



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**ELABORACIÓN DE UN MODELO DE MACHINE LEARNING PARA
PREDECIR LA PROBABILIDAD DE EMBARAZO CON TRATAMIENTOS
DE ALTA COMPLEJIDAD PARA PACIENTES DEL INSTITUTO DE
INVESTIGACIONES MATERNO INFANTIL (IDIMI)**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

EDUARDO RAÚL PÉREZ LIZAMA

PROFESOR GUÍA:
ROCÍO RUIZ MORENO

MIEMBROS DE LA COMISIÓN:
ALEJANDRA PUENTE CHANDÍA
FELIPE VERA CID

SANTIAGO DE CHILE
2024

ELABORACIÓN DE UN MODELO DE MACHINE LEARNING PARA PREDECIR LA PROBABILIDAD DE EMBARAZO CON TRATAMIENTOS DE ALTA COMPLEJIDAD PARA PACIENTES DEL INSTITUTO DE INVESTIGACIONES MATERNO INFANTIL (IDIMI)

En un escenario marcado por el aumento significativo de casos de infertilidad a nivel mundial, con una prevalencia de infertilidad a lo largo de la vida estimada en un 17.5 % a nivel global y, en particular en Chile, la fertilización asistida ha emergido como una opción crucial. Sin embargo, el proceso actual de derivación y clasificación de tratamientos en nuestro país y, en específico, en el Instituto de Investigaciones Materno Infantil (IDIMI) y el Hospital Clínico San Borja Arriarán presenta desafíos como derivaciones tardías o inadecuadas, generando gastos significativos y disminuyendo la tasa de éxito de los tratamientos.

En este contexto, el objetivo general de este proyecto es determinar la probabilidad de éxito de un tratamiento de alta complejidad mediante el desarrollo de modelos de machine learning que utilicen parámetros clínicos y marcadores bioquímicos relacionados con la paciente, con el propósito de lograr resultados más efectivos en los tratamientos de fertilización asistida para pacientes de IDIMI. La metodología incluyó la revisión del estado del arte, reuniones con profesionales de la salud, análisis de una base de datos histórica de pacientes sometidos a tratamientos entre los años 2010 al 2019 en 1609 pacientes, y la aplicación de la metodología KDD para la preparación de datos y desarrollo de modelos. Se utilizaron algoritmos de inteligencia artificial como Random Forest, Super Vector Machine y XGBoost.

La interpretación y evaluación de datos se llevó a cabo mediante la aplicación de SHAP. Además, se realizó un análisis del impacto social y económico revelando los costos de ciclos fallidos en los tratamientos y el impacto de la solución propuesta en el sistema de salud. Independientemente de las métricas, se ofrecieron recomendaciones para futuras iteraciones del modelo.

El modelo de Red Neuronal con validación cruzada mostró el mejor desempeño, con un AUC ROC de 0.82 y un F1 Score de 0.75. Random Forest, por otro lado, destacó entre los modelos no neuronales con un AUC ROC de 0.75 y un F1 Score de 0.68, reflejando un equilibrio entre precisión y capacidad de interpretación.

Entre las variables más importantes para la predicción del éxito del tratamiento, la Tasa de Blastulación fue la más influyente, seguida por el Recuento de Folículos Antrales (AFC) y la Edad de la paciente, donde un aumento en la edad se asocia a una disminución en la probabilidad de éxito. Además, altos niveles de Hormona Antimülleriana (AMH) y el número de embriones transferidos también mostraron relaciones significativas con los resultados.

La implementación de estos modelos tiene el potencial de reducir los ciclos fallidos entre un 10 % y un 15 %, lo que se traduciría en ahorros estimados de entre 29 y 57 millones de pesos chilenos al año. El uso de estos modelos también fortalecerá la capacidad del personal médico para tomar decisiones informadas, maximizando la probabilidad de éxito en los tratamientos.

Tabla de Contenido

1. Introducción	1
1.1. Antecedentes generales	1
1.2. Apoyos institucionales	2
1.2.1. Web Intelligence Centre (WIC)	2
1.2.2. Instituto de Investigaciones Materno Infantil (IDIMI)	2
1.3. Planteamiento del problema y justificación	3
1.4. Hipótesis de investigación	7
1.5. Objetivos	7
1.5.1. Objetivo general	7
1.5.2. Objetivos específicos	7
1.6. Metodología	7
1.7. Alcances	8
1.8. Resultados esperados	9
2. Marco conceptual	10
2.1. Conceptos médicos	10
2.1.1. Infertilidad	10
2.1.2. Fertilización in vitro	10
2.2. Descubrimiento de conocimiento en bases de datos - Metodología KDD	11
2.3. Conceptos relevantes para la evaluación de desempeño de modelos	13
2.3.1. Matriz de confusión	13
2.3.2. Métricas clave	13
2.3.3. Machine Learning	14
2.3.4. Validación cruzada	15
2.3.5. Método SHAP y valores Shapley	16
3. Estado del arte	17
3.1. Diversos enfoques para identificar las variables más relevantes en los tratamientos y/o predecir el éxito del embarazo	18
4. Desarrollo	23
4.1. Los datos	23
4.2. Preprocesamiento de los datos	23
4.2.1. Unificación bases de datos	23
4.2.2. Tratamiento de datos faltantes	24
4.2.3. Transformación de datos	25
4.2.3.1. Distintos diagnósticos	25
4.2.3.2. Distintas inducciones de ovulación	26

4.3.	Análisis exploratorio de los datos	27
4.3.1.	Análisis de correlaciones	27
4.3.2.	Análisis grafico de distribuciones	30
4.4.	Desarrollo de modelos	37
4.4.1.	Datos a utilizar	37
4.4.2.	Resultados	38
4.4.2.1.	Random Forest (mejor modelo)	39
4.4.2.2.	Redes neuronales	41
5.	Discusión	43
5.1.	Interpretación de los resultados generales	43
5.2.	Importancia de las variables predictoras	43
5.3.	Limitaciones y consideraciones	44
6.	Impacto social y económico	46
6.1.	Impacto económico	46
6.2.	Impacto social	48
7.	Conclusiones	49
7.1.	Desarrollo metodológico	49
7.1.1.	Importancia de las variables predictoras	49
7.1.2.	Validación cruzada y su importancia en la predicción de resultados clínicos	49
7.2.	Desempeño de los modelos y su relevancia en el contexto clínico	49
7.3.	Impacto económico y social del uso de modelos predictivos	50
7.4.	Conclusiones del trabajo	50
7.5.	Recomendaciones para iteraciones futuras	51
	Bibliografía	52

Índice de Tablas

2.1.	Matriz de confusión. Fuente: Elaboración propia	13
4.1.	Categorías variable EMBARAZO	25
4.2.	Categorías diagnóstico. Fuente: Elaboración propia.	26
4.3.	Categorías inducción. Fuente: Elaboración propia.	27
4.4.	Variables de caracterización de ovocitos obtenidos. Fuente: Elaboración propia.	29
4.5.	Variables de caracterización de fecundación. Fuente: Elaboración propia. . . .	29
4.6.	Set de variables final. Fuente: Elaboración propia.	38
4.7.	Resultados de los modelos. Fuente: Elaboración propia.	39
4.8.	Mejores Hiperparámetros modelo Random Forest. Fuente: Elaboración propia.	40
6.1.	Financiamiento otorgado por FONASA al IDIMI para tratamientos de fertiliza- ción asistida de alta complejidad en 2019. Fuente: [53].	46

Índice de Ilustraciones

1.1.	Costos directos por un ciclo de ART y PIB per cápita. La figura muestra el costo de un ciclo de ART en USD PPP en relación con el PIB per cápita, también en USD PPP. Fuente: [14].	4
1.2.	Árbol de problemas. Fuente: Elaboración propia.	6
2.1.	Proceso de fertilización in Vitro (FIV). Fuente: [18].	11
2.2.	Proceso KDD. Fuente: Elaboración propia.	12
2.3.	División por registros y división por materias. Fuente: Elaboración propia.	16
4.1.	Mapa de correlación entre variables. Fuente: Elaboración propia.	28
4.2.	Número de muestras por etiqueta de la variable objetivo. Fuente: Elaboración propia.	31
4.3.	Histogramas de cada variable respecto a la variable objetivo. Parte 1	32
4.4.	Histogramas de cada variable respecto a la variable objetivo. Parte 2	33
4.5.	Histogramas de cada variable respecto a la variable objetivo. Parte 3	34
4.6.	Gráficos de barra de cada variable respecto a la variable objetivo. Parte 1	35
4.7.	Gráficos de barra de cada variable respecto a la variable objetivo. Parte 2	36
4.8.	Resultados algoritmo SHAP modelo Random Forest. Fuente: Elaboración propia.	41
6.1.	Índice de precios al consumidor 2019–2023. Fuente: Banco mundial.	47

Capítulo 1

Introducción

1.1. Antecedentes generales

La Organización Mundial de la Salud define la infertilidad como una condición del sistema reproductivo, tanto masculino como femenino, caracterizada por la incapacidad de concebir después de 12 meses o más de relaciones sexuales habituales sin protección. Esta problemática puede derivar de factores masculinos, femeninos o combinados, afectando aproximadamente al 40–45 % de las parejas, distribuido en un 35–40 % para hombres, 40–45 % para mujeres y 20–30 % para ambas partes de la pareja [1]. A nivel global, se estima que entre 48.5 y 72.4 millones de parejas sufren de infertilidad, con tasas de prevalencia más elevadas en regiones como América, el Pacífico occidental, África y algunas áreas de Europa [2][3].

En el contexto chileno, según el Instituto de Investigaciones Materno Infantil (IDIMI), se estima que, de las 3 957 257 mujeres chilenas en edad fértil (entre 15 y 44 años), alrededor de un 10 % podrían enfrentar problemas de infertilidad, de las cuales cerca del 40 % podrían experimentar una infertilidad de larga duración en sus vidas de pareja, traduciéndose en más de 150 mil mujeres afectadas [1]. El aumento de la prevalencia en Chile en las últimas décadas se atribuye a diversos factores, como la disminución de la concentración y calidad de espermatozoides a nivel mundial y el retraso en la maternidad hacia edades avanzadas en las mujeres [4].

En síntesis, la infertilidad se presenta como una problemática de salud pública en Chile, afectando significativamente la tasa de crecimiento poblacional. Para abordar esta situación, se han desarrollado diversas técnicas de fertilidad asistida, disponibles tanto en el sistema público como privado de salud. Estas técnicas se dividen en procedimientos de baja complejidad (BC), como la inducción de ovulación (IO) con o sin inseminación intrauterina (IIU), y de alta complejidad (AC), donde destaca la fertilización in vitro (FIV), que consiste en un proceso en el cual se extraen óvulos de los ovarios y se fertilizan con espermatozoides en un laboratorio. Este procedimiento se denomina ciclo FIV y abarca desde la estimulación ovárica hasta la transferencia de embriones al útero. Solo en el año 2019, IDIMI realizó más de 360 ciclos FIV. En la última década se han llevado a cabo casi 5000 ciclos de alta complejidad.

En Chile, la regulación de los tratamientos de fertilización asistida se encuentra respaldada por un marco legal específico. Desde el 17 de marzo de 2014 las normativas incorporadas al Arancel de la Modalidad de Libre Elección del Fonasa bajo “Pago Asociado a Diagnóstico” (PAD) han influido significativamente en el acceso a los tratamientos de fertilidad [5].

Según estas disposiciones legales, las Isapres están obligadas a contemplar, al menos, las prestaciones y cobertura financiera establecidas por el Fonasa para la modalidad de Libre

Elección. En el sector público, Fonasa cubre los tratamientos de fertilización asistida de alta complejidad bajo condiciones específicas. La mujer beneficiaria de Fonasa (tramo B, C o D) debe tener diagnóstico previo de infertilidad al igual que su pareja si es que es asegurado de Fonasa en los mismos tramos. Para acceder a los Procedimientos Asociados a Diagnóstico es necesario que la entidad de salud elegida tenga la prestación en convenio. Además, se exige que la pareja o beneficiaria haya agotado previamente todas las posibilidades clínicas en tratamientos de fertilidad de baja complejidad, o que por indicación clínica sea candidata directa solo a procedimientos de alta complejidad.

1.2. Apoyos institucionales

1.2.1. Web Intelligence Centre (WIC)

El Web Intelligence Centre es el centro de investigación aplicada a la resolución de problemas que demanden soluciones basadas en tecnologías de información y comunicaciones, ciencia de datos e inteligencia artificial. El WIC, fue fundado en el año 2010 por el profesor Juan Velázquez, quién dirige el centro junto a Rocío Ruiz, directora ejecutiva. En la última década el centro se ha especializado en el desarrollo de soluciones en el ámbito de la salud, que se traducen en la creación de servicios y productos para apoyar la diagnosis y mejorar el bienestar de la población, de los cuales destacan Kefuri, sistema de aviso automático de posibles donantes de órganos; Proyecto VIDA, plataforma informática basada en algoritmos de machine learning e inteligencia artificial para la extracción de patrones desde datos relacionados a perfiles de pacientes con VIH, para caracterizarlos e identificar el grado de adherencia y monitoreo constante del paciente; Proyecto Test Virtual, test de navegación espacial virtual que detecta de forma rápida y precisa el deterioro cognitivo inicial en adultos mayores (AM), mediante un puntaje calculado en base al rendimiento de los pacientes en esta prueba virtual; y Sonama, plataforma de análisis de redes sociales para estudiar la prevalencia de alcohol y marihuana en Chile [6].

El trabajo se realizó con la tutela del WIC, en específico con Rocío Ruiz, en el marco del desarrollo de una plataforma informática para la selección del tratamiento de fertilización asistida de los pacientes del IDIMI, donde la memoria es parte del desarrollo de este proyecto.

1.2.2. Instituto de Investigaciones Materno Infantil (IDIMI)

El Instituto de Investigaciones Materno Infantil es un organismo de la Facultad de Medicina de la Universidad de Chile creado el año 1988 bajo la decanatura del Prof. Dr. Alejandro Goic G. El instituto conjuga la práctica clínica con la investigación en las áreas de Medicina Reproductiva, Crecimiento y Desarrollo, Endocrinología Infantil y Genética. En el área clínica ambulatoria reciben niños y niñas con problemas de crecimiento, desarrollo puberal, diabetes y otras enfermedades hormonales. También atienden a hombres y mujeres con patologías que afectan su capacidad reproductiva. En el área de la investigación, se centran en los problemas de salud que enfrentan sus pacientes y buscan comprender las etiologías detrás de las enfermedades vinculadas al crecimiento, pubertad y la reproducción. La vinculación con la Facultad de Medicina de la Universidad de Chile es una de las características más relevantes del IDIMI, dado que añade a la práctica clínica y atención al paciente su marcado carácter docente e investigador [7].

El presente trabajo se desarrolla junto a este equipo, quienes proveerán el enfoque clínico al trabajo y los datos necesarios para cumplir con los objetivos propuestos provenientes de más de 1600 pacientes en control en el Hospital Clínico San Borja Arriarán e IDIMI desde el año 2015 hasta el 2019.

1.3. Planteamiento del problema y justificación

Desde la creación del Programa Nacional de Fertilización Asistida de Alta Complejidad en 1992 y el Programa de Baja Complejidad en 2005, el Instituto de Investigaciones Materno Infantil (IDIMI), junto con el Hospital Clínico San Borja Arriarán, ha brindado apoyo técnico en el desarrollo de ambos programas y en la formación de centros de medicina reproductiva a lo largo del país. Actualmente, existen aproximadamente 10 centros de Baja Complejidad (BC) y 4 de Alta Complejidad (AC) en Chile. Los centros de baja complejidad se enfocan en procedimientos menos invasivos, como la inducción de ovulación y la inseminación intrauterina (IIU), que requieren menor tecnología y especialización. Por otro lado, los centros de alta complejidad realizan técnicas más avanzadas, como la fertilización in vitro (FIV), un proceso que incluye múltiples etapas: desde la estimulación ovárica y la recuperación de óvulos hasta la fertilización en laboratorio y la transferencia de embriones al útero. Todo este conjunto de etapas constituye lo que se denomina un ciclo FIV para un paciente. Actualmente, el 90 % de los ciclos de FIV en Chile se llevan a cabo en el IDIMI y el Hospital Clínico San Borja Arriarán. En total, los 10 centros de Baja Complejidad y 4 de Alta Complejidad en el país realizan más de 3000 ciclos anualmente.

Entre los años 2010 y 2019, el IDIMI llevó a cabo más de 4000 ciclos de alta complejidad, de los cuales un 70 % fueron realizados con ovocitos frescos. Este proceso involucra varias etapas complejas, como la estimulación ovárica, recuperación de ovocitos, fecundación, cultivo de embriones y transferencia embrionaria al útero. Sin embargo, solo en un 40 % de estos ciclos se logró el embarazo. Las fallas en el resto de los ciclos se debieron a diversos factores, como la falla en la fecundación, la no recuperación de ovocitos o fallas en la estructura del endometrio. Estos factores no solo afectan las tasas de éxito, sino que también representan una carga emocional y económica significativa para las parejas involucradas.

La infertilidad y los tratamientos asociados tienen un impacto profundo y multifacético en el bienestar psicosocial de las personas y parejas afectadas. Este impacto abarca los efectos mentales, emocionales, sociales y espirituales que surgen como resultado de la infertilidad [8]. El estrés emocional es una experiencia común, caracterizada por sentimientos de tristeza, desesperación y frustración ante la incapacidad de concebir, lo que genera una carga psicológica abrumadora [9]. Esta situación se ve exacerbada por la falta de comprensión y empatía en el entorno social, lo que puede llevar al aislamiento y a sentimientos de soledad. A menudo, las mujeres son las protagonistas de este proceso. Aunque la problemática de fertilidad puede residir en la mujer o en su pareja, son ellas quienes frecuentemente se someten a procedimientos médicos invasivos, monitorean sus ciclos menstruales y ajustan sus rutinas para seguir estrictamente los tratamientos [10]. La presión por obtener resultados positivos, combinada con tratamientos tardíos o ineficaces, puede deteriorar aún más su bienestar.

La presión financiera sobre los pacientes es considerable, ya que los tratamientos de fertilidad son costosos y no siempre están cubiertos por seguros médicos [5]. En Chile, existen restricciones directas para acceder a ayudas gubernamentales a través de Fonasa, permitiendo solo una oportunidad de acceso, excluyendo a mujeres solteras y sin incluir la criopreservación

embrionaria. Este último servicio puede alcanzar los 2 millones de pesos en el sector privado si se desea criopreservar más de seis embriones [11]. Hasta el 2019, solo un 10 % del total de ciclos en el país se realizaron bajo Fonasa [11]. La pandemia de COVID-19 exacerbó esta situación, con la atención de fertilidad relegada en el sistema público. Además, el riesgo de infección planteó nuevas preguntas sobre la seguridad de los procedimientos [12].

