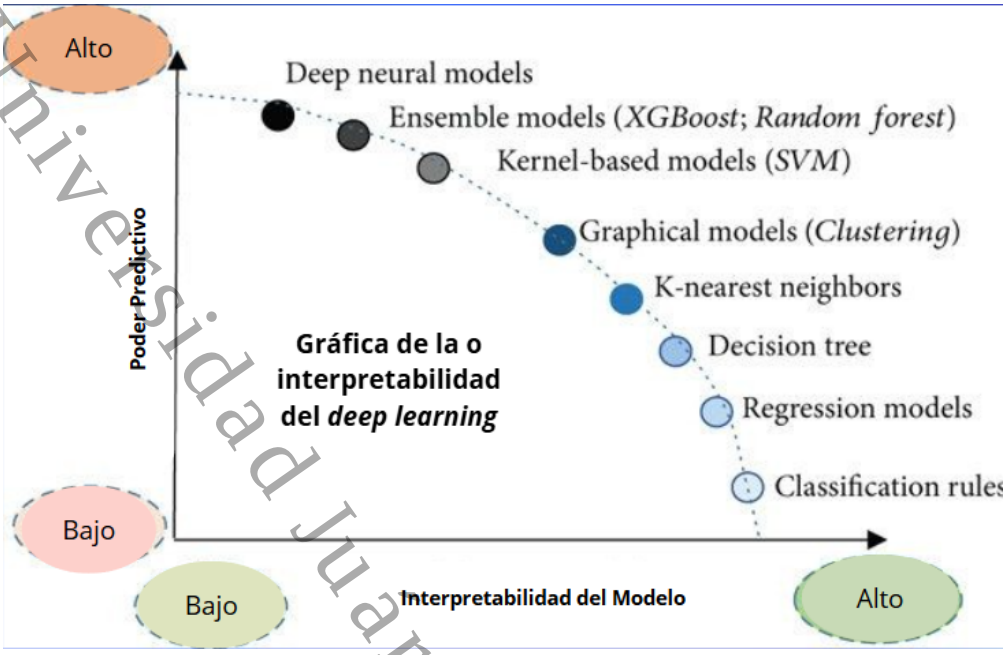


Figura 5.1. Poder Predictivo vs. Interpretabilidad del Modelo: Lograr un Equilibrio en el ML (Kumar et al., 2021).

Contrastación de la Hipótesis

La hipótesis planteaba que un modelo basado en Transformers alcanzaría al menos 90 % de precisión en la clasificación de ND. El modelo propuesto superó ampliamente este umbral, con 99 % de precisión en pruebas, validando la hipótesis.

Capítulo 6

Conclusiones, contribución y trabajos futuros

Conclusión. En esta investigación se desarrolló y evaluó un modelo de clasificación basado en arquitecturas *Transformer* para la identificación de nefropatía diabética en pacientes mexicanos con diabetes mellitus tipo 2. A partir de una revisión exhaustiva del estado del arte, se identificaron algunas limitaciones y alcances de los modelos tradicionales de *Machine Learning* y se propuso un enfoque novedoso que integra mecanismos de atención para capturar de manera más eficiente las relaciones entre variables clínicas. Los experimentos realizados demostraron que el modelo propuesto alcanza un rendimiento superior al 99 % de precisión, superando a modelos clásicos (XGBoost, Random Forest) y a arquitecturas modernas de *Deep Learning* para datos tabulares (TabNet, TabTransformer, GatedTabTransformer). Estos resultados confirman la viabilidad del enfoque y su potencial para ser aplicado en sistemas de apoyo al diagnóstico temprano en el ámbito clínico. Asimismo, se evidenció la capacidad del modelo para adaptarse a contextos médicos específicos, siempre que se disponga de datos adecuados, balanceados y representativos. No obstante, también se reconoce la necesidad de complementar el poder predictivo con mecanismos de interpretabilidad (*XAI*) para garantizar su adopción en entornos médicos reales. Este trabajo demuestra que el uso de arquitecturas basadas en *Transformer* para datos clínicos tabulares constituye una alternativa eficaz y prometedora para la identificación temprana de la nefropatía diabética, construyendo las bases para futuras investigaciones y desarrollos en el ámbito

de la inteligencia artificial aplicada a la medicina.