El costo de los tratamientos de fertilidad y de las tecnologías de reproducción asistida (TRA) es un desafío en países desarrollados y en desarrollo. En Canadá, por ejemplo, los tratamientos de TRA pueden representar hasta el 25 % del gasto promedio anual de un hogar, mientras que en EE.UU. llegan a consumir el 50 % del ingreso anual disponible de una persona [13]. En un estudio publicado en 2023, se evidenció que países con un PIB per cápita más alto no necesariamente tienen costos más bajos en los tratamientos de fertilidad, lo que resalta la complejidad del acceso a estos servicios en diferentes contextos económicos [14]. La figura 1.1 muestra que Chile, a pesar de su PIB per cápita relativamente alto, enfrenta costos ART significativamente elevados, lo cual sugiere barreras económicas similares a las de países con menores ingresos.

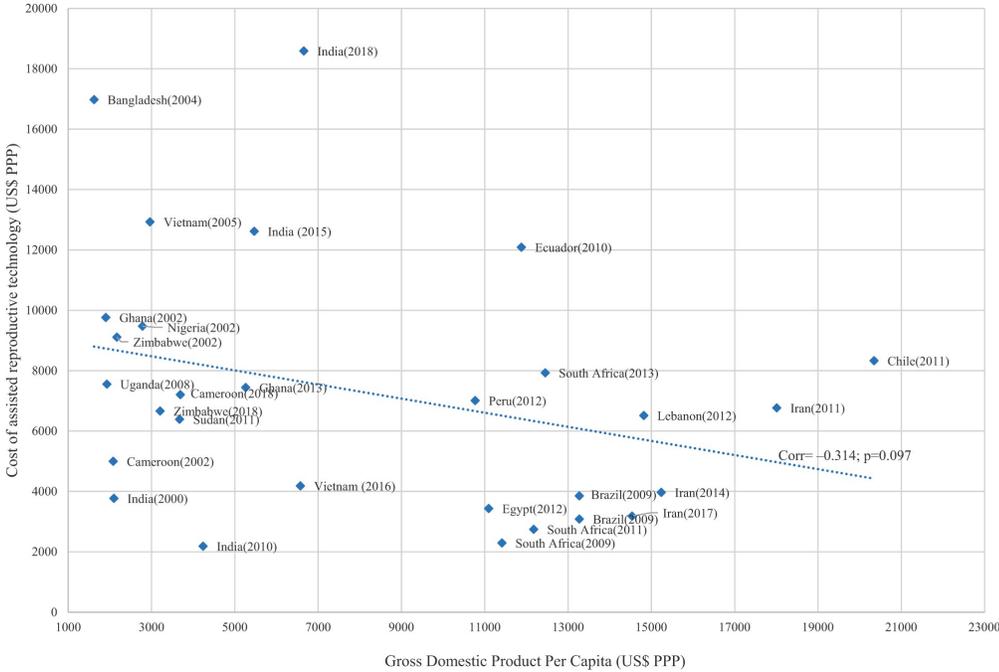


Figura 1.1: Costos directos por un ciclo de ART y PIB per cápita. La figura muestra el costo de un ciclo de ART en USD PPP en relación con el PIB per cápita, también en USD PPP. Fuente: [14].

El apoyo actual para el personal médico está normado en las Guías Ministeriales para el estudio y tratamiento de la pareja infértil [1], en cuya elaboración participaron miembros del IDIMI. Estas guías incluyen estudios previos para cada tratamiento, indicaciones específicas, factores de buen y mal pronóstico, y el número de ciclos recomendados. Las recomendaciones están basadas en análisis estadísticos como la regresión logística y las curvas ROC de factores aislados. No obstante, estas no ofrecen recomendaciones personalizadas para cada pareja que ingresa al programa, lo que puede resultar en derivaciones tardías o erróneas, aumentando el gasto público y disminuyendo la tasa de éxito de los tratamientos. Esta situación se resume en el árbol de problemas presentado en la figura 1.2.

La personalización del tratamiento es crucial debido a la gran variabilidad de las causas de infertilidad y de la respuesta de cada individuo a los tratamientos. Las guías actuales, al no considerar esta variabilidad, pueden llevar a la aplicación de tratamientos subóptimos para ciertos pacientes. Además, la falta de personalización puede contribuir a una mayor incertidumbre y frustración en las parejas, las cuales a menudo deben someterse a múltiples ciclos de tratamiento sin garantía de éxito.

En otras palabras, las soluciones actuales, basadas en herramientas clásicas, han alcanzado sus límites. La estandarización de las guías clínicas basadas en datos de poblaciones de Estados Unidos y Europa no resulta suficiente para abordar las necesidades del contexto chileno. Por lo tanto, surge la necesidad de herramientas que permitan al personal médico tomar decisiones más informadas y adaptadas a la realidad local.

El presente trabajo busca abordar esta problemática mediante el desarrollo de un modelo de machine learning que permita predecir la probabilidad de éxito de los pacientes del IDIMI en tratamientos de alta complejidad. Esta solución será de gran utilidad para el personal clínico, apoyando la toma de decisiones basada en los antecedentes y exámenes locales, además de identificar las variables más importantes que determinan el éxito o fracaso de un tratamiento de estas características. Con esta herramienta, se espera mejorar la precisión y eficiencia en la selección y manejo de los tratamientos, reduciendo así el tiempo y los costos asociados, y aumentando las tasas de éxito y la satisfacción de las parejas.

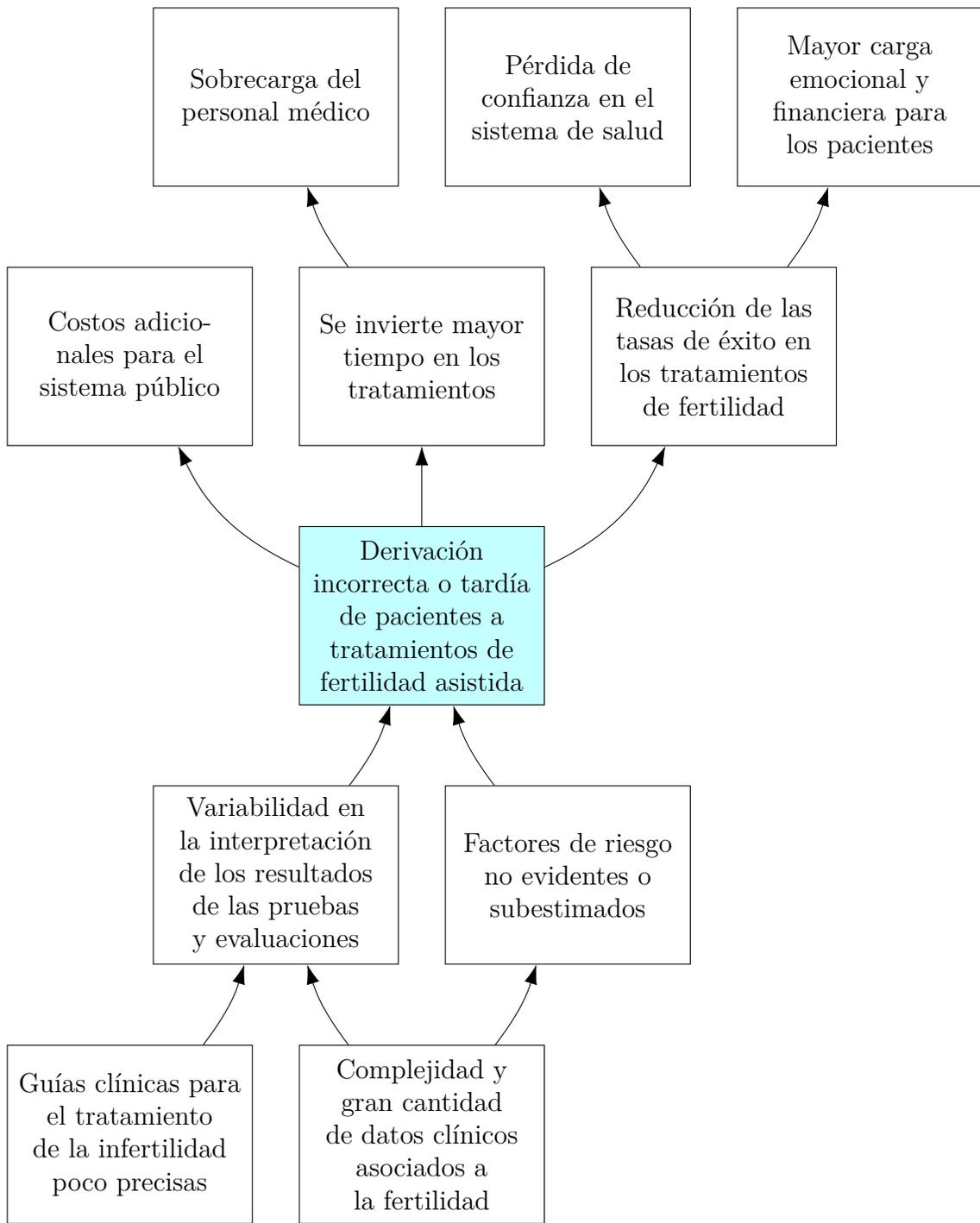


Figura 1.2: Árbol de problemas. Fuente: Elaboración propia.

1.4. Hipótesis de investigación

En base a los antecedentes generales y justificación del problema se plantea la siguiente hipótesis de investigación:

Es posible determinar la probabilidad de éxito de un tratamiento de alta complejidad para pacientes del IDIMI mediante el desarrollo de un modelo de clasificación basado en algoritmos de machine learning, utilizando parámetros clínicos y marcadores bioquímicos relacionados con la paciente.

1.5. Objetivos

1.5.1. Objetivo general

Elaborar un modelo de machine learning, utilizando parámetros clínicos y marcadores bioquímicos relacionados con la paciente para determinar la probabilidad de embarazo con tratamientos de alta complejidad en pacientes con infertilidad.

1.5.2. Objetivos específicos

1. Elaborar el estado del arte respecto al uso de algoritmos de machine learning para tratamientos de infertilidad, con el fin de forjar una base sólida de conocimientos y experiencias previas identificando enfoques exitosos y mejores prácticas.
2. Construir y evaluar distintos modelos machine learning, utilizando métricas como especificidad, sensibilidad y robustez.
3. Identificar y seleccionar las variables más relevantes y significativas en la tasa de éxito de un tratamiento de alta complejidad para los pacientes del IDIMI, con el fin de ser de utilidad para el personal médico de asignación.
4. Valorizar el impacto social y económico de la solución, con el propósito de cuantificar el ahorro económico y los beneficios sociales derivados de la implementación del modelo.

1.6. Metodología

En primera instancia se realizará una investigación exhaustiva del estado del arte respecto a la utilización de algoritmos de machine learning en medicina y posteriormente su uso específico en tratamientos de fertilidad asistida, esto permitirá identificar las variables más relevantes utilizadas en los tratamientos y encontrar los instrumentos o modelos más pertinentes para la clasificación de los tratamientos. Junto con esto, se realizarán reuniones con distintos profesionales de la salud pertenecientes al IDIMI, con el fin de conocer cómo es el proceso actual que lleva a cabo una paciente y cómo se realiza el tratamiento de los datos obtenidos de los exámenes. En cuanto al desarrollo de los modelos de clasificación, se cuenta con una base de datos con un registro de más de 1600 pacientes en control en el Hospital Clínico San Borja Arriarán e IDIMI desde el año 2015 hasta el 2019. Esta base de datos incluye información detallada de pacientes, características clínicas, resultados de pruebas y tratamientos anteriores, entre otros datos relevantes. Una vez obtenida, se hará un

preprocesamiento de los datos para asegurar la coherencia y la calidad de estos siguiendo la metodología de Knowledge Discovery in Database (KDD).

En este contexto, la etapa siguiente implica realizar un análisis exploratorio de los datos para identificar patrones, tendencias y relaciones significativas entre las variables. Este análisis inicial proporcionará una comprensión profunda de la estructura de los datos, y guiará las decisiones sobre los modelos y la transformación de los mismos.

En cuanto al tratamiento de los datos, se abordará el desbalance de clases en la variable objetivo en caso de ser necesario y se aplicarán técnicas de preprocesamiento como la codificación de variables categóricas y la normalización de variables numéricas. Además, se seleccionarán modelos de machine learning adecuados con toda la información obtenida en el estado del arte, tales como Gradient Boosting, Random Forest, Support Vector Machines, entre otros, y se entrenarán para determinar la probabilidad de embarazo de un tratamiento de alta complejidad. La implementación de los modelos se realizará con todas las variables disponibles luego del preprocesamiento y se determinará la importancia de estas variables en la clasificación posterior a esta mediante SHAP (SHapley Additive exPlanations).

La interpretación y evaluación de los modelos jugará un papel crucial en la toma de decisiones informadas. Se aplicarán métricas de evaluación como precisión, sensibilidad y área bajo la curva ROC.

Finalmente, se realizará una evaluación del impacto social y económico tomando como base el cálculo de costos directos asociados a los tratamientos de fertilidad asistida no exitosos, evaluando además el impacto dadas las limitaciones que surjan de los modelos generados. Para esto, se revisará a detalle la estructura de costos actual de los procedimientos de fertilización asistida, incluyendo tanto los costos estatales, como los incurridos por las pacientes que acceden a estos tratamientos.

1.7. Alcances

Los alcances del trabajo se centran en la utilización de datos digitales provenientes exclusivamente del IDIMI y del Hospital Clínico San Borja Arriarán, los cuales serán utilizados de manera anónima y protegida para garantizar la confidencialidad y privacidad de los pacientes. Se hará uso exclusivo del registro de los tratamientos de alta complejidad, quedando fuera los de baja complejidad, ya que no existen datos digitalizados suficientes que permitan realizar modelos robustos que incluyan esta información.

Se excluye cualquier información de fichas clínicas físicas con el objetivo de preservar la integridad y confiabilidad de los datos. Además, se establecen límites claros en cuanto a la recopilación, uso y manejo de la información, asegurando el cumplimiento de las normativas y políticas internas de protección de datos de las instituciones involucradas. En cuanto a los modelos de machine learning, el trabajo se centra en métodos no neuronales, incluyendo específicamente Random Forest, Support Vector Machine, Gradient Boosting y XGBoost. Aunque se exploran modelos neuronales de manera preliminar, queda fuera del alcance estudiar su interpretabilidad, y se ofrecerán recomendaciones para futuros trabajos que deseen emplear estas herramientas. Cabe destacar que este proyecto actúa como punto de partida para definir el enfoque de la plataforma que se desea implementar en el proyecto completo, brindando herramientas de apoyo para el análisis y tratamiento de datos en el contexto de la fertilidad asistida, quedando fuera de este proyecto la construcción, evaluación y validación

de la plataforma.

1.8. Resultados esperados

Se espera que este proyecto genere un modelo de machine learning capaz de identificar de manera precisa cuál es la probabilidad de éxito de la paciente si adopta un tratamiento de alta complejidad junto con entregar las variables más importantes en esta predicción. Este modelo permitirá una asignación más precisa y personalizada de los tratamientos, adaptándolos a las características individuales de cada paciente. Esto mejoraría los resultados clínicos al reducir el tiempo de espera para el embarazo exitoso y evitar tratamientos prolongados si el pronóstico no es favorable. Además, optimizaría los recursos del sistema de salud al priorizar a las pacientes con mayores posibilidades de éxito.

La aplicación práctica de estos modelos estará respaldada por un análisis exhaustivo del impacto social y económico. Este análisis permitirá evaluar de manera detallada cómo la solución propuesta influirá en el sistema actual de tratamientos de fertilización asistida, tanto en términos de eficiencia clínica como en el uso de recursos económicos y humanos. Esto proporcionará una visión global de los beneficios y desafíos asociados con la implementación de la solución propuesta.

Adicionalmente, más allá de los resultados métricos obtenidos por los modelos, se generarán recomendaciones fundamentadas para el desarrollo futuro de estos modelos, lo que permitirá tener una base sólida para trabajar en iteraciones futuras con la información de los tratamientos de baja complejidad efectuados por el IDIMI que no fueron considerados en este proyecto. Estas recomendaciones se basarán en los resultados obtenidos, así como en el estado del arte y las mejores prácticas identificadas a lo largo del proyecto. Esto asegurará que el modelo evolucione de manera efectiva y se mantenga relevante en un contexto en constante cambio de la medicina reproductiva e inteligencia artificial aplicada a la salud.

Capítulo 2

Marco conceptual

2.1. Conceptos médicos

2.1.1. Infertilidad

La infertilidad se define como una afección del sistema reproductivo, tanto masculino como femenino, que se manifiesta en la incapacidad de lograr un embarazo después de 12 meses o más de relaciones sexuales sin protección. Las razones detrás de la infertilidad son extremadamente variadas, pudiendo afectar a uno o ambos miembros de la pareja, y en cada caso pueden existir múltiples factores contribuyentes, lo que complica significativamente el proceso de tratamiento clínico. Esta condición tiene consecuencias sustanciales en la vida de las parejas afectadas. En el caso de las mujeres, se enfrentan a un mayor riesgo de violencia, divorcio, estigmatización social, así como a niveles elevados de estrés emocional, depresión, ansiedad y baja autoestima. Además, la infertilidad puede tener impactos económicos negativos cuando los hogares deben afrontar costos financieros significativos para acceder a tratamientos, alcanzando niveles catastróficos en algunos casos [15].

2.1.2. Fertilización in vitro

El tratamiento de fertilización in vitro (FIV) es un procedimiento complejo y prolongado que implica múltiples etapas. Este proceso comienza con la estimulación ovárica mediante la administración de hormonas gonadotrópicas a través de inyecciones subcutáneas, lo cual estimula el crecimiento de varios folículos en los ovarios. Durante esta fase, la paciente debe asistir a consultas frecuentes para realizar ultrasonidos transvaginales y análisis de sangre, que permiten monitorear el desarrollo folicular y ajustar las dosis hormonales según sea necesario.

Una vez que los folículos han alcanzado la madurez, se realiza la punción folicular (Paso 1 en la figura 2.1), que consiste en extraer los óvulos mediante un procedimiento mínimamente invasivo guiado por ecografía. Esta intervención, aunque rápida, puede resultar incómoda y, en algunos casos, requiere anestesia local o sedación.

La siguiente etapa es la fecundación de los óvulos, que puede llevarse a cabo de dos maneras (Paso 2 en la figura 2.1): por medio de la FIV convencional, donde los óvulos y espermatozoides se colocan juntos en un medio de cultivo, o mediante la técnica de inyección intracitoplasmática de espermatozoides (ICSI), en la cual se introduce un espermatozoide directamente en el óvulo. La técnica ICSI se utiliza especialmente en casos de infertilidad masculina severa.

Después de la fecundación, los embriones se mantienen en un cultivo embrionario (Paso 3 en la figura 2.1) en el laboratorio durante un período de 3 a 5 días, mientras se evalúa su desarrollo. Solo los embriones de mejor calidad serán seleccionados para el siguiente paso.

Finalmente, los embriones seleccionados se transfieren al útero de la paciente (Paso 4 en la figura 2.1) mediante un procedimiento conocido como transferencia embrionaria, que, aunque simple, requiere precisión para optimizar las probabilidades de implantación.

Este tratamiento es considerablemente más costoso y prolongado que otros métodos de reproducción asistida, como la inseminación intrauterina (IIU), y puede requerir hospitalización durante al menos 3 días en promedio debido a las intervenciones involucradas. Adicionalmente, un pequeño porcentaje de las pacientes (alrededor del 2%) puede experimentar una complicación grave conocida como Síndrome de Hiperestimulación Ovárica, la cual puede requerir hospitalización en la Unidad de Cuidados Intensivos (UCI) con una duración media de 5 días para estabilizar a la paciente [16][17].

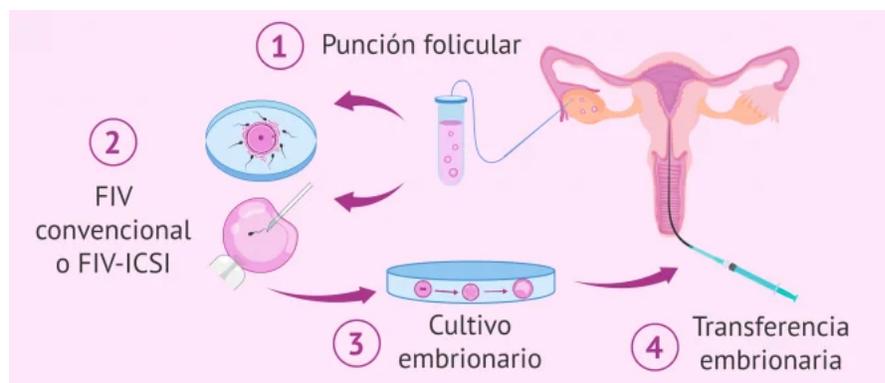


Figura 2.1: Proceso de fertilización in Vitro (FIV). Fuente: [18].

2.2. Descubrimiento de conocimiento en bases de datos - Metodología KDD

Knowledge Discovery in Databases (KDD) es un proceso interactivo e iterativo utilizado para la extracción de información útil, previamente desconocida y potencialmente valiosa a partir de grandes volúmenes de datos. Los pasos involucrados varían respecto de las decisiones tomadas por el usuario, pero se pueden resumir a grandes rasgos de la siguiente manera [19]:

- 1. Aprendizaje del ámbito de aplicación:** Con el propósito de fundamentar sólidamente el conocimiento en el área de estudio y garantizar una aplicación coherente de los modelos, el primer paso de la metodología KDD implica realizar una exhaustiva investigación del estado del arte y definir claramente los objetivos de la aplicación.
- 2. Selección del conjunto de datos:** Este paso implica consolidar las diversas fuentes de datos y seleccionar los datos pertinentes para abordar el problema específico que se desea resolver.
- 3. Limpieza y preprocesamiento de los datos:** Este proceso involucra operaciones básicas como eliminar ruido u outliers si es necesario, recopilar la información necesaria

para modelar o tener en cuenta el ruido y decidir estrategias para manejar campos de datos faltantes.

4. **Reducción de datos:** En este paso se busca identificar características relevantes para representar los datos, según el objetivo de la tarea, además de emplear técnicas para reducir la dimensionalidad o transformar los datos con el fin de disminuir el número de variables consideradas o encontrar representaciones más simplificadas.
5. **Definir el objetivo de la minería de datos:** Este paso involucra la selección del objetivo o propósito del modelo generado por el algoritmo de minería de datos, como por ejemplo: clasificación, regresión o agrupación.
6. **Elección del algoritmo de minería de datos:** Se elige el algoritmo más adecuado para cumplir con el objetivo del proyecto, junto con los parámetros más apropiados asociados a dicho algoritmo.
7. **Minería de datos:** Comprende la exploración y análisis de patrones, tendencias o estructuras significativas en los datos mediante la ejecución de los modelos y algoritmos seleccionados.
8. **Interpretación:** Incluye la interpretación de los patrones descubiertos y la posibilidad de volver a cualquiera de los pasos anteriores, así como la visualización de los patrones extraídos y la eliminación de patrones redundantes o irrelevantes.
9. **Usar el conocimiento adquirido:** Este último punto comprende la incorporación de este conocimiento en el sistema de rendimiento, la toma de acciones basadas en el conocimiento, o simplemente la documentación e información a las partes interesadas, así como verificar y resolver posibles conflictos con el conocimiento previamente creído.

Aunque existen otros procesos de extracción de datos, como CRISP-DM y SEMMA, en este trabajo se optó por utilizar KDD debido a su carácter más completo y preciso. CRISP-DM y SEMMA están orientados principalmente a aplicaciones empresariales, especialmente SEMMA, que fue desarrollado y es utilizado por la empresa de minería de datos SAS en conjunto con su software. Por otro lado, el enfoque iterativo de KDD permite ajustar y mejorar continuamente los resultados a lo largo de cada fase del proceso, lo que contribuye a un descubrimiento de conocimiento más efectivo y robusto [20].

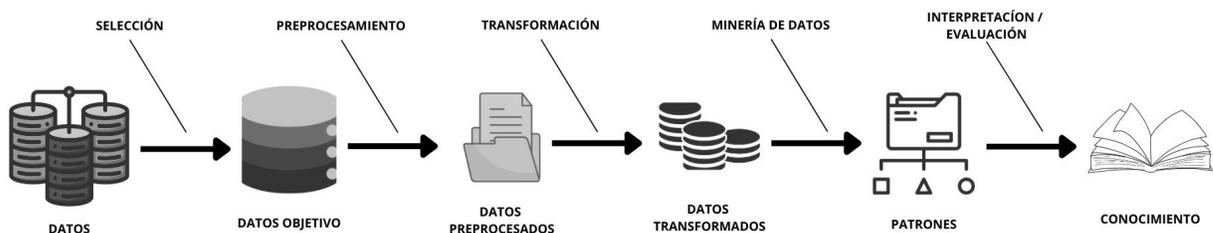


Figura 2.2: Proceso KDD. Fuente: Elaboración propia.

2.3. Conceptos relevantes para la evaluación de desempeño de modelos

2.3.1. Matriz de confusión

Cuando se trata de evaluar si un modelo de clasificación está haciendo bien su trabajo, se tienen distintas herramientas para validar aquello. Una de ellas es realizar una representación visual a través de la matriz de confusión, la cual está compuesta por cuatro clasificaciones [21]:

- **Verdaderos negativos (VN):** Valor real negativo y el modelo predijo también que el resultado era negativo.
- **Verdaderos positivos (VP):** Valor real positivo y el modelo predijo que también el resultado era positivo.
- **Falsos negativos (FN):** Valor real positivo y el modelo predijo erróneamente un resultado negativo.
- **Falsos positivos (FP):** Valor real negativo y el modelo predijo erróneamente un resultado positivo.

Esto se resume comúnmente de manera visual como se muestra a continuación:

Tabla 2.1: Matriz de confusión. Fuente: Elaboración propia

		Valor Real	
		Positivo	Negativo
Valor predicho	Positivo	VP	FP
	Negativo	FN	VN

La matriz de confusión da pie a diversas métricas que permiten evaluar los modelos de clasificación en distintas aristas.

2.3.2. Métricas clave

- **Accuracy:** Indica la proporción de los datos que fueron clasificados correctamente.

$$Accuracy = \frac{VP + VN}{\text{Todos los casos}} \quad (2.1)$$

- **Recall:** Indica la proporción de casos positivos que fueron clasificados correctamente. Esta es una de las métricas más importantes, ya que en el área de la medicina es fundamental minimizar la cantidad de pacientes mal diagnosticados.

$$Recall = \frac{VP}{VP + FN} \quad (2.2)$$

- **Precision:** Indica la proporción de clasificaciones positivas hechas correctamente.

$$Precision = \frac{VP}{VP + FP} \quad (2.3)$$

- **F1-Score:** Es la media armónica entre la precisión y la sensibilidad (recall). Un valor alto de F1 indica un buen equilibrio entre precision y recall.

$$F1-Score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (2.4)$$

2.3.3. Machine Learning

El aprendizaje de máquinas (ML, del inglés *machine learning*) es un subconjunto de la inteligencia artificial (IA) que se caracteriza por aprender de manera experimental, similar a la inteligencia humana. Además, tiene la capacidad de mejorar sus análisis mediante el uso de algoritmos computacionales. Estos algoritmos utilizan grandes conjuntos de datos de entrada y salida para reconocer patrones y que la máquina aprenda de forma eficaz para hacer recomendaciones o tomar decisiones autónomas. Tras suficientes repeticiones y modificaciones del algoritmo, la máquina llega a ser capaz de tomar una entrada y predecir un resultado [22].

Existen dos grandes tipos de enfoques dentro del ML: los algoritmos de aprendizaje supervisado y no supervisado. El tipo de algoritmo empleado depende del tipo de problema que se quiera resolver, del número de variables, y del enfoque que se le quiera dar a este [23].

Aprendizaje supervisado

Los algoritmos supervisados son aquellos que requieren ayuda externa para aprender en el contexto del aprendizaje automático. Estos algoritmos utilizan pares de entrada-salida para inferir una función a partir de datos etiquetados, lo que se conoce como el conjunto de entrenamiento. Este conjunto de entrenamiento representa una parte de los datos de entrada. La otra parte, denominada conjunto de prueba, se utiliza para evaluar la efectividad del modelo en predecir o clasificar, según el objetivo deseado. Los algoritmos más famosos de este tipo son la regresión lineal, regresión logística, Máquinas de Vectores de Soporte (SVM), Bosque Aleatorio (RF), entre otros [23].

Aprendizaje no supervisado

Los algoritmos de aprendizaje no supervisado se denominan así porque, a diferencia del aprendizaje supervisado, no cuentan con respuestas correctas ni un maestro que los guíe. Estos algoritmos exploran los datos por sí mismos para descubrir y presentar estructuras interesantes. Aprenden características inherentes de los datos sin necesidad de etiquetas previas. Cuando se les presentan nuevos datos, utilizan lo aprendido para reconocer patrones o agrupar elementos. Los algoritmos más famosos de este tipo son el Análisis de Componentes Principales (PCA), Agrupación de K-medias (K-means), Reglas de Asociación, entre otros [23].

Redes neuronales

Las redes neuronales artificiales (RNA) son un subconjunto importante de los algoritmos de aprendizaje automático, inspirados en la estructura y funcionamiento del cerebro humano. Estos tipos de modelo consisten en capas de nodos o “neuronas” conectadas entre sí, donde cada capa procesa los datos y ajusta los pesos de sus conexiones mediante un proceso iterativo de entrenamiento. Su capacidad para aprender patrones complejos y no lineales los convierte en una herramienta poderosa para tareas de clasificación y predicción, especialmente en problemas donde las relaciones entre las variables no son evidentes a simple vista [23].

Las redes neuronales se componen principalmente de tres tipos de capas:

- **Capa de entrada:** Recibe los datos.
- **Capas ocultas:** Donde se realiza la mayor parte del procesamiento de la información a través de funciones de activación no lineales, como por ejemplo, la función ReLU (Rectified Linear Unit).
- **Capa de salida:** Produce la predicción final.

Dependiendo de la complejidad del problema, una red neuronal puede contener múltiples capas ocultas. Durante el entrenamiento, se utiliza el algoritmo de retropropagación (back-propagation) para ajustar los pesos de las conexiones internas y reducir el error entre las predicciones del modelo y las etiquetas reales. Este proceso es iterativo y asegura que la red neuronal sea capaz de generalizar bien a nuevos datos [24].

2.3.4. Validación cruzada

Con el fin de evaluar multidisciplinariamente el desempeño de un modelo de clasificación existen otras técnicas como la validación cruzada, la cual permite evaluar el rendimiento del modelo con nuevos datos que no fueron utilizados en el entrenamiento. Existen distintas variaciones de esta validación con distintos enfoques, por ejemplo la división por registros (*Record-wise*) o la división por materias (*Subject-wise*). Esta última es la que se usará en este caso, ya que un estudio reciente reveló que es más apropiada para datos médicos [25]. Esta técnica consiste en dividir el conjunto de entrenamiento y prueba, de tal manera que se tengan distintos sujetos en ambos conjuntos, simulando un estudio clínico. La representación gráfica de esta se puede ver en la figura 2.3.

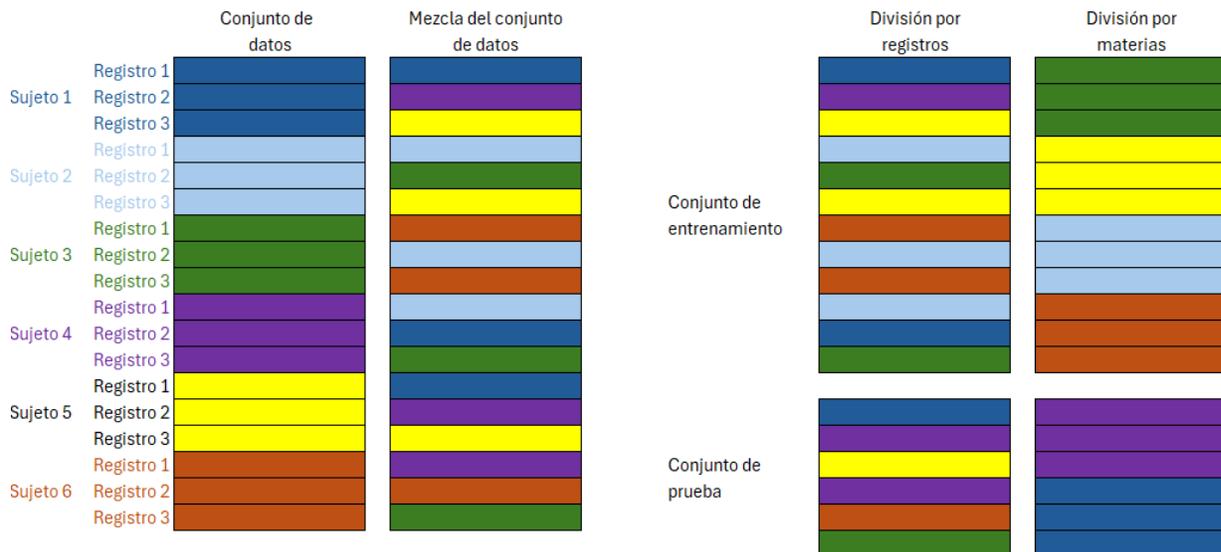


Figura 2.3: División por registros y división por materias. Fuente: Elaboración propia.

2.3.5. Método SHAP y valores Shapley

El método SHAP (SHapley Additive exPlanations) se basa en los valores Shapley, originados en la teoría de juegos cooperativos. Estos valores se utilizan para calcular la contribución individual de cada característica a la predicción de un modelo, de manera similar a cómo se distribuye el valor entre jugadores en un juego colaborativo.

Los valores Shapley garantizan que la predicción del modelo se descomponga de manera justa entre las características, considerando todas sus combinaciones posibles y sus contribuciones marginales. SHAP cumple tres propiedades clave: precisión local, que asegura que la suma de las contribuciones individuales iguale la predicción del modelo; consistencia, que garantiza que una característica con mayor impacto en un modelo se refleje con mayor importancia; y efecto nulo, donde las características que no influyen en la predicción tienen un valor SHAP de cero [26].

En la práctica, SHAP genera muestras artificiales basadas en los datos de referencia y las pondera utilizando un núcleo SHAP. Luego, se entrena un modelo de regresión local para estimar la importancia de las características. Este método permite interpretar de manera clara el papel de cada variable en las predicciones del modelo, algo crucial en aplicaciones donde la interpretabilidad es esencial, como lo es en los tratamientos de fertilización asistida.

Capítulo 3

Estado del arte

El estudio de la fertilidad y la efectividad de los tratamientos de fertilización asistida es un campo complejo que ha generado una variedad de enfoques y herramientas de medición para evaluar diferentes aspectos de este proceso. En este contexto, la medición de variables clave y el análisis de datos son fundamentales para comprender mejor los factores que influyen en la fertilidad y en el éxito de los tratamientos.

Como se mencionó anteriormente, el MINSAL en el año 2015 normó el estudio de la pareja infértil mediante la creación de guías ministeriales. Esta guía está orientada a profesionales responsables de la atención de las parejas que sufren infertilidad y otras formas de discapacidad reproductiva, tales como Médicos Generales, Matronas y Matrones, Endocrinólogos, entre otros. El objetivo principal de esta es establecer recomendaciones técnicas generales para el Estudio y Tratamiento de la Infertilidad y otras discapacidades reproductivas, orientadas a un óptimo funcionamiento de cada Nivel de Atención de la Red Asistencial, según su capacidad resolutoria [1].

Sin embargo, la variabilidad de los factores relevantes para estimar el éxito o el fracaso de un tratamiento de estas características para cada paciente hace necesario ampliar el espectro de soluciones, por lo que este estado del arte está orientado a conocer los distintos enfoques en el ámbito que hay hasta la fecha.

Diversos estudios señalaron que entre las causas más comunes de infertilidad femenina encontradas, el Síndrome de Ovario Poliquístico (SOP) y la patología tubárica fueron las más frecuentes, contribuyendo con un 46 % y un 33.8 % respectivamente, a todos los casos atribuidos a la infertilidad femenina. El SOP es un trastorno hormonal que afecta el equilibrio de las hormonas reproductivas, lo que dificulta la ovulación regular y puede causar la aparición de quistes en los ovarios, interfiriendo con la fertilidad. Por otro lado, la patología tubárica se refiere a alteraciones o bloqueos en las trompas de Falopio, lo que impide el encuentro entre el óvulo y el espermatozoide, causando infertilidad.

En el 7.6 % de los casos se encontró endometriosis, una enfermedad donde el tejido que normalmente recubre el interior del útero crece fuera de él, afectando órganos como las trompas de Falopio y los ovarios, y causando distorsión tubárica y adherencias que dificultan la concepción. Así pues, la detección precoz y el tratamiento oportuno de la Enfermedad Pélvica Inflamatoria (EPI), una infección que afecta los órganos reproductores femeninos y que puede provocar daños permanentes en el útero, ovarios y trompas de Falopio, son muy esenciales para abordar esta causa de infertilidad [27] [28].

Además, se ha comprobado que la edad de la mujer y la duración de la infertilidad también son factores determinantes. En estudios recientes, se ha observado que la Tasa Acumulativa de Embarazo por Ciclo Iniciado (TAECI) disminuye en un 15 % y un 35.9 % cuando la edad de la mujer supera los 35 años y los 39 años, respectivamente. Asimismo, la duración de la infertilidad también influye, con una disminución en la TAECI de hasta un 50 % cuando la infertilidad persiste por más de 36 meses [29].