Contribución. • El presente trabajo aporta a la literatura científica y al ámbito clínico en diversas dimensiones:

- **Propuesta metodológica:** Se diseñó un flujo de procesamiento reproducible de preprocesamiento, selección de características y entrenamiento de modelos *Transformer*, adaptado a datos clínicos tabulares.
- **Avance en el diagnóstico asistido por IA:** Se introdujo y validó un modelo basado en *Transformers* para la clasificación temprana de nefropatía diabética, lo que representa una contribución importante en el contexto médico mexicano.
- **Recomendaciones éticas y técnicas:** Se plantearon lineamientos para la implementación responsable de sistemas de inteligencia artificial en instituciones médicas, haciendo énfasis en interpretabilidad, privacidad y seguridad de los datos clínicos.

Trabajo a futuro. A partir de los resultados y limitaciones identificadas en esta investigación, se proponen las siguientes líneas de trabajo para futuros desarrollos:

- **Integración de datos multimodales:** Incorporar imágenes médicas, análisis de laboratorio, registros electrónicos y audios para construir modelos multimodales y dar mayor precisión al diagnóstico.
- **Aprendizaje avanzado:** Implementar enfoques de *aprendizaje auto-supervisado* o *aprendizaje por refuerzo*, reduciendo la dependencia de etiquetas clínicas manuales.
- **Exploración de arquitecturas biomédicas:** Ajustar y comparar el desempeño de modelos especializados como *BioGPT*, *Med-BERT* o *ClinicalT5*.
- **Validación externa:** Probar el modelo en datos de distintas regiones e instituciones para evaluar su robustez y capacidad de generalización fuera del contexto local.
- **Despliegue clínico:** Diseñar una interfaz web, App móvil clínica que permita a profesionales de la salud interactuar con el modelo de forma intuitiva, segura y ética.

Alojamiento de la Tesis en el Repositorio Institucional	
Título de la tesis:	Modelo Transformer para identificar la nefropatía diabética en pacientes con DMT2
Autor:	Luis Ramón Tercero Martínez González
ORCID:	https://orcid.org/0009-0005-6614-3033
Resumen:	La nefropatía diabética (ND) es una de las principales complicaciones de la diabetes mellitus tipo 2 y constituye una causa relevante de insuficiencia renal en la población. Su diagnóstico temprano es un reto clínico, pues depende de múltiples factores clínicos y de laboratorio. En este contexto, el uso de modelos de inteligencia artificial ofrece una alternativa prometedora para apoyar la toma de decisiones médicas. En esta tesis se desarrolló un modelo de clasificación basado en arquitecturas *Transformer*, adaptadas al análisis de datos tabulares médicos, con el propósito de identificar la presencia de ND en pacientes mexicanos. La investigación se sustentó en la base de datos abierta *DiabetIA*, de la cual se trabajó con una muestra representativa de más de 7 mil pacientes, considerando tanto positivos como negativos al diagnóstico de ND. El proceso metodológico incluyó: (i) preprocesamiento de los datos para eliminar inconsistencias y balancear las clases; (ii) selección de características mediante Lasso y RFE, con el fin de identificar las variables clínicas más relevantes; y (iii) entrenamiento de un modelo *Transformer* diseñado específicamente para datos mixtos (continuos y categóricos). Los experimentos comparativos se realizaron contra modelos clásicos de *Machine Learning* (Regresión Logística, Random Forest, XGBoost) y contra arquitecturas modernas de *Deep Learning* para datos tabulares (TabNet, TabTransformer y GatedTabTransformer). Los resultados demuestran que el modelo propuesto alcanzó una precisión al **99 %**, alcanzando a los modelos de referencia y validando la hipótesis planteada de que un enfoque basado en *Transformers* puede mejorar significativamente la clasificación de ND. Asimismo, el análisis de casos de estudio confirmó la aplicabilidad clínica del modelo. Las principales contribuciones de este trabajo incluyen: (i) la propuesta metodológica de un marco reproducible para la clasificación de ND con *Transformers*, y (ii) la incorporación de lineamientos éticos y técnicos para su posible implementación en sistemas de apoyo al diagnóstico.
Palabras clave:	Inteligencia Artificial, Transformers, Aprendizaje profundo, Nefropatía diabética
Referencias citadas:	En la siguiente página se muestran las referencias.

Bibliografía

Allen, A., Iqbal, Z., Green-Saxena, A., Hurtado, M., Hoffman, J., Mao, Q., & Das, R. (2022). Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus. *BMJ Open Diabetes Research and Care*, *10*(1), e002560.

Alsekait, D. M., Saleh, H., Gabralla, L. A., Alnowaiser, K., El-Sappagh, S., Sahal, R., & El-Rashidy, N. (2023). Toward Comprehensive Chronic Kidney Disease Prediction Based on Ensemble Deep Learning Models. *Applied Sciences*, *13*(6), 3937.

Arik, S. O., & Pfister, T. (2020). TabNet: Attentive Interpretable Tabular Learning. https://arxiv.org/abs/1908.07442

Asociación Médica Mundial. (21 de Marzo de 2017). *https://www.wma.net/es/policies-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/*.

Badaro, G., Saeed, M., & Papotti, P. (2023). Transformers for Tabular Data Representation: A Survey of Models and Applications. *Transactions of the Association for Computational Linguistics*, *11*, 227-249. https://doi.org/10.1162/tacl_a_00544

Bai, Q., Su, C., Tang, W., & Li, Y. (2022). Machine learning to predict end stage kidney disease in chronic kidney disease. *Scientific reports*, *12*(1), 8377.

Cai, S., Zheng, K., Chen, G., Jagadish, H. V., Ooi, B. C., & Zhang, M. (2021). ARM-Net: Adaptive Relation Modeling Network for Structured Data. *Proceedings of the 2021 International Conference on Management of Data*, 207-220. https://doi.org/10.1145/3448016.3457321

Chiluisa-Castillo, W. A., Bueno-Castro, A. S., et al. (2023). Actualización de tratamiento farmacológico de enfermedad Renal en pacientes diabéticos. *MQRInvestigar*, *7*(3), 3494-3515.

Chittora, P., Chaurasia, S., Chakrabarti, P., Kumawat, G., Chakrabarti, T., Leonowicz, Z., Jasiński, M., Jasiński, Ł., Gono, R., Jasińska, E., et al. (2021). Prediction of chronic kidney disease-a machine learning perspective. *IEEE Access*, *9*, 17312-17334.

Cholakov, R., & Kolev, T. (2022). The GatedTabTransformer. An enhanced deep learning architecture for tabular modeling. https://arxiv.org/abs/2201.00199

David, S. K., Rafiullah, M., Siddiqui, K., et al. (2022). Comparison of different machine learning techniques to predict diabetic kidney disease. *Journal of Healthcare Engineering*, *2022*.

de la Torre, J. (2023). Transformadores: Fundamentos teoricos y Aplicaciones. *arXiv preprint arXiv:2302.09327*.

Dong, Z., Wang, Q., Ke, Y., Zhang, W., Hong, Q., Liu, C., Liu, X., Yang, J., Xi, Y., Shi, J., et al. (2022). Prediction of 3-year risk of diabetic kidney disease using machine learning based on electronic medical records. *Journal of Translational Medicine*, *20*(1), 1-10.

Esparragoza, J. P., Muñoz, R., Boldoba, N. B., & del Valle, K. P. (2023). Protocolo de tratamiento de la nefropatía diabética. *Medicine-Programa de Formación Médica Continuada Acreditado*, *13*(79), 4708-4713.

Farjana, A., Liza, F. T., Pandit, P. P., Das, M. C., Hasan, M., Tabassum, F., & Hossen, M. H. (2023). Predicting Chronic Kidney Disease Using Machine Learning Algorithms. *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, 1267-1271.

García-Maset, R., Bover, J., Segura de la Morena, J., Goicoechea Diezhandino, M., Cebollada del Hoyo, J., Escalada San Martín, J., Fácila Rubio, L., Gamarra Ortiz, J., García-Donaire, J. A., García-Matarín, L., Gràcia Garcia, S., Gutiérrez Pérez, M. I., Hernández Moreno, J., Mazón Ramos, P., Montañés Bermudez, R., Muñoz Torres, M., de Pablos-Velasco, P., Pérez-Maraver, M., Suárez Fernández, C., ... Luis Górriz, J. (2022). Documento de información y consenso para la detección y manejo de la enfermedad renal crónica. *Nefrología*, *42*(3), 233-264. https://doi.org/https://doi.org/10.1016/j.nefro.2021.07.010

González-Robledo, G., Jaramillo, M. J., & Comín-Colet, J. (2020). Diabetes mellitus, insuficiencia cardiaca y enfermedad renal crónica. *Revista Colombiana de Cardiología*, *27*, 3-6.