Otros estudios realizados en otros lugares del mundo también muestran que el SOP es la causa más común de infertilidad femenina. También se descubrió que no existen diferencias significativas entre pacientes con SOP con sobrepeso u obesidad y aquellos con índice de masa corporal normal, lo que indica que las pacientes con SOP delgadas son igualmente propensas a la infertilidad que las pacientes con SOP obesas [27].

Por otro lado, para el factor masculino, existe una relación directa entre tasa de embarazo y el número de espermatozoides motiles (NEM) a inseminar. La tasa de embarazo es del 18 % con un NEM > 10 millones/mL, 12 % con un NEM entre 9 a 2 millones/mL y es menor de 3 % cuando el NEM es menor de 1 millón/mL [29]. Las causas y los factores de riesgo de la infertilidad masculina son numerosos y pueden determinar una alteración pretesticular (hipotalámica o hipofisaria, con niveles bajos de gonadotropina), testicular (caracterizada por niveles elevados de gonadotropina) y post-testicular (con niveles de gonadotropina generalmente en el normales) [30] [31].

Diversos autores han observado que la probabilidad de éxito en tratamientos de fertilización asistida, especialmente los de alta complejidad como la Fecundación In Vitro (IVF por sus siglas en inglés), aumenta significativamente al realizar múltiples ciclos. Se ha reportado una probabilidad de éxito del 38 % al 57 % para dos y tres ciclos, respectivamente, dependiendo de las características individuales del paciente [32] [33]. Por tanto, considerar la tasa acumulativa de embarazo al estimar la probabilidad de éxito en pacientes con ciclos previos resulta crucial [34].

Estos hallazgos evidencian la complejidad y la importancia de considerar múltiples factores al momento de evaluar la eficacia de los tratamientos de fertilización asistida, lo cual constituye una base sólida para el desarrollo de enfoques más personalizados y efectivos en este campo.

A continuación se mostrarán una serie de soluciones existentes en el área, con distintos enfoques que han intentado atacar esta problemática.

3.1. Diversos enfoques para identificar las variables más relevantes en los tratamientos y/o predecir el éxito del embarazo

La capacidad de estos algoritmos de aprender rápidamente, ha hecho que en las últimas décadas su popularidad aumente en el área de la medicina. Sobre todo en la infertilidad, donde se ha utilizado para predecir el éxito de los tratamientos o conocer la importancia que tienen las variables involucradas en el éxito o el fracaso de los procedimientos.

En el caso de los tratamientos IVF, el enfoque principal para estimar la probabilidad de embarazo de estos, se los ha llevado el enfoque supervisado, en específico las regresiones logísticas. Esta herramienta ha sido de gran importancia para la investigación clínica y epi-

demiológica, ya que permite modelar la probabilidad de que ocurra un evento en función de los factores de riesgo, al cuantificarlos con significación estadística [35]. Hasta el año 2020, la mayoría de los estudios (92 de 142, equivalente al 64.8 %) emplearon regresiones logísticas para desarrollar modelos predictivos. Por otro lado, un número significativo de estudios que no utilizaron este enfoque optaron por emplear redes neuronales artificiales (20 de 142, equivalente al 14.1 %) [36].

Por ejemplo, en [37] se emplearon modelos de regresión logística multivariable (MLR por sus siglas en inglés) para anticipar los desenlaces de los procedimientos de transferencia de embriones, considerando las restricciones inherentes a los datos obtenidos de manera habitual. Gracias a este extenso conjunto de datos, el modelo desarrollado facilita la identificación de factores específicos que influyen en la viabilidad del embrión o la receptividad materna. La variabilidad observada en ambos componentes entre distintos centros con resultados generales similares sugiere la posibilidad de continuar mejorando el tratamiento de FIV, es decir, la utilización de estos modelos a sido una herramientas fundamental previo al desarrollo de modelos de machine learning para identificar las variables más influyentes en los tratamientos. Sin embargo, en [38] se comparó la eficacia de la MLR y los modelos de perceptrón (MLP por sus siglas en inglés) en la predicción de embarazos durante tratamientos FIV. Los resultados revelaron que, si bien el análisis MLR resultaba más apropiado para cuestiones teóricas, el clasificador MLP demostró ser más efectivo en términos de predicción clínica. Utilizando un clasificador MLP con 26 entradas y tres capas ocultas de estructura 40-6-2, se logró superar al MLR en términos de la medida AUC por un margen de 0.061.

Aunque el enfoque supervisado no solo se ha reducido a utilizar regresiones logísticas. Por ejemplo en [39] se desarrollaron los algoritmos de regresión logística, RF, XG Boost y SVM para predecir el resultado de tratamientos IVF en el servicio de salud pública de Brasil. En este estudio el algoritmo que presentó mejor desempeño fue el RF con una precisión, sensibilidad y especificidad de 0.61 , 0.61 y 0.59 respectivamente. Además se utilizó el algoritmo SHAP [40] para identificar las variables más relevantes al momento de realizar la predicción, donde se identificó que la edad de la mujer y las diversas métricas de calidad embrionaria tales como el número de blastocistos o embriones tipo A son las más influyentes en los modelos. Este algoritmo y el Hill Climbing Approach han sido las principales herramientas tanto para mejorar el desempeño de los modelos, como para identificar las variables más relevantes en la probabilidad de éxito de tratamientos IVF, en [41] se utilizó el Hill Climbing Approach y se logró mejorar la precisión de un algoritmo SVM de 97.42 % a 98.38 % y un RF de 90.54 % a 98.25 % de precisión.

Otros estudios, han utilizado como herramienta para mejorar el desempeño de los modelos al algoritmo genético (GA por sus siglas en inglés), que se basa en operadores de inspiración biológica como la mutación, el cruce y la selección para generar soluciones de alta calidad a los problemas de optimización y búsqueda [42]. En [43]se desarrollo un modelo dual utilizando GA y los árboles de decisión para predecir el resultado de un tratamiento IVF, alcanzándose una precisión de 0.73. Además se propuso una serie de reglas *if-then* sobre las relaciones entre los atributos con el fin de ser de utilidad para el personal médico, permitiendo adaptarse a las necesidad y/o características de cada paciente. Siguiendo los mismos lineamientos, en un estudio reciente se desarrollaron 3 modelos: árboles de decisión, RF y XG boost. Y se utilizó GA para realizar simultáneamente el entrenamiento de múltiples modelos de predicción para simplificar y acelerar el proceso de formación. Donde se obtuvo que el modelo con mayor precisión fue XG boost con un 0.65 [44].

La metodología KDD es la principal herramienta utilizada al momento del desarrollo de modelos de machine learning [19]. Esta plantea diversos pasos fundamentales para crear modelos robustos y capaces de cumplir con los objetivos de manera óptima. Uno de los primeros pasos de esta, responde a la necesidad de preparar la información para ser utilizada en los modelos. La limpieza y preprocesamiento de información en el área de la medicina no es trabajo fácil, ya que, usualmente las bases de datos se encuentran desbalanceadas. Para afrontar esto se han estudiado diversas técnicas de remuestreo [45] [46] [47]. En [47] se compararon estas técnicas aplicando las metodologías en dos bases de datos representativas de cáncer de pulmón utilizadas para predecir la incidencia de esta enfermedad, ambas con un desbalance de datos donde los ratios de desbalance (la proporción de muestras de la clase mayoritaria con respecto a las de la clase minoritaria) son de 24.7 % y 25 % para cada una de las bases. De los 23 métodos de remuestreo utilizadas, se demostró que el *Over-sampling* es un método estable, y el AUC del modelo suele ser superior al de otras formas, por consiguiente su habilidad para predecir o clasificar es superior a comparación de no utilizar estas técnicas.

En cuanto a la validación cruzada, principal técnica utilizada en la aplicación de modelos de ML que busca dar lugar a un rendimiento medio superior al de la aplicación de una única estrategia de clasificación, y también reducir el riesgo de un rendimiento deficiente [48]. Es común realizar una división porcentual en la base de datos final a utilizar para implementar los modelos de un ratio del 70 % para el set de entrenamiento (parte de la información con el cual el modelo aprende) y el 30 % restante al set de testeo (datos etiquetados con los cual un modelo de ML supervisado evalúa el rendimiento de la predicción o clasificación). Pero a la hora de trabajar con bases de datos con información médica como es el caso de los tratamientos IVF, la validación cruzada por materias (*Subject-wise*) a demostrado tener mejor adaptación [25] a este tipo de información, esta validación cruzada consiste en no repetir los mismos pacientes en el set de entrenamiento y testeo, con el fin de no sesgar los resultados del modelo. En otros estudios también se ha optado por realizar distintas validaciones cruzadas, diviendo el set de entrenamiento y testeo en diferentes proporciones y comparando el AUC del modelo resultante, en [49] se realizaron estas distintas divisiones del set de datos, resultando que para el caso donde un 80 % de la información se deja para entrenamiento y el 20 % restante para testeo el modelo entregó mejores resultados con un AUC de 0.976, levemente superior a las otras divisiones.

Finalmente, otro enfoque estudiado se basa en las cadenas de Markov, un proceso estocástico que cumple con la propiedad de Markov. Esta propiedad establece que, dado un sistema en el que los estados están representados por variables aleatorias ($X_1, X_2, X_3, \dots, X_n$) en distintos momentos de tiempo, la probabilidad del estado futuro depende únicamente del estado presente y no de los estados anteriores. En términos formales, para cualquier $n \geq 1$ se cumple que [50]:

$$P(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_{n+1} = x | X_n = x_n) \quad (3.1)$$

Una de las diferencias clave del enfoque de modelado de Markov respecto a otros enfoques líderes (como modelos basados en ecuaciones diferenciales o modelos estadísticos) es su consideración de las probabilidades de eventos del pasado reciente, Como se explica en la ecuación 3.1 el estado siguiente (X_{n+1}) solo depende del estado actual (X_n), y no de cómo se llegó a él. Este tipo de modelado encuentra aplicación en diversos campos como biología,

medicina y otros [51].

En el caso de los tratamientos de infertilidad, Rao Y Diamond propusieron utilizar un tipo de análisis de modelización de Markov para comprender los datos generados por los tratamientos para la infertilidad en mujeres que reciben estimulación ovárica. Para esto definieron inicialmente 4 estados [52]:

- **Estado 0:** La mujer ha comenzado el tratamiento para la infertilidad y lo está continuando.
- **Estado 1:** La mujer logra concebir.
- **Estado 2:** La mujer sigue siendo infértil después de completar un número determinado de ciclos de tratamiento y se ha detenido el tratamiento.
- **Estado 3:** El embarazo concebido termina en un aborto espontáneo.
- **Estado 4:** El embarazo concebido culmina en el nacimiento de un bebé vivo.

Luego, por las características de estos estados, se definieron las 4 posibles transiciones entre estos:

- i) Estado 0 \rightarrow Estado 0
- ii) Estado 0 \rightarrow Estado 1 \rightarrow Estado 4
- iii) Estado 0 \rightarrow Estado 2
- iv) Estado 0 \rightarrow Estado 1 \rightarrow Estado 3

Iterativamente fueron añadiendo más estados intermedios, considerando las distintas complicaciones o procesos más específicos dentro del tratamiento de infertilidad. A partir de todo esto pudieron concluir que en ocasiones, la cantidad de datos sobre los cambios entre ciertos estados es insuficiente para establecer probabilidades que reflejen con exactitud a grupos de mujeres divididos por edades. Además, el supuesto de que no se requiere recordar eventos pasados para predecir futuros (propiedad de falta de memoria) podría no ser válido en casos de mujeres mayores o aquellas que ya han experimentado varios ciclos de fertilización in vitro antes de iniciar su tratamiento actual.

Esto significa que, para algunos segmentos de la población, como las mujeres de mayor edad o con tratamientos previos, puede ser difícil calcular las probabilidades de transición entre estados con precisión debido a la falta de datos representativos. Esto, a su vez, puede limitar la efectividad del modelo de Markov para predecir los resultados del tratamiento en estos grupos específicos.

A modo de conclusión, este proyecto presenta una propuesta novedosa que destaca por su adaptación a la realidad chilena y su enfoque en apoyar al personal médico de asignación. La mayoría de las herramientas y modelos actuales en el ámbito médico y de la salud son desarrollados en contextos internacionales, los cuales no necesariamente reflejan ni se adaptan de manera adecuada a las necesidades y particularidades de la población chilena. La especificidad de este proyecto, en cambio, reside en su diseño, orientado a responder y ajustarse a los patrones y datos específicos de la población local.

Además, la herramienta propuesta tiene un valor práctico relevante, al facilitar el trabajo del personal médico encargado de la asignación de tratamientos, proporcionándoles un soporte directo y contextualizado. Este aspecto no solo ayuda a optimizar la eficiencia en la toma de decisiones, sino que también se convierte en un referente en el país, promoviendo un enfoque personalizado y basado en evidencia que puede sentar precedentes para futuras investigaciones y desarrollos en el área de salud en Chile.

Capítulo 4

Desarrollo

4.1. Los datos

Para el presente trabajo, se dispuso de una serie de bases de datos que contienen los registros anuales de los ciclos realizados a los pacientes del IDIMI desde el 2010 hasta el año 2019. Cada una de estas bases cuenta con la información de los tratamientos con ovocitos descongelados (D) o frescos (F), con un total de 1353 y 3406 registros respectivamente.

Se utilizarán únicamente los registros de ovocitos frescos desde el año 2015 hasta el año 2019 por tres razones principales:

- **Nuevas variables:** Dependiendo del origen de los embriones utilizados en el tratamiento (D o F), se añaden nuevas variables. Por lo tanto, no es posible integrar una base de datos que contenga datos para todos los atributos de ambos tipos de embriones.
- **Clase Mayoritaria:** Se elige trabajar con los datos F porque representan el 70% del total de ciclos realizados en el periodo.
- **Equipamiento y Calidad Embrionaria:** Se decide excluir los registros desde el año 2010 hasta el 2014 debido a que en ese periodo se utilizaba diferente equipamiento relacionado al tratamiento. Además, en esos años no se registraba el número de blastocistos obtenidos, una característica importante para evaluar la calidad embrionaria. Siguiendo las recomendaciones expuestas en el estado del arte, se prefiere trabajar con las bases que cuentan con esta métrica.

Considerando lo anterior, previo al procesamiento de los datos se cuenta con un total de 1674 registros, dividido en 5 bases de datos, correspondiente a los ciclos con embriones frescos realizados por el IDIMI desde el año 2015 hasta 2019 en 1619 pacientes.

4.2. Preprocesamiento de los datos

4.2.1. Unificación bases de datos

El primer paso consiste en consolidar toda la información en una sola base de datos. Para lograrlo, fue necesario establecer un conjunto de variables comunes en todas las fuentes de información disponibles. Inicialmente, estas bases de datos contaban con un total de 54

columnas para la más reciente (2019) y 133 columnas para la que tenía el mayor número de variables (2017). Estas diferencias se debieron a los cambios en el registro de variables a lo largo del tiempo, reflejando las prioridades y nuevos enfoques en la recopilación de información.

De todas las variables disponibles, se identificó un subconjunto de mayor cardinalidad común entre todas las bases de datos, compuesto por 62 variables. Sin embargo, después de un proceso de selección riguroso en conjunto con el equipo del IDIMI, se redujo este conjunto a un total de 34 variables.

Esta reducción implicó:

1. **Eliminación de identificadores del paciente:** Se excluyeron variables como el nombre y los apellidos del paciente. Únicamente se mantuvo el RUT (Rol Único Tributario) como identificador, ya que será útil en la validación cruzada definida en la figura 2.3.
2. **Eliminación de variables no informativas:** Se descartaron variables relacionadas con los distintos especialistas involucrados en los procedimientos. También se excluyó la información sobre la incubadora utilizada y número de embriones criopreservados ya que no aportaba datos relevantes para el estudio.
3. **Variables en blanco:** Algunas variables venían completamente en blanco debido a que correspondían a características específicas de tratamientos realizados con embriones descongelados. Por ejemplo, se excluyeron datos sobre el donante o la cantidad de embriones descongelados. Es importante destacar que las variables con algunos registros en blanco no se eliminaron en este paso. Posteriormente, se estudiarán las razones detrás de esta falta de información y se tratarán caso a caso.

Con esta primera limpieza, la base de datos consolidada que contiene todos los periodos estudiados consta de 1504 registros y 32 columnas.

4.2.2. Tratamiento de datos faltantes

Como se mencionó en la sección anterior, al unificar la base de datos, hubo ciertas columnas que presentaban datos faltantes, pero no fueron eliminadas en esa instancia. A continuación, se mostrará el tratamiento específico que se realizó para cada una de estas variables, considerando el contexto del significado de cada una.

Inicialmente, se definió la variable objetivo del modelo como la variable EMBARAZO, que indica si en un ciclo específico de tratamiento se logró el embarazo o no. Originalmente, esta variable estaba representada por $\beta 1$, que corresponde al nivel de gonadotropina coriónica humana (hCG), una hormona producida por la placenta durante el embarazo. Si esta variable registra un valor mayor a 0, se considera que se logró el embarazo; cuando es 0, no se logró. Por conveniencia, se transformó esta variable a 1 para indicar el logro del embarazo y a 0 para indicar la ausencia de embarazo en el ciclo correspondiente, como se resume en la tabla 4.1.

Tabla 4.1: Categorías variable EMBARAZO

Valor	Embarazo
0	No
1	Si

En segundo lugar, para el caso de las variables antropométricas correspondientes a *TALLA* (altura en metros), *PESO* (peso en kg) e *IMC* (índice de masa corporal en kg/m²), se eliminaron los valores en blanco en los registros donde no se disponía de la talla, el peso ni el IMC. Sin embargo, en los casos donde faltaba el IMC pero se tenían los valores de la talla y el peso, se reconstruyó esta columna utilizando la ecuación 4.1.

$$IMC = \frac{PESO}{TALLA^2} \quad (4.1)$$

También existían casos donde se tenía el IMC, pero el peso o la talla no; para cada uno de estos casos, se utilizó la misma ecuación para reconstruir los datos faltantes.

En el caso de las variables *DÍAS DE ESTIMULACIÓN*, *EDAD*, *INDUCCIÓN*, *#TRANSF* Y $\beta 1$, se eliminaron los datos en blanco o con registros “NT”, que corresponden a los ciclos sin transferencia de embriones. Esto se debe a que no es posible reconstruir estas variables debido a su significado específico. Estos registros constituyen gran parte de la información que se tiene del período, reduciendo así enormemente el set de datos.

En relación a las variables *DIAG.1*, *DIAG.2* y *DIAG.3*, que corresponden a los distintos diagnósticos asociados a la infertilidad del paciente registrados por el equipo médico, se encontraron datos en blanco o faltantes para *DIAG.2* y *DIAG.3*. En estos casos, se tomó la decisión de asignar una categoría específica “0” para identificarlos durante la exploración posterior de los datos. Hasta este punto, no se han eliminado estos registros.

Finalmente, para las variables *OVO*, *MII*, *MI*, *VG* y otras que corresponden al número de ovocitos en cada etapa de maduración, se reemplazaron los datos faltantes por el valor 0. Esta misma estrategia se aplicó a las variables que registran el número de embriones en sus distintas etapas, así como al número de embriones criopreservados o al número de embriones transferidos. Todos los registros en blanco en estas variables fueron rellenados con 0.

4.2.3. Transformación de datos

4.2.3.1. Distintos diagnósticos

Como se mencionó en el tratamiento de datos faltantes, existen tres variables, *DIAG.1*, *DIAG.2* y *DIAG.3*, cada una de estas viene como un texto, que corresponde al nombre del diagnóstico. Para facilitar su uso en los modelos, se transformaron estas variables a numéricas, bajo las categorías de la tabla 4.2.

Tabla 4.2: Categorías diagnóstico. Fuente: Elaboración propia.

Valor numérico	Categoría
0	Sin diagnóstico
1	Factor masculino
2	Factor femenino anatómico (miomas, adherencias)
3	Factor femenino endocrino (anovulación)
4	Insuficiencia ovárica (low responder)
5	Endometriosis
6	ISCA
7	Otros

4.2.3.2. Distintas inducciones de ovulación

La variable *INDUCCION*, presente en el consolidado, aparece numéricamente en los registros de los periodos 2015 a 2018, pero en los registros del 2019, se presenta como texto. La traducción se realizó en base a la tabla 4.3, donde se muestra el valor numérico correspondiente a cada método de inducción. Esta nomenclatura es la misma que se utilizó en los registros de los periodos 2015 a 2018.

Tabla 4.3: Categorías inducción. Fuente: Elaboración propia.

Valor numérico	Categoría
1	Sin inducción
2	Agonista sub dermico + FSHr + HMG
3	Agonista + HMG
4	Antagonista + CC + HMG
5	Antagonista + FSHr + HMG
6	Antagonista + HMG
7	Antagonista FSHr
8	Antagonista + Femara
9	Antagonista + CC + HMG + FSHr
10	FSH r solo
11	Agonista + FSHr
12	Antagonista + HMG + CC
13	CC + HMG + FSHr
14	Antagonista + FSHr + MENOPUR
15	Antagonista + FSH + ELONVA
16	FSH + MENOPUR
17	MENOPUR
18	Antagonista + ELONVA + MENOPUR

4.3. Análisis exploratorio de los datos

La exploración de datos se dividió en dos grandes partes. La primera corresponde al análisis de correlaciones mediante un mapa de correlación. El objetivo de esta etapa es identificar y manejar variables independientes altamente correlacionadas entre sí, ya sea eliminándolas, reemplazándolas o transformándolas, para evitar problemas de multicolinealidad en los modelos a emplear.

La segunda parte del análisis se enfoca en el análisis gráfico de las variables restantes después de la primera etapa de exploración de datos, incluyendo el conjunto de variables que serán transformadas al final de esta etapa. El objetivo principal de esta fase es observar la distribución de los datos y tratar los *outliers* en el caso que sea necesario, y establecer finalmente el conjunto de datos que se utilizará en los modelos.

4.3.1. Análisis de correlaciones

Para iniciar la exploración de los datos, se generó un mapa de correlación (Figura 4.1) para comprender las relaciones entre las variables.

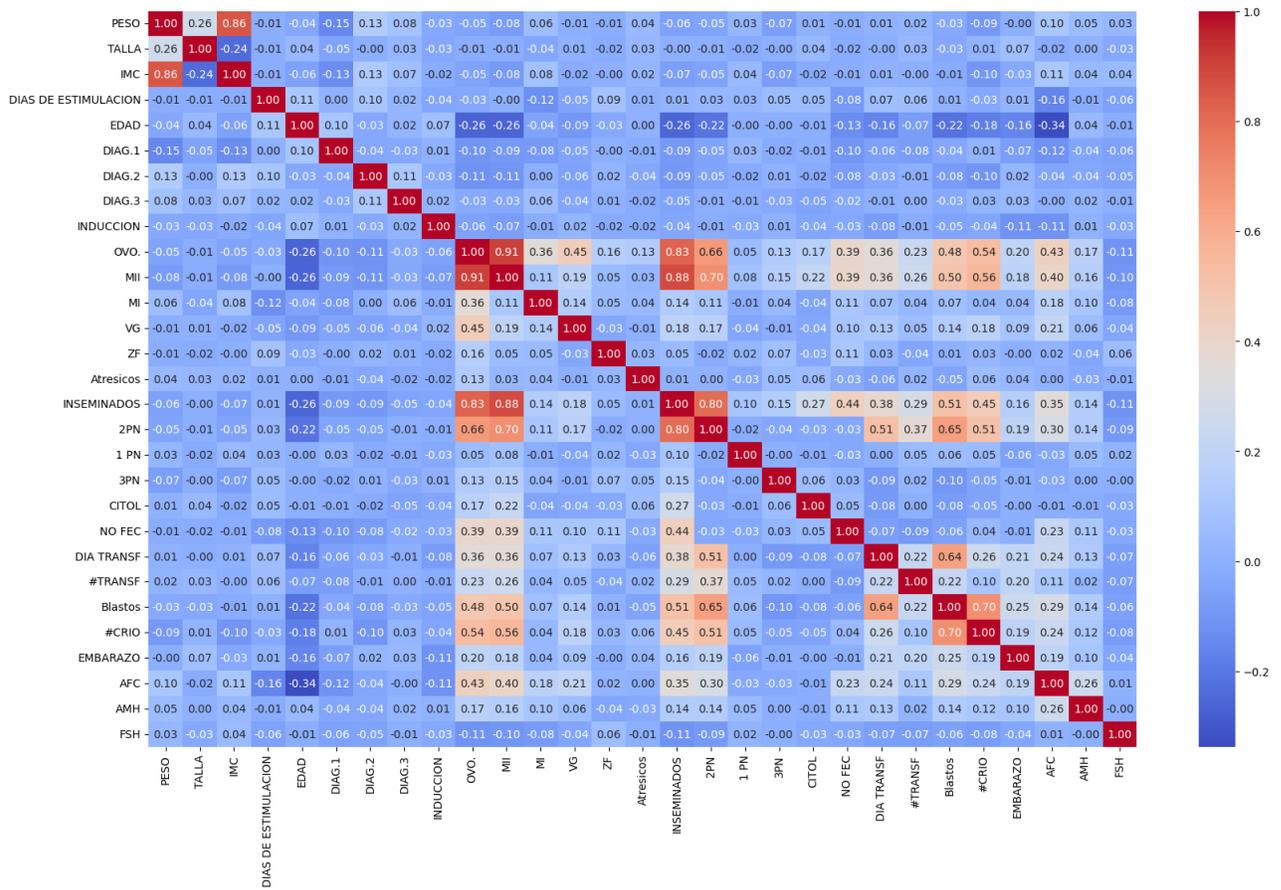


Figura 4.1: Mapa de correlación entre variables. Fuente: Elaboración propia.

A partir de este análisis, se observó una alta correlación entre las variables *PESO*, *TALLA* e *IMC*. Dado que estas variables están directamente relacionadas entre sí (ver ecuación 4.1), se decidió eliminar las primeras dos variables y mantener el *IMC* como el indicador antropométrico del paciente.

Como parte importante de los procedimiento IVF, se lleva un registros del número de ovocitos obtenidos, y las distintas etapas de maduración o estados de estos ovocitos, que les permiten ser aptos para ser transferidos (AT) o no aptos (NAT). Esta información está representada por las variables que se muestran en la tabla 4.4 y siguen la relación descrita en 4.2.

Tabla 4.4: Variables de caracterización de ovocitos obtenidos. Fuente: Elaboración propia.

Variable	Descripción
OVO	Número de ovocitos obtenidos
MII	Número de ovocitos en metafase II obtenidos (AT)
MI	Número de ovocitos en metafase I obtenidos (NAT)
VG	Número de ovocitos en vesícula germinal obtenidos (NAT)
ZF	Número de ovocitos zona fracturada obtenidos (NAT)
Atresicos	Número de ovocitos zona atresicos obtenidos (NAT)

$$OVO = MII + MI + VG + ZF + \text{Atresicos} \quad (4.2)$$

En este contexto, y con base en la información proporcionada por el mapa de correlación, se observa una alta correlación entre este subconjunto de variables. En particular, las variables *OVO* y *MII* muestran una correlación casi perfecta de 0.91. Mantener ambas variables en su forma actual podría generar efectos adversos en los modelos.

Por esta razón, se decidió eliminar las variables *OVO*, *MI*, *VG*, *ZG*, y *Atresicos*, conservando únicamente la información relacionada con el número de ovocitos en metafase II obtenidos, ya que estos son los que finalmente están aptos para el tratamiento.

En la siguiente etapa de los procedimientos, se registra el número de ovocitos inseminados en la variable *INSEMINADOS*. Esta cantidad de ovocitos, en la mayoría de los casos, corresponde al total de ovocitos MII obtenidos en la etapa anterior. Esta relación explica la alta correlación positiva entre las variables *INSEMINADOS* y *MII*. De los ovocitos inseminados, se registra el número que ha fecundado normalmente (FN), aquellos con algún grado de anormalidad (FA), o aquellos que no han fecundado. Estos datos están representados por las variables que se muestran en la tabla 4.5, siguiendo la relación descrita en la ecuación 4.3.

Tabla 4.5: Variables de caracterización de fecundación. Fuente: Elaboración propia.

Variable	Fecundación
2PN	Fecundación normal
1PN	Fecundación anormal
3PN	Fecundación anormal
CITOL	Fecundación anormal
NO FEC.	No fecundados

$$INSEMINADOS = 2PN + 1PN + 3PN + CITOL + NO FEC. \quad (4.3)$$

En consecuencia, en conjunto con el personal médico se ha optado por introducir una nueva variable, que fue denominada *RatioFecundacionNormal*, en lugar de las variables *INSEMINADOS* y *2PN*. Esta nueva variable se define como la proporción de ovocitos que han alcanzado la etapa de 2 pronúcleos (2PN) respecto al total de ovocitos inseminados (ver ecuación 4.4). Esta decisión va acompañada de la eliminación de las variables *1PN*, *3PN*, *CITOL* y *NOFEC*.

$$\text{RatioFecundacionNormal} = \frac{2PN}{INSEMINADOS} \quad (4.4)$$

Otra relación positiva notablemente alta se observa entre las variables *Blastos* y *2PN*. Esta correlación tiene sentido, ya que dentro del conjunto de ovocitos que logran una fecundación normal, existe un subgrupo que alcanza el grado de desarrollo embrionario conocido como blastocisto. Por lo tanto, a medida que aumenta el número de ovocitos que experimentan una fecundación exitosa, se espera que también aumente la cantidad de blastocistos desarrollados.

A raíz de esto, y con el fin de no perder información relevante y evitar problemas de multicolinealidad, y por recomendación del personal médico del IDIMI, es que se crea una nueva variable denominada *TasaBlastulacion*, que relaciona ambas variables como se muestra en la ecuación 4.5.

$$\text{TasaBlastulacion} = \frac{\text{Blastos}}{2PN} \quad (4.5)$$

Por otro lado, la variable *2PN* e *INSEMINADOS*, ambas independientes, En cuanto a la variable objetivo *EMBARAZO* notamos que esta tiene una mayor correlación con *Blastos*, *#TRANSF* y *EDAD* respectivamente.

4.3.2. Análisis grafico de distribuciones

Para comenzar, se realizó un gráfico de barras de la variable dependiente *EMBARAZO*, con el fin de ver el número de registros por clase que se tiene para esta.

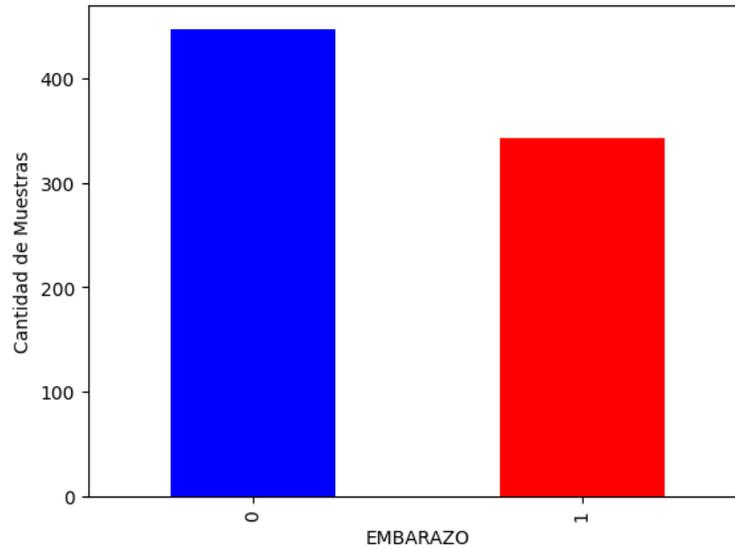


Figura 4.2: Número de muestras por etiqueta de la variable objetivo. Fuente: Elaboración propia.

El gráfico de la figura 4.2 muestra una presencia de 450 muestras para pacientes que no lograron el embarazo (56%) y 348 (44%) para los de embarazo positivo. Debido a que no existe un claro desbalance en los datos, se optará por no utilizar técnicas de balanceo de datos en este caso.

A continuación se extenderá el análisis al resto de variables. Para el caso de las variables continuas se observará mediante histograma, a excepción de variables que por sus características propias, sea mejor trabajarlas visualizarlas en gráficos de barras, como es el caso de las variables de tipo categóricas.

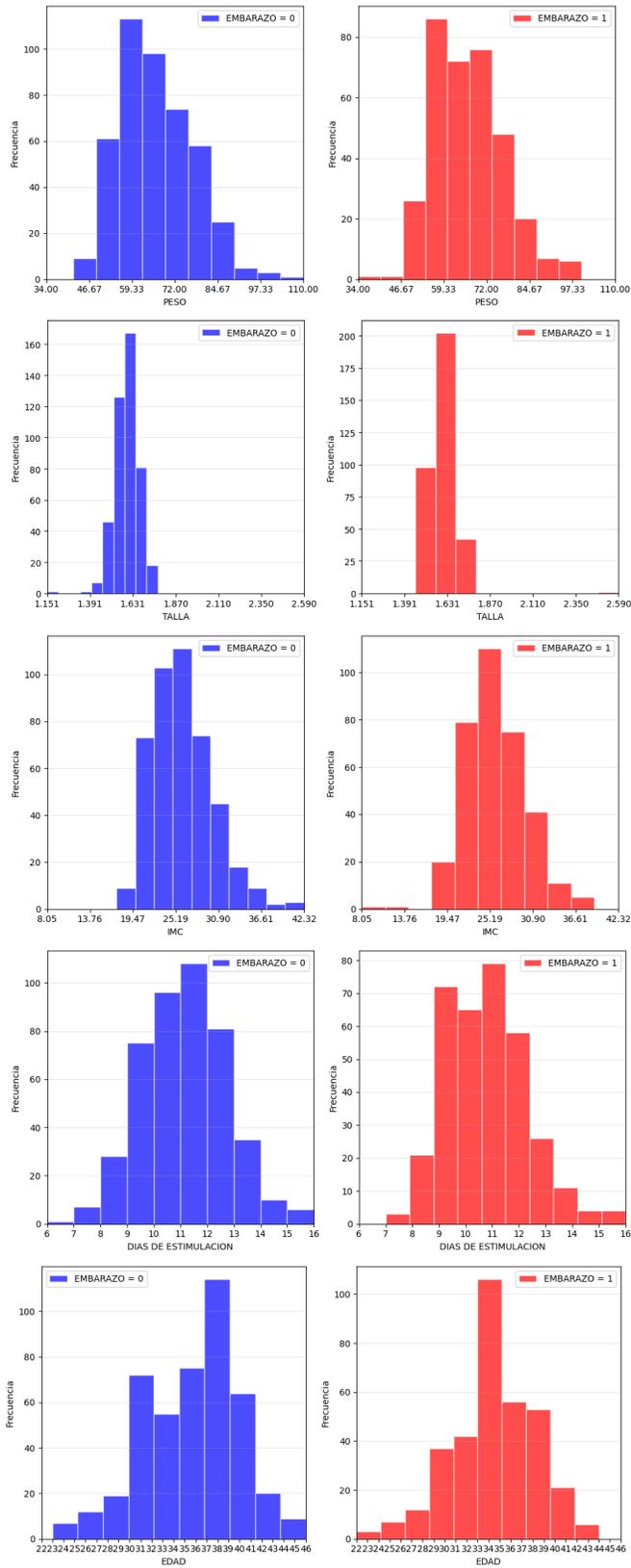


Figura 4.3: Histogramas de cada variable respecto a la variable objetivo.
Parte 1

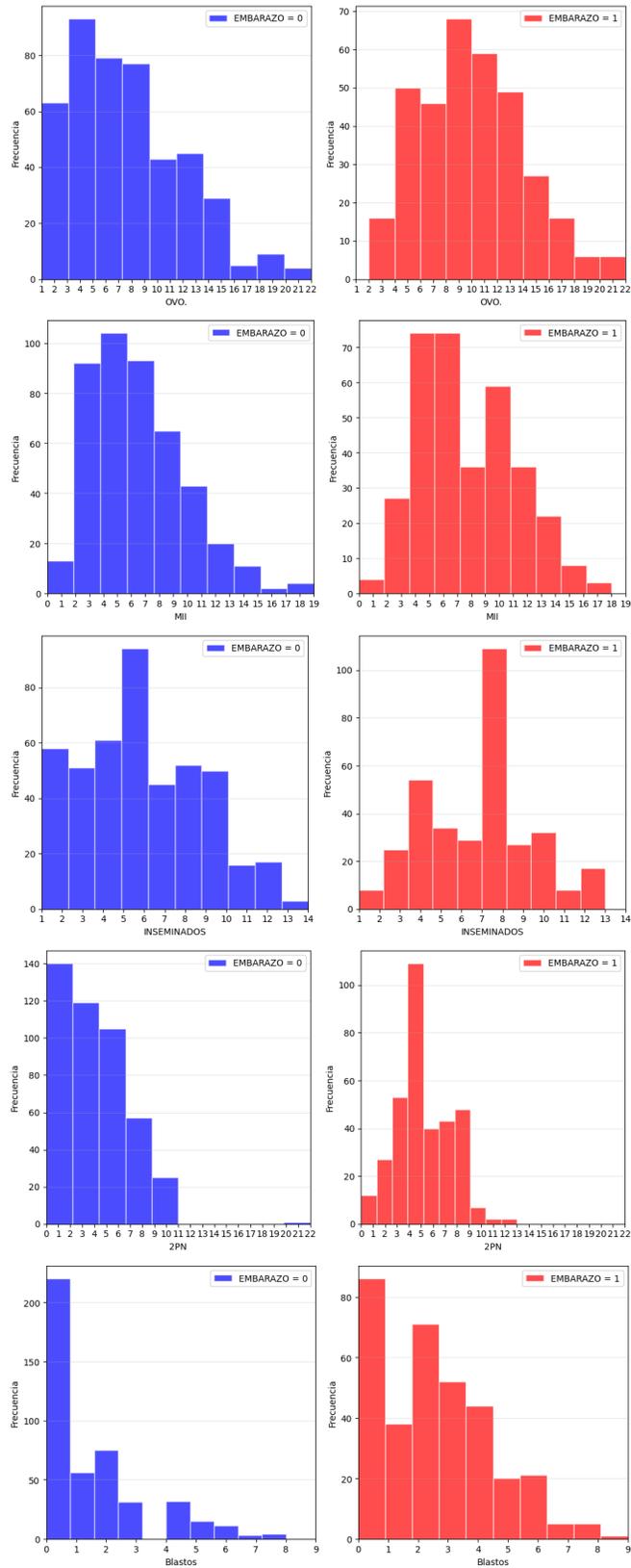


Figura 4.4: Histogramas de cada variable respecto a la variable objetivo.
Parte 2