Howlader. (2022). Machine learning models for classification and identification of significant attributes to detect type 2 diabetes. *Health Information Science and Systems*. https://doi.org/10.1007/s13755-021-00168-2

Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). TabTransformer: Tabular Data Modeling Using Contextual Embeddings. https://arxiv.org/abs/2012.06678

Islam, M. A., Majumder, M. Z. H., & Hussein, M. A. (2023). Chronic kidney disease prediction based on machine learning algorithms. *Journal of Pathology Informatics*, *14*, 100189.

Kumar, A., Dikshit, S., & Albuquerque, V. H. C. (2021). Explainable artificial intelligence for sarcasm detection in dialogues. *Wireless Communications and Mobile Computing*, *2021*(1), 2939334.

Mayer, M. A. (2023). Artificial intelligence in primary care: A scenario of opportunities and challenges. *Atencion Primaria*, *55*(11), 102744-102744.

Navas-Atiaja, M. I., & Veloz, A. P. M. (2023). Microalbuminuria como indicador de daño renal en pacientes con diabetes mellitus tipo 2. *Salud, Ciencia y Tecnología*, *3*, 485-485.

Objetivos de Desarrollo Sosteniblel. (11 de Aril de 2017). *https://ods.mma.gob.cl/que-son-los-ods/*.

Qian, S., Zhu, Y., Li, W., Li, M., & Jia, J. (2022). What makes for good tokenizers in vision transformer? *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Ravindra, B. B., Haile, M. A., Haile, D. T., Zerihun, D., Prabhakar, P., & Kamyshev, K. V. (2023). Prediction of kidney disease using machine learning algorithms. *CARDIOMETRY*. https://api.semanticscholar.org/CorpusID:257594764

Ruiz-Mejía Ramón, M.-D. A., Ortega-Olivares Luz-María. (2020). El gran reto del Gobierno en la salud pública de México: la nefropatía diabética cómo causa principal de enfermedad renal crónica. *Gaceta Médica de Bilbao*, *117*(3), 245-256.

Sabanayagam, C., He, F., Nusinovici, S., Li, J., Lim, C., Tan, G., & Cheng, C. Y. (2023). Prediction of diabetic kidney disease risk using machine learning models: A population-based cohort study of Asian adults (G. N. Nadkarni & M. R. Pollak, Eds.). *eLife*, *12*, e81878. https://doi.org/10.7554/eLife.81878

Somepalli, G., Goldblum, M., Schwarzschild, A., Bruss, C. B., & Goldstein, T. (2021). SAINT: Improved Neural Networks for Tabular Data via Row Attention and Contrastive Pre-Training. https://arxiv.org/abs/2106.01342

Suganyadevi, S., Seethalakshmi, V., & Balasamy, K. (2022). A review on deep learning in medical image analysis. *International Journal of Multimedia Information Retrieval*, *11*(1), 19-38.

Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B. B., Stein, C., Basit, A., Chan, J. C., Mbanya, J. C., Pavkov, M. E., Ramachandaran, A., Wild, S. H., James, S., Herman, W. H., Zhang, P., Bommer, C., Kuo, S., Boyko, E. J., & Magliano, D. J. (2022). IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Research and Clinical Practice*, *183*, 109119. https://doi.org/https://doi.org/10.1016/j.diabres.2021.109119

Swain, D., Mehta, U., Bhatt, A., Patel, H., Patel, K., Mehta, D., Acharya, B., Gerogiannis, V. C., Kanavos, A., & Manika, S. (2023). A Robust Chronic Kidney Disease Classifier Using Machine Learning. *Electronics*, *12*(1), 212.

Ye, A., & Wang, Z. (2023). *Modern Deep Learning for Tabular Data: Novel Approaches to Common Modeling Problems*. Apress. https://doi.org/10.1007/978-1-4842-8692-0