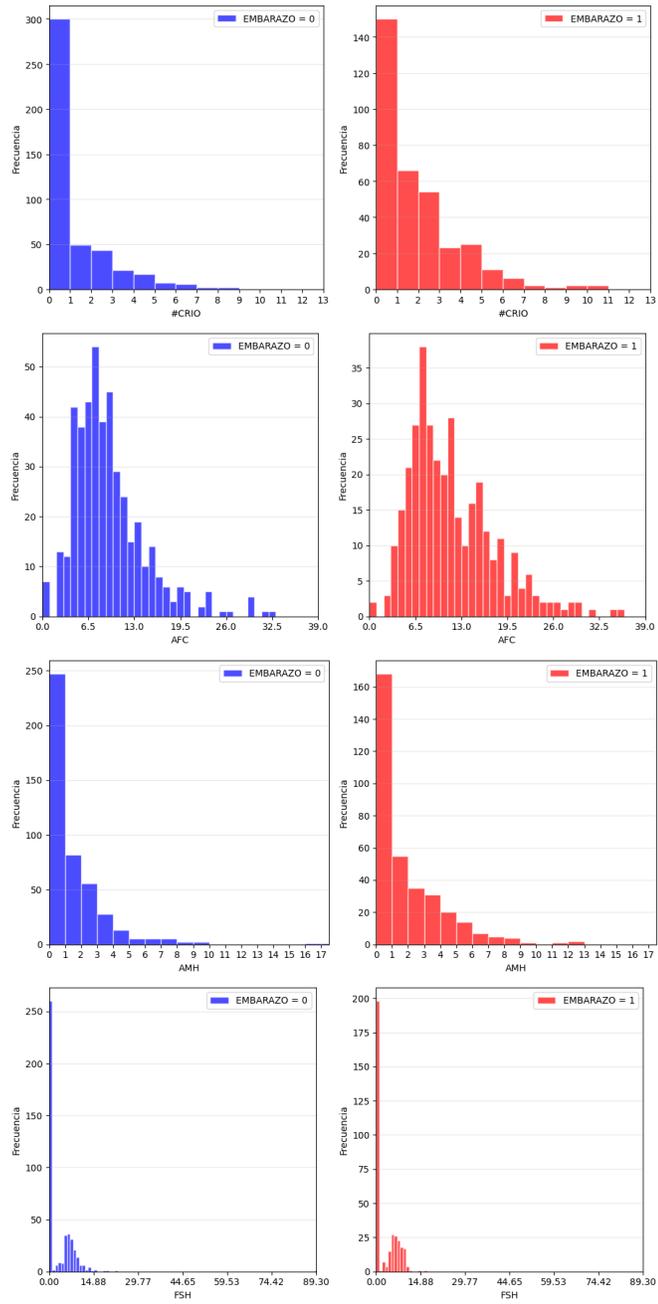


Figura 4.5: Histogramas de cada variable respecto a la variable objetivo.
Parte 3

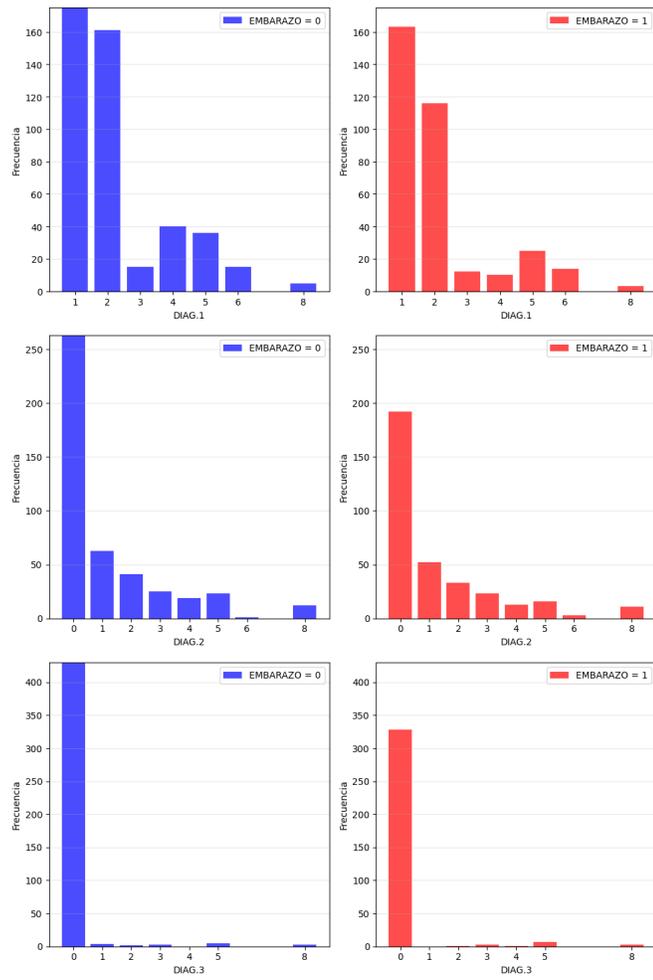


Figura 4.6: Gráficos de barra de cada variable respecto a la variable objetivo.
Parte 1

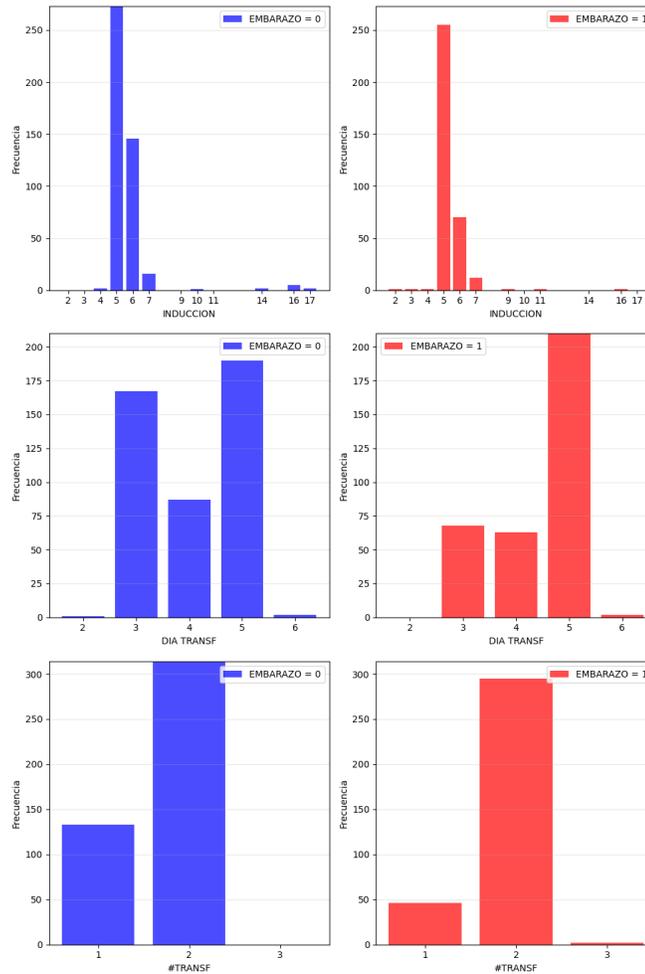


Figura 4.7: Gráficos de barra de cada variable respecto a la variable objetivo.
Parte 2

En relación al primer grupo de datos representado mediante histogramas en la figura 4.3, se identificó la presencia de *outliers* en las variables de *PESO*, *TALLA* e *IMC*, tanto para el grupo que logró el embarazo como para el que no lo logró. Dado que estos valores atípicos constituyen una proporción mínima del total de registros, se procederá a eliminarlos de la base de datos.

Aunque en la sección anterior, luego del análisis de correlaciones, se decidió en mantener únicamente la variable *IMC*, se optará por eliminar los *outliers* de las tres variables mencionadas. Esto se fundamenta en la posibilidad de que la presencia de datos extraños haya afectado la reconstrucción de algunos registros, ya que parte de los datos fueron reconstruidos en función de la relación entre estas variables. Por lo tanto, es importante destacar que los *outliers* detectados en estas variables no necesariamente corresponden al mismo ciclo, lo que podría distorsionar el análisis.

En cuanto a la variable que indica los días de estimulación, observamos que oscila entre 6 y 16 días, con la mayoría de los registros concentrados alrededor de los 11 días. Respecto a la edad de las pacientes, estas varían desde los 22 hasta los 46 años, destacándose una mayor participación de mujeres en el rango de edad entre los 30 y 37 años.

El segundo grupo de variables, en la figura 4.4 muestra la similitud entre las variables

OVO, *MII* e *INSEMINADOS*, tanto en distribución como en amplitud del intervalo, que oscila entre 0 y 22.

Por otro lado, la variable *2PN* presenta una pequeña cantidad de muestras para el grupo de embarazo negativo, que se escapa de la media de la característica. Estos registros serán eliminados, dado que comprenden solo un par de datos.

En cuanto a la cantidad de blastos, esta se mueve entre 0 y 9 blastos, donde se puede notar que ambos grupos, embarazo y no embarazo, tienen como mayoría cantidad de blastos cercanos a 0.

La Hormona Antimülleriana (AMH), el recuento de folículos antrales (AFC) y la Hormona Folículo Estimulante (FSH) son importantes biomarcadores utilizados en la evaluación de la reserva ovárica y la planificación de los tratamientos FIV. Estas variables presentes en la figura 4.5, cuentan con la presencia de *outliers*, donde en esta última existe solo un registro que se escapa completamente del resto de información. Por lo que será eliminada.

Continuando con el análisis gráfico, se puede notar que la cantidad de embriones criopreservados, representado por la variable *CRIO* tiene bastante presencia con valor 0, en el grupo que no logró el embarazo, pese a que este es el que presenta menos registros totales.

Por último, el grupo de variables categóricas que almacena la información de los 3 diagnósticos de infertilidad que se les registra a los pacientes (ver figura 4.6, muestra una presencia significativa de registros con diagnóstico 0, es decir, que no poseen un registro de diagnóstico asociado, en las variables *DIAG.2* y *DIAG.3*. Por lo que estas variables no están entregando información relevante adicional, es por esto que se eliminarán del modelo, y en representación de estas características solo quedará la variable *DIAG.1*

4.4. Desarrollo de modelos

4.4.1. Datos a utilizar

Luego de la etapa de preprocesamiento que implicó la transformación o eliminación de variables, se determinó como set de datos final el que se muestra en la tabla 4.6 junto con sus definiciones, conformado por 15 variables y 780 registros.

Tabla 4.6: Set de variables final. Fuente: Elaboración propia.

Nombre variable	Descripción
IMC	Índice de masa corporal (kg/m ²)
DIAS DE ESTIMULACION	Días de estimulación ovárica
EDAD	Edad paciente en años
DIAG.1	Diagnóstico principal de infertilidad paciente
INDUCCION	Categoría inducción
MII	Número de ovocitos MII obtenidos
RatioFecundacionNormal	Relación ovocitos fecundados normales sobre total inseminados
#TRANSF	Número de embriones transferidos
Categoria_1	Igual a 1 si el primer embrión transferido es de calidad A o B. 0 si no.
TasaBlastulación	Relación número de blastos y total fecundados normales
EMBARAZO	Igual a 1 si se logró el embarazo. 0 si no.
AFC	Cantidad de folículos antrales
AMH	Nivel de hormona antimülleriana
FSH	Indicador de hormona folículo estimulante

Todos los modelos desarrollados en este estudio han seguido una metodología uniforme para garantizar la coherencia y permitir comparaciones directas entre ellos. En primer lugar, los datos fueron divididos en conjuntos de entrenamiento y prueba, manteniendo la integridad de los grupos por paciente mediante la técnica de StratifiedGroupKFold. Este enfoque garantiza que los datos de un mismo paciente no se mezclen entre los conjuntos de entrenamiento y prueba, minimizando así el riesgo de fuga de datos y obteniendo una evaluación más realista del rendimiento de los modelos.

Posteriormente, se llevó a cabo una normalización de las variables numéricas y una codificación de las variables categóricas, utilizando un pipeline que integra el StandardScaler y el OneHotEncoder con el parámetro `drop='first'` para evitar la multicolinealidad. La optimización de los hiperparámetros se realizó mediante GridSearchCV con validación cruzada estratificada, utilizando diferentes universos de hiperparámetros adaptados a la naturaleza de cada modelo. El F1 Score se utilizó como métrica principal para la optimización, dado su balance entre precisión y recall, crucial en problemas de clasificación donde ambas métricas son importantes.

4.4.2. Resultados

Los resultados de las métricas para todos los modelos implementados se muestran a continuación.

Tabla 4.7: Resultados de los modelos. Fuente: Elaboración propia.

Model	AUC ROC	F1 Score	Accuracy	Precision	Recall
Random Forest	0.75	0.68	0.70	0.72	0.64
Gradient Boosting	0.67	0.60	0.61	0.61	0.59
Extreme Gradient Boosting	0.73	0.64	0.67	0.70	0.59
Super Vector Machine	0.73	0.67	0.66	0.64	0.69
Red Neuronal 1 ⁽¹⁾	0.81	0.75	0.75	0.73	0.77
Red Neuronal 2 ⁽²⁾	0.82	0.75	0.73	0.71	0.79

⁽¹⁾ Corresponde a la red neuronal con la arquitectura inicial

⁽²⁾ Corresponde a la red neuronal con validación cruzada

Después de evaluar el desempeño de todos los modelos, se observa que el modelo Random Forest obtuvo el mejor rendimiento entre los modelos no neuronales, con un AUC ROC de 0.75 y una precisión de 0.72. A continuación, se presenta una descripción detallada de la configuración utilizada para este modelo. Además, dado que los modelos neuronales lograron los mejores resultados generales, alcanzando un AUC ROC de 0.81 y 0.82 respectivamente, también se describe la configuración de las dos arquitecturas de redes neuronales implementadas, incluyendo la inicial y la versión con validación cruzada.

4.4.2.1. Random Forest (mejor modelo)

EL mejor modelo de clasificación no neuronal, el Random Forest, siguió la metodología explicada anteriormente para asegurar la consistencia en la evaluación y comparación de los modelos. La búsqueda de hiperparámetros se centró en maximizar el F1 Score. A continuación se muestra el universo de parámetros estudiados y sus valores.

1. **Número de Estimadores (n_estimators):** Este parámetro controla el número total de árboles en el ensemble. Se evaluaron valores entre 50 y 150 para encontrar un equilibrio óptimo entre el ajuste del modelo y el riesgo de sobreajuste.
2. **Profundidad Máxima (max_depth):** Limita la profundidad de los árboles individuales, afectando la capacidad del modelo para captar patrones complejos y la propensión al sobreajuste. Se probaron profundidades de 10, 20, y 30.
3. **Mínimo de Muestras por División (min_samples_split):** Determina el número mínimo de muestras necesarias para dividir un nodo. Se evaluaron valores de 2, 5, y 10 para ajustar la sensibilidad del modelo a las características del conjunto de datos.
4. **Mínimo de Muestras por Hoja (min_samples_leaf):** Define el número mínimo de muestras que debe haber en una hoja final. Se evaluaron valores de 1, 2, y 4 para controlar la granularidad del modelo.
5. **Bootstrap:** Indica si se deben usar muestras de bootstrap para construir cada árbol. Se probaron dos configuraciones: True y False, para evaluar el impacto de la técnica de muestreo en el rendimiento del modelo.

Los resultados de los hiperparámetros obtenidos para el mejor modelo se muestran en la tabla 4.8

Tabla 4.8: Mejores Hiperparámetros modelo Random Forest. Fuente: Elaboración propia.

Hiperparámetro	Valor
n_estimators	70
max_depth	20
min_samples_split	5
min_samples_leaf	1
Bootstrap:	True

El gráfico resumen de SHAP (ver figura 4.8) muestra la importancia relativa de las características en las predicciones del modelo, con las variables más influyentes ubicadas en la parte superior. De todas las características incluidas, la Tasa de Blastulación se destaca como la de mayor impacto. Los valores altos de esta variable tienden a aumentar significativamente la probabilidad de éxito en el tratamiento, ya que los puntos rojos (valores altos) se encuentran predominantemente hacia el lado derecho del gráfico, indicando una contribución positiva. Esto es congruente con la evidencia clínica que sugiere que una alta tasa de blastulación está asociada con mejores resultados en los tratamientos de fertilidad.

Otra característica crucial es el Recuento de Folículos Antrales (AFC), cuyos valores altos también están relacionados con un aumento en la probabilidad de embarazo. Esto se refleja en los puntos rojos, que indican que pacientes con un mayor recuento de folículos tienen una mayor probabilidad de éxito.

La Edad de la paciente, como era de esperarse, tiene un impacto significativo en el modelo. Los valores altos (más cercano a los 40 años) de esta característica (indicados en rojo) disminuyen considerablemente la probabilidad de éxito, ya que los puntos tienden a estar hacia el lado izquierdo del eje de valores SHAP. Este comportamiento refleja la relación bien documentada entre la edad avanzada y una menor tasa de éxito en los tratamientos de fertilización in vitro (IVF). El modelo captura correctamente este patrón, otorgando un mayor peso a la edad como factor clave para predecir los resultados de embarazo.

Otras variables como los Ovocitos Maduros (MII), los niveles de AMH (Hormona Antimülleriana) y el Número de embriones transferidos también muestran una influencia significativa en el modelo. En particular, altos valores de AMH están asociados positivamente con la probabilidad de éxito, lo que es coherente con su uso clínico como marcador de reserva ovárica. Por otro lado, el número de embriones transferidos muestra una relación más compleja con el resultado, ya que la dispersión de los puntos en el gráfico indica que tanto valores altos como bajos pueden tener efectos variables según el contexto, recordar que los valores de esta variable oscilan entre 1 y 3 embriones.

VARIABLES COMO EL Índice de Masa Corporal (IMC), el número de días de estimulación y otros indicadores más específicos relacionados con la inducción o diagnósticos previos (como INDUCCION_5 o DIAG_1_4) tienen un impacto más moderado, con valores SHAP cercanos a cero en la mayoría de los casos. Esto sugiere que, si bien estas características pueden aportar cierta información, su contribución a la predicción del modelo es menor comparada con las variables clínicas principales como la Tasa de Blastulación, AFC y Edad.

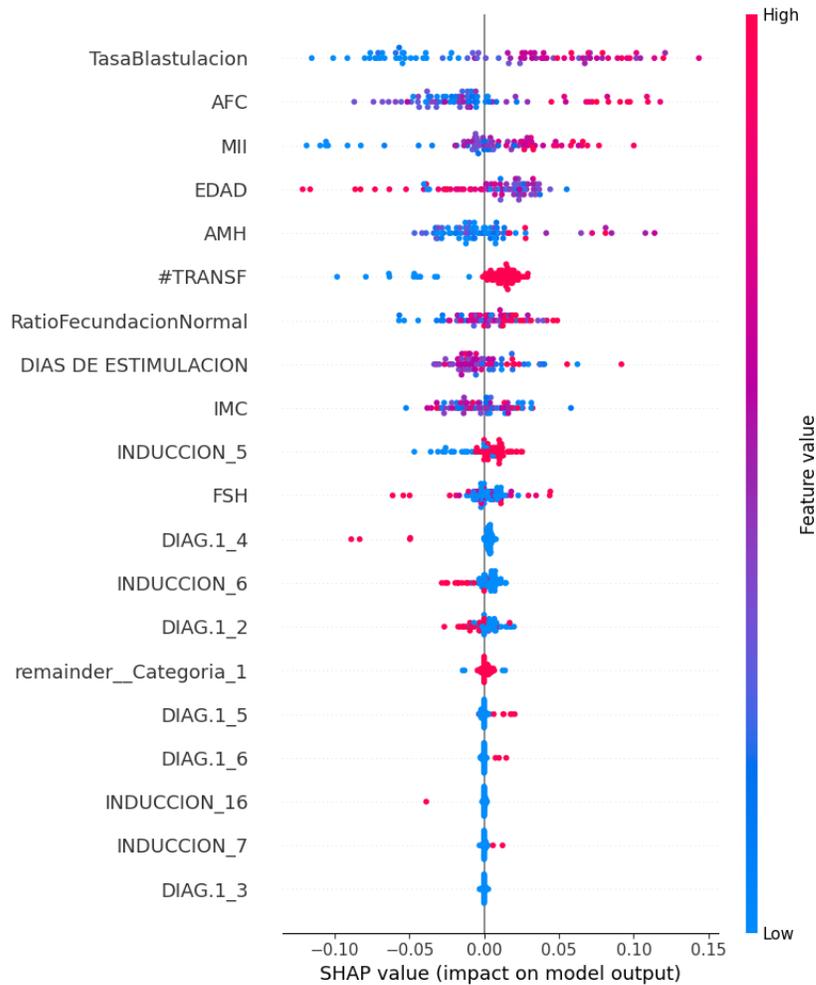


Figura 4.8: Resultados algoritmo SHAP modelo Random Forest. Fuente: Elaboración propia.

4.4.2.2. Redes neuronales

Con el fin de explorar otros enfoques, se implementaron dos redes neuronales densamente conectadas. Ambas redes fueron diseñadas con capas ocultas, regularización mediante dropout, optimizador Adam y función de activación ReLU. No obstante, cada red varía en su enfoque de validación y en los hiperparámetros de configuración. Como se detalló en los alcances, la interpretabilidad de estos modelos queda fuera para este trabajo. A continuación se muestra cada configuración y se analiza su relación.

Primera Red Neuronal: Arquitectura Inicial

La primera red fue diseñada para abordar el problema de clasificación binaria (éxito o no del embarazo) con una arquitectura sencilla pero efectiva. La red consta de tres capas densas ocultas seguidas de una capa de salida. Cada capa oculta utiliza la función de activación ReLU, mientras que la capa de salida emplea una función sigmoide para generar probabilidades que permiten tomar decisiones de clasificación. La estructura de la red es la siguiente:

1. **Primera capa oculta:** 32 unidades (neuronas) con activación ReLU.

2. **Segunda capa oculta:** 16 unidades con activación ReLU.
3. **Tercera capa oculta:** 8 unidades con activación ReLU.
4. **Capa de salida:** 1 unidad con activación sigmoide.

La regularización de esta red se llevó a cabo mediante la implementación de un Dropout con un 20 % de probabilidad de desactivación en las capas ocultas, con el objetivo de reducir el sobreajuste. El modelo fue entrenado utilizando el optimizador Adam, conocido por ser eficiente en la mayoría de los problemas de clasificación, ya que ajusta los pesos de manera adaptativa durante el proceso de entrenamiento.

En cuanto al entrenamiento, se utilizó el 90 % para esta tarea, mientras que el 10 % restante se utilizó para la validación. Se implementó la técnica de early stopping para detener el entrenamiento cuando no hubo mejora en la pérdida de validación durante 30 épocas consecutivas. Este enfoque ayudó a prevenir el sobreentrenamiento.

Segunda Red Neuronal: Configuración Mejorada con Validación Cruzada

La segunda red neuronal, aunque basada en la primera, incorpora ajustes clave, entre ellos la implementación de validación cruzada estratificada (Stratified K-Fold), lo que mejora la capacidad de evaluar el rendimiento del modelo de manera más robusta. Además, se optimizó la estructura de la red neuronal, manteniendo la misma función de activación y optimizador, pero ajustando el número de neuronas en las capas y utilizando un número diferente de unidades en la capa oculta. La estructura de la red es la siguiente:

1. **Primera capa oculta:** 25 unidades con activación ReLU.
2. **Segunda capa oculta:** 20 unidades con activación ReLU.
3. **Tercera capa oculta:** 10 unidades con activación ReLU.
4. **Capa de salida:** 1 unidad con activación sigmoide.

Al igual que en la primera red, se utilizó un Dropout del 20 % en cada capa oculta, reforzando la capacidad del modelo para generalizar mejor y evitar el sobreajuste. Además, se mantuvo el optimizador Adam, que ha demostrado ser eficaz para problemas de clasificación binaria.

Una de las diferencias clave de esta red es el uso de *Stratified K-Fold Cross Validation*, con 5 particiones (folds). En cada iteración, el conjunto de datos se divide en cinco subconjuntos, utilizando cuatro de ellos para el entrenamiento y el restante para la validación. Este enfoque asegura que la proporción de clases se mantenga en cada fold, proporcionando evaluaciones más representativas de la capacidad del modelo.

Una vez completada la validación cruzada, el modelo fue entrenado y probado en el conjunto de datos de prueba, que representó el 10 % de los datos. El early stopping también se aplicó aquí, deteniendo el entrenamiento si no se observaba mejora en la pérdida de validación después de 30 épocas.

Capítulo 5

Discusión

5.1. Interpretación de los resultados generales

En términos generales, los modelos de machine learning no neuronales (Random Forest, Gradient Boosting, XGBoost y SVM) lograron un rendimiento sólido, en particular el modelo de Random Forest, que mostró un AUC ROC de 0.75 y un equilibrio entre precisión y recall. Esto es coherente con la naturaleza del problema, ya que Random Forest es eficaz en la identificación de patrones complejos en datos no lineales, especialmente cuando hay variables con relaciones complejas. Tanto Gradient Boosting como XGBoost tuvieron un rendimiento levemente inferior en comparación con Random Forest. El AUC ROC fue de 0.67 y 0.73, respectivamente. Si bien estos modelos son generalmente más flexibles y potentes, en este caso específico pueden haber sido más sensibles al ruido en los datos, lo que explica los resultados algo más bajos en comparación con Random Forest. Sin embargo, su capacidad para modelar relaciones no lineales complejas también es un factor relevante, lo cual se refleja en su rendimiento aceptable. Finalmente, el modelo de SVM también obtuvo un AUC ROC moderado (0.72), lo cual es esperable dado que el kernel sigmoid fue usado, lo que permite capturar ciertas no linealidades. Sin embargo, los resultados ligeramente más bajos en términos de precisión (0.69) pueden deberse a la dificultad de ajustar este modelo de manera óptima para datos más complejos o con más ruido.

Respecto a los resultados de las redes neuronales, la primera de ellas mostró un rendimiento razonable (AUC ROC de 0.81), lo que indica que logró capturar patrones más complejos de los datos que los modelos basados en boosting y SVM, y fue similar a Random Forest. Sin embargo, la variabilidad en precisión y recall sugiere que podría ser sensible a pequeñas variaciones en los datos, lo que podría estar asociado a la falta de optimización a través de técnicas como la validación cruzada.

5.2. Importancia de las variables predictoras

El análisis de importancia de las variables a través de SHAP en el modelo Random Forest mostró resultados alineados con la literatura clínica. La tasa de blastulación, la edad de la paciente y el Recuento de Folículos Antrales (AFC) emergieron como las variables más relevantes para predecir el éxito de un tratamiento.

La tasa de blastulación fue consistentemente la variable más importante en todos los modelos, incluido el Random Forest. Este indicador refleja la calidad y viabilidad del embrión,

ya que un mayor número de embriones que alcanzan la fase de blastocisto aumenta el potencial de implantación, y por ende, la probabilidad de éxito del embarazo. Los resultados de SHAP muestran una correlación positiva entre una alta tasa de blastulación y el éxito del tratamiento, lo que subraya su relevancia en la toma de decisiones clínicas.

Estos hallazgos también refuerzan estudios previos que destacan la edad como un factor clave en la capacidad reproductiva y el AFC como un indicador confiable de la reserva ovárica. La visualización proporcionada por SHAP clarifica que tanto un aumento en el AFC como una tasa de blastulación alta aumentan la probabilidad de éxito, mientras que la edad avanzada reduce significativamente esta probabilidad.

Aunque modelos como Gradient Boosting y Extreme Gradient Boosting (XGBoost) ofrecieron un rendimiento aceptable en términos de AUC ROC y F1 Score, estos modelos son generalmente más sensibles a los hiperparámetros y al ruido en los datos, lo que puede limitar su capacidad de generalización. Además, aunque los modelos de redes neuronales como los implementados en este estudio mostraron un rendimiento competitivo, carecen de la misma capacidad de interpretabilidad que ofrece SHAP en combinación con Random Forest. Esto es especialmente relevante en un contexto clínico, donde es vital no solo predecir con precisión, sino también entender qué factores están influyendo en las decisiones del modelo. Por ello, el uso de SHAP con Random Forest ofrece una ventaja significativa, al permitir que el personal médico comprenda qué aspectos son más relevantes para maximizar las probabilidades de un tratamiento exitoso, contribuyendo a un enfoque más personalizado y efectivo.

5.3. Limitaciones y consideraciones

A pesar de los resultados prometedores obtenidos en este estudio, es importante señalar algunas limitaciones que podrían haber influido en los hallazgos. En primer lugar, el tamaño de la muestra utilizada (aproximadamente 780 casos) es relativamente limitado, lo que podría afectar la capacidad de generalización de los modelos, particularmente aquellos más complejos como las redes neuronales. Un conjunto de datos más grande podría mejorar el rendimiento y la robustez de estos modelos, al permitirles captar patrones más complejos y reducir el riesgo de sobreajuste.

Además, la naturaleza de los datos recolectados podría haber introducido sesgos inherentes que influyen en la capacidad predictiva de los modelos. Variables no incluidas en el conjunto de datos, como factores socioeconómicos o antecedentes de salud más detallados, podrían jugar un papel importante en los resultados de los tratamientos de fertilización asistida. La omisión de estas variables puede haber limitado la capacidad de los modelos para ofrecer una visión más integral y precisa de los factores que influyen en el éxito de los tratamientos.

Una limitación adicional es que los modelos actuales no consideran explícitamente si el paciente ha tenido ciclos previos de tratamientos de fertilización asistida. Este es un factor crítico, ya que la experiencia y los resultados de ciclos anteriores podrían afectar significativamente las probabilidades de éxito en ciclos futuros. Incluir esta variable en análisis futuros podría mejorar sustancialmente la capacidad predictiva de los modelos.

En cuanto a la interpretabilidad, si bien los modelos avanzados como las redes neuronales mostraron un buen rendimiento predictivo, su capacidad para explicar de manera clara las decisiones tomadas es menor en comparación con modelos más interpretables como el

Random Forest. Esta falta de transparencia representa un desafío cuando se pretende aplicar estos resultados en un entorno clínico, donde es fundamental que los especialistas comprendan por qué y cómo se toman ciertas decisiones para confiar en las recomendaciones del modelo y adaptarlas a cada paciente. Por esto, explorar la opción de entender a profundidad como realizan las predicciones estos modelos es fundamental.

Capítulo 6

Impacto social y económico

6.1. Impacto económico

El impacto económico de la implementación de modelos predictivos basados en machine learning en tratamientos de fertilización asistida es altamente significativo, especialmente considerando los costos asociados con estos tratamientos en el sistema de salud pública chileno. Para demostrar este impacto, se utilizó como referencia el financiamiento otorgado por el Ministerio de Salud al Instituto de Investigación Materno Infantil (IDIMI) en los años 2019 y 2020 para tratamientos de fertilización asistida de alta complejidad (IVF)[53]. Durante esos años, el IDIMI recibió un total de \$620 116 400 pesos chilenos, destinados a 292 ciclos de tratamiento (234 correspondientes al primer ciclo y 58 a ciclos de continuidad). El costo promedio por ciclo fue de \$2 467 440 para los primeros intentos y \$485 630 para ciclos subsecuentes. Este financiamiento refleja la inversión pública directa para cubrir los tratamientos de fertilización asistida.

Tabla 6.1: Financiamiento otorgado por FONASA al IDIMI para tratamientos de fertilización asistida de alta complejidad en 2019. Fuente: [53].

Tipo de Cupos	Nº	Monto por cupo (\$)	Monto total (\$)
Nuevos Primer Ciclo (I intento)	234	2 467 440	577 380 960
Nuevos Segundo Ciclo (II intento)	58	485 630	28 166 540
Continuidad Primer Ciclo año anterior	30	485 630	14 568 900
Total	322		620 116 400

Se estima que entre el 30 % y el 40 % de los tratamientos realizados durante este período resultaron en fallos. Esto implica que un número considerable de recursos destinados a los tratamientos no condujo a un embarazo exitoso. Aplicando este porcentaje de fallos a los ciclos financiados en 2019:

1. **Número de tratamientos fallidos (30 %-40 %):** De los 292 ciclos financiados, se estima que entre 88 y 117 tratamientos no alcanzaron un embarazo exitoso.
2. **Costo estimado de ciclos fallidos:** Dado que cada ciclo fallido tuvo un costo de \$2 467 440, el gasto asociado con los ciclos fallidos en 2019 varió entre \$217 135 000 y

\$288 179 000 pesos chilenos.

- 3. Proyección de crecimiento del financiamiento al 2023:** Con el aumento de la prevalencia de la infertilidad y el retraso en la maternidad, se estima que la demanda por tratamientos de fertilización asistida ha crecido. Asumiendo un crecimiento del 10 % en los ciclos financiados entre 2019 y 2023, se estima que en 2023 se financiarían aproximadamente 321 ciclos
- 4. Ajuste por inflación:** Considerando los efectos de la inflación entre 2019 y 2023 (ver figura 6.1, y utilizando los índices de precios al consumidor reales de esos años (131.9 en 2019 y 170.7 en 2023), los costos por ciclo en 2023 habrán aumentado aproximadamente un 29.4%. Al ajustar los montos otorgados en 2019 por esta inflación acumulada, el financiamiento necesario en 2023 para cubrir un ciclo de fertilización in vitro (IVF) sería de aproximadamente \$3 192 570 pesos chilenos.
- 5. Financiamiento total proyectado en 2023:** Si consideramos el financiamiento para 321 ciclos al costo ajustado por inflación, se necesitaría un total aproximado de \$1024 millones de pesos chilenos para cubrir estos tratamientos en 2023.

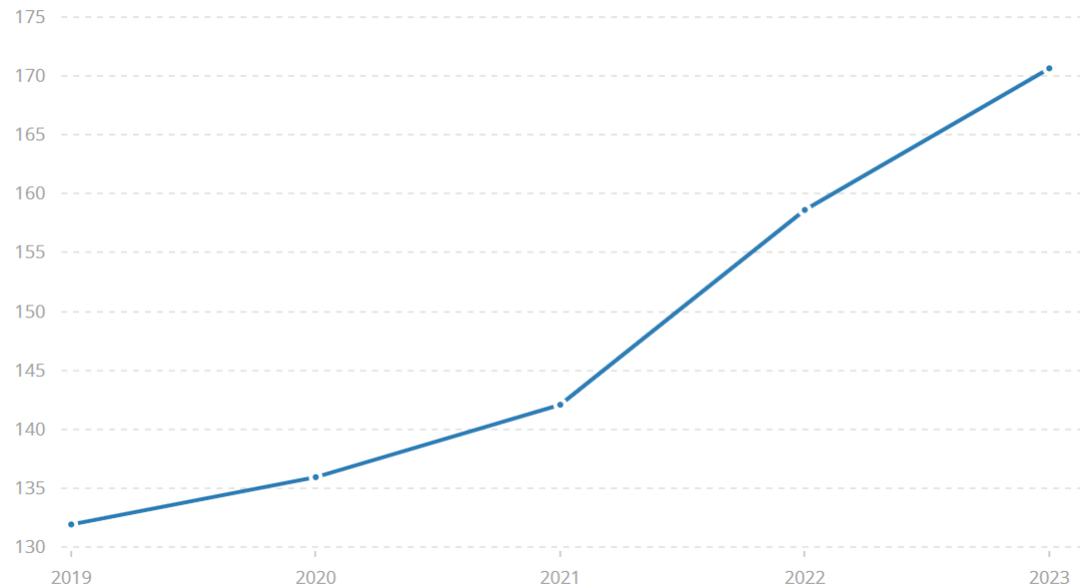


Figura 6.1: Índice de precios al consumidor 2019–2023. Fuente: Banco mundial.

Con todo esto, la implementación de modelos de machine learning tiene el potencial de mejorar la tasa de éxito en los tratamientos de fertilización asistida. Si asumimos una reducción de los ciclos fallidos en un 10 % a 15 % con su implementación. Aplicando esta reducción a los ciclos fallidos estimados en 2019 (88–117 ciclos), podríamos evitar entre 9 y 18 ciclos fallidos al año, lo que se traduciría en un ahorro estimado de entre 29 y 57 millones de pesos chilenos en 2023.

6.2. Impacto social

La implementación de modelos de machine learning para predecir el éxito de un embarazo en tratamientos de fertilización asistida no solo tendrá un impacto económico significativo, sino también un profundo impacto social, tanto para los pacientes como para el sistema de salud público.

1. **Mejora de la Equidad en el Acceso a Tratamientos:** Al integrar estos modelos predictivos en los servicios de salud pública, se espera reducir las disparidades en el acceso a tratamientos efectivos de fertilización asistida. Los modelos proporcionarán recomendaciones personalizadas, independientemente del hospital o región, asegurando que los pacientes de zonas rurales o con menos acceso a especialistas reciban el mismo nivel de precisión en la predicción de éxito de los tratamientos que aquellos en zonas urbanas. Esta estandarización puede ayudar a cerrar la brecha en el acceso a tratamientos de calidad, permitiendo que más personas tengan oportunidades de éxito, sin importar su ubicación geográfica o nivel socioeconómico.
2. **Aumento de las Tasas de Éxito en los Tratamientos:** La capacidad de los modelos de machine learning para identificar las variables más relevantes, como la tasa de blastulación, el recuento de folículos antrales (AFC), y la edad de la paciente, permite una predicción más precisa de los ciclos de tratamiento más efectivos. Esto significa que los pacientes podrán beneficiarse de recomendaciones basadas en datos reales que mejoran significativamente las probabilidades de éxito en cada intento de fertilización. Al reducir los intentos fallidos, no solo se reduce la carga emocional en los pacientes, sino que también se optimiza el uso de los recursos clínicos, aumentando la eficiencia general del sistema.
3. **Reducción de Costos para los Pacientes y el Sistema de Salud:** Al predecir con mayor precisión qué pacientes tienen mayores probabilidades de éxito en los tratamientos y qué características específicas de los ciclos son más efectivas, los modelos ayudarán a reducir la cantidad de ciclos de tratamiento innecesarios. Esto se traducirá en una disminución de los costos tanto para los pacientes, quienes evitarán gastos adicionales en ciclos fallidos, como para el sistema de salud. Cada ciclo evitado representa un ahorro sustancial, tanto en términos económicos como en términos de recursos médicos y tiempo de los profesionales de salud.
4. **Contribución a la Sostenibilidad del Sistema de Salud:** La reducción de ciclos fallidos también contribuirá a la sostenibilidad a largo plazo del sistema de salud. Al utilizar machine learning para optimizar los tratamientos, el sistema de salud pública podrá manejar de manera más eficiente la creciente demanda de servicios de fertilización asistida, derivada del envejecimiento de la población y el retraso en la maternidad. Esto permitirá a las instituciones públicas planificar y asignar recursos de manera más efectiva, mejorando el acceso a estos tratamientos a largo plazo y asegurando un uso más racional de los recursos disponibles.

Además, los modelos predictivos no solo optimizan los tratamientos en términos clínicos, sino que también aportan transparencia y confianza a los pacientes y especialistas, al ofrecer explicaciones claras sobre las variables más importantes que afectan el éxito del tratamiento.

Capítulo 7

Conclusiones

7.1. Desarrollo metodológico

7.1.1. Importancia de las variables predictoras

Uno de los aportes más significativos de este trabajo es la identificación de las variables más influyentes en la predicción del éxito de los tratamientos de fertilización asistida. A través de técnicas como SHAP, se confirmó que la tasa de blastulación fue la variable más relevante en todos los modelos. Esta tasa refleja la calidad del embrión y su capacidad para desarrollarse hasta la fase de blastocisto, siendo un indicador clave de éxito en la implantación. Además, se destacó el papel del Recuento de Folículos Antrales (AFC), que mide la reserva ovárica de la paciente, y la edad, que sigue siendo uno de los principales factores limitantes en los tratamientos de fertilización. La relación inversa entre la edad avanzada y la tasa de éxito sugiere que cuanto antes se intervenga, mayor será la probabilidad de éxito, alineándose con la literatura existente. Estos hallazgos no solo contribuyen a mejorar la predicción de los resultados, sino que también proporcionan una base sólida para la toma de decisiones clínicas, permitiendo personalizar los tratamientos.

7.1.2. Validación cruzada y su importancia en la predicción de resultados clínicos

A lo largo de este estudio, se demostró que la validación cruzada es una técnica esencial en la evaluación del rendimiento de los modelos de machine learning, especialmente cuando se trabaja con conjuntos de datos pequeños. El uso de validación cruzada permitió mejorar la estabilidad de los resultados obtenidos al mitigar el riesgo de sobreajuste, garantizando que los modelos pudieran generalizar mejor en datos no vistos. En el ámbito clínico, donde los resultados predictivos deben ser lo más precisos posible, esta metodología fue particularmente útil para asegurar que los modelos no solo tuvieran un buen desempeño en los datos de entrenamiento, sino que también fueran robustos y confiables en el uso real.

7.2. Desempeño de los modelos y su relevancia en el contexto clínico

Los resultados obtenidos muestran que la Red Neuronal con validación cruzada fue el modelo con mejor desempeño, logrando un AUC ROC de 0.82 y un F1 Score de 0.75. En el caso

de los modelos no neuronales, el Random Forest destacó con un AUC ROC de 0.75 y un F1 Score de 0.68. Estos resultados son especialmente significativos en el contexto clínico, ya que métricas más altas sugieren una mayor precisión en la identificación de pacientes con mayores probabilidades de éxito en los tratamientos de fertilización asistida de alta complejidad.

Los modelos de caja negra, como las redes neuronales, aunque son menos interpretables, han demostrado un gran potencial en la generación de predicciones precisas. Este tipo de modelos puede implementarse para mejorar la toma de decisiones clínicas, aportando beneficios en la personalización de los tratamientos, aunque su uso debe complementarse con otros enfoques que ayuden a explicar sus predicciones.

Por otro lado, modelos como Random Forest, que ofrecen un equilibrio entre rendimiento e interpretabilidad, también tienen un valor considerable. Si bien no logran la misma precisión que las redes neuronales, su capacidad para ofrecer explicaciones claras sobre los factores que influyen en el éxito del tratamiento los convierte en herramientas útiles cuando es necesario un análisis más detallado. En particular, la capacidad de estos modelos para identificar variables clave permite a los clínicos tomar decisiones informadas basadas en las características más influyentes del paciente.

7.3. Impacto económico y social del uso de modelos predictivos

La implementación de modelos de machine learning en tratamientos de fertilización asistida no solo tiene un impacto económico positivo, al reducir los costos asociados a los ciclos fallidos, sino también un impacto social significativo. Desde un punto de vista económico, se estimó que la reducción de los ciclos fallidos podría traducirse en un ahorro anual de entre 29 y 57 millones de pesos chilenos para el sistema de salud. Esta optimización de recursos permitiría una mayor eficiencia en el uso de los fondos destinados a los tratamientos de alta complejidad, facilitando que más pacientes accedan a estos tratamientos. En cuanto al impacto social, la reducción de ciclos fallidos no solo aliviaría la carga económica en los pacientes, sino también el impacto emocional que estos tratamientos suelen tener. Al mejorar la precisión en la predicción de resultados y personalizar los tratamientos, se podría reducir la cantidad de ciclos necesarios para lograr un embarazo exitoso, lo que aumentaría significativamente la satisfacción de los pacientes y mejoraría su calidad de vida.

7.4. Conclusiones del trabajo

Se ha demostrado que es posible determinar la probabilidad de éxito de un tratamiento de alta complejidad utilizando modelos de machine learning, cumpliendo así con la hipótesis planteada. Los modelos desarrollados han logrado identificar las variables más relevantes y predecir con alta precisión el éxito de los tratamientos de fertilización asistida.

Sin embargo, pese a los resultados prometedores obtenidos, aún existe un espacio significativo para la mejora. Los modelos implementados han demostrado ser moderadamente buenos, pero no excelentes, en términos de precisión y robustez. Esto indica que, si bien los algoritmos propuestos son útiles para apoyar la toma de decisiones clínicas, aún no alcanzan el nivel de confianza que se requiere para ser considerados herramientas definitivas en un entorno clínico. Por esto, profundizar y continuar con estudios en esta área es fundamental

para establecer una herramienta útil y eficaz para el personal médico de asignación de tratamientos de fertilidad asistida.

7.5. Recomendaciones para iteraciones futuras

Este estudio sienta las bases para futuras investigaciones en el área de la fertilización asistida utilizando machine learning. A continuación, se presentan algunas recomendaciones que podrían fortalecer y ampliar los enfoques de investigación en este campo.

Un aspecto crucial a mejorar en futuras iteraciones es la inclusión de variables andrológicas, las cuales están relacionadas con el factor masculino. Estas variables son esenciales para una evaluación más completa de la fertilización asistida. Aunque este trabajo no pudo considerarlas debido a la escasa información disponible, su inclusión en estudios futuros podría proporcionar un entendimiento más holístico de los factores que afectan los resultados.

Además, la incorporación de más variables relacionadas con el endometrio podría ofrecer una visión más precisa del tratamiento. Aspectos como el grosor endometrial, la receptividad endometrial y otros factores hormonales influyen en el éxito de la implantación. Esto permitiría a los médicos personalizar aún más los tratamientos y adaptarse a los escenarios reales que enfrentan.

Otro punto importante es la integración de la historia clínica del paciente en los modelos futuros. Incluir esta información podría contribuir a desarrollar modelos más robustos. La evidencia en el estado del arte sugiere que la tasa de éxito del tratamiento está relacionada con la cantidad de ciclos acumulados del paciente. Capturar esta información permitiría identificar patrones que mejorarían la capacidad predictiva de los modelos, facilitando un enfoque más individualizado en el tratamiento.

Por otro lado, aunque las redes neuronales mostraron un gran potencial, es recomendable optimizar estos modelos utilizando técnicas avanzadas de ajuste de hiperparámetros. Esto podría llevar a mejoras significativas en el rendimiento del modelo, reduciendo el riesgo de sobreajuste y aumentando la interpretabilidad de los resultados.

También es fundamental contar con un mayor número de registros de datos clínicos para mejorar la generalización de los modelos. La recolección de más datos permitirá que los modelos sean más robustos y reflejen mejor la diversidad de los casos clínicos, resultando en predicciones más precisas y aplicables a diferentes poblaciones.

Bibliografía

- [1] de Salud. Subsecretaría de Salud Pública. División de Prevención y Control de Enfermedades. Departamento de Ciclo Vital, C. M., “Guía para el estudio y tratamiento de la infertilidad: Programa nacional salud de la mujer - 2015”, 2015.
- [2] Cox, C., Thoma, M., Tchangalova, N., Mburu, G., Bornstein, M., Johnson, C., y Kiarie, J., “Infertility prevalence and the methods of estimation from 1990 to 2021: a systematic review and meta-analysis”, *Human Reproduction Open*, vol. 2022, no. 4, p. hoac051, 2022.
- [3] World Health Organization, “Infertility prevalence estimates, 1990–2021”, 2023, <https://www.who.int/publications/i/item/9789240063721>. Licence: CC BY-NC-SA 3.0 IGO.
- [4] Levine, H., Jørgensen, N., Martino-Andrade, A., Mendiola, J., Weksler-Derri, D., Jolles, M., Pinotti, R., y Swan, S. H., “Temporal trends in sperm count: a systematic review and meta-regression analysis of samples collected globally in the 20th and 21st centuries”, *Human reproduction update*, vol. 29, no. 2, pp. 157–176, 2023.
- [5] de Salud; Subsecretaría de Salud Pública, M., “Modifica resolución exenta n° 277/2011 del ministerio de salud, que aprobó las normas técnico administrativas para la aplicación del arancel del régimen de prestaciones de salud del libro ii dfl n° 1, de 2005, del ministerio de salud, en la modalidad de libre elección”, 2019.
- [6] WIC, “Web intelligence centre”. URL <https://wic.uchile.cl>. 2023.
- [7] IDIMI, “Insituto de investigaciones materno infantil”. URL <https://www.idimi.cl/nuestroinstituto/>. 2023.
- [8] NIH, “Instituto nacional del cáncer”. URL <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/psicosocial>. 2024.
- [9] Ezzell, W., “The impact of infertility on women’s mental health”, *North Carolina medical journal*, vol. 77, no. 6, pp. 427–428, 2016.
- [10] El Kissi, Y., Romdhane, A. B., Hidar, S., Bannour, S., Idrissi, K. A., Khairi, H., y Ali, B. B. H., “General psychopathology, anxiety, depression and self-esteem in couples undergoing infertility treatment: a comparative study between men and women”, *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 167, no. 2, pp. 185–189, 2013.
- [11] Zegers, F., “La infertilidad y el acceso a las técnicas de reproducción médicamente asistida (tra) como un problema de salud pública en Chile”, *Hacia una legislación que regule las técnicas de reproducción médicamente asistida en Chile*, p. 14, 2022.
- [12] Rojo, M. A. E., Fernández-Pascual, E., y Martínez-Salamanca, J. I., “Impacto de la pandemia covid-19 en andrología. recomendaciones en la práctica clínica diaria”, *Revista*

- Internacional De Andrologia, vol. 18, no. 3, p. 124, 2020.
- [13] Connolly, M. P., Hoorens, S., y Chambers, G. M., “The costs and consequences of assisted reproductive technology: an economic perspective”, *Human reproduction update*, vol. 16, no. 6, pp. 603–613, 2010.
- [14] Njagi, P., Groot, W., Arsenijevic, J., Dyer, S., Mburu, G., y Kiarie, J., “Financial costs of assisted reproductive technology for patients in low-and middle-income countries: a systematic review”, *Human reproduction open*, vol. 2023, no. 2, p. hoad007, 2023.
- [15] OMS, “Temas de salud,infertilidad”. URL https://www.who.int/es/health-topics/infertility#tab=tab_1. 2023.
- [16] Van Voorhis, B. J., “In vitro fertilization”, *New England Journal of Medicine*, vol. 356, no. 4, pp. 379–386, 2007.
- [17] Zegers-Hochschild, F., Adamson, G. D., Dyer, S., Racowsky, C., De Mouzon, J., Sokol, R., Rienzi, L., Sunde, A., Schmidt, L., Cooke, I. D., *et al.*, “The international glossary on infertility and fertility care, 2017”, *Human reproduction*, vol. 32, no. 9, pp. 1786–1801, 2017.
- [18] ORG, R. A., “¿cómo es el proceso de la fecundación in vitro paso a paso?”. URL <https://www.reproduccionasistida.org/el-proceso-de-la-fecundacion-in-vitro/>. 2024.
- [19] Fayyad, U., Piatetsky-Shapiro, G., y Smyth, P., “The kdd process for extracting useful knowledge from volumes of data”, *Communications of the ACM*, vol. 39, no. 11, pp. 27–34, 1996.
- [20] Shafique, U. y Qaiser, H., “A comparative study of data mining process models (kdd, crisp-dm and semma)”, *International Journal of Innovation and Scientific Research*, vol. 12, no. 1, pp. 217–222, 2014.
- [21] Santra, A. y Christy, C. J., “Genetic algorithm and confusion matrix for document cluste pages=322,ring”, *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 1, 2012.
- [22] Bini, S. A., “Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care?”, *The Journal of arthroplasty*, vol. 33, no. 8, pp. 2358–2361, 2018.
- [23] Mahesh, B., “Machine learning algorithms-a review”, *International Journal of Science and Research (IJSR)*. [Internet], vol. 9, no. 1, pp. 381–386, 2020.
- [24] Kramer, M. A. y Leonard, J., “Diagnosis using backpropagation neural networks—analysis and criticism”, *Computers & chemical engineering*, vol. 14, no. 12, pp. 1323–1338, 1990.
- [25] Tougui, I., Jilbab, A., y El Mhamdi, J., “Impact of the choice of cross-validation techniques on the results of machine learning-based diagnostic applications”, *Healthcare informatics research*, vol. 27, no. 3, p. 189, 2021.
- [26] Rodríguez-Pérez, R. y Bajorath, J., “Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values”, *Journal of medicinal chemistry*, vol. 63, no. 16, pp. 8761–8777, 2019.
- [27] Deshpande, P. S. y Gupta, A. S., “Causes and prevalence of factors causing infertility in a public health facility”, *Journal of human reproductive sciences*, vol. 12, no. 4,

pp. 287–293, 2019.

- [28] Patel, A., Sharma, P., Narayan, P., Binu, V., Dinesh, N., y Pai, P. J., “Prevalence and predictors of infertility-specific stress in women diagnosed with primary infertility: A clinic-based study”, *Journal of human reproductive sciences*, vol. 9, no. 1, pp. 28–34, 2016.
- [29] Gubert, P. G., Pudwell, J., Van Vugt, D., Reid, R. L., y Velez, M. P., “Number of motile spermatozoa inseminated and pregnancy outcomes in intrauterine insemination”, *Fertility research and practice*, vol. 5, pp. 1–9, 2019.
- [30] Tournaye, H., “Male factor infertility and art”, *Asian Journal of andrology*, vol. 14, no. 1, p. 103, 2012.
- [31] Devroey, P., Fauser, B., Diedrich, K., y 2008, E. A. R. E. W. G., “Approaches to improve the diagnosis and management of infertility”, *Human Reproduction Update*, vol. 15, no. 4, pp. 391–408, 2009.
- [32] Khalife, D., Nassar, A., Khalil, A., Awwad, J., Musa, A. A., Hannoun, A., El Taha, L., Khalifeh, F., Abiad, M., y Ghazeeri, G., “Cumulative live-birth rates by maternal age after one or multiple in vitro fertilization cycles: An institutional experience”, *International Journal of Fertility & Sterility*, vol. 14, no. 1, p. 34, 2020.
- [33] Engmann, L., Maconochie, N., Bekir, J. S., Jacobs, H. S., y Tan, S. L., “Cumulative probability of clinical pregnancy and live birth after a multiple cycle ivf package: a more realistic assessment of overall and age-specific success rates?”, *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 106, no. 2, pp. 165–170, 1999.
- [34] Leijdekkers, J., Eijkemans, M., Van Tilborg, T., Oudshoorn, S., McLernon, D., Bhattacharya, S., Mol, B., Broekmans, F., Torrance, H., y group, O., “Predicting the cumulative chance of live birth over multiple complete cycles of in vitro fertilization: an external validation study”, *Human Reproduction*, vol. 33, no. 9, pp. 1684–1695, 2018.
- [35] Díaz-Narváez, V. P., “Regresión logística y decisiones clínicas”, *Nutrición Hospitalaria*, vol. 34, no. 6, pp. 1505–1505, 2017.
- [36] Sufriyana, H., Husnayain, A., Chen, Y.-L., Kuo, C.-Y., Singh, O., Yeh, T.-Y., Wu, Y.-W., Su, E. C.-Y., *et al.*, “Comparison of multivariable logistic regression and other machine learning algorithms for prognostic prediction studies in pregnancy care: systematic review and meta-analysis”, *JMIR medical informatics*, vol. 8, no. 11, p. e16503, 2020.
- [37] Roberts, S., Hirst, W., Brison, D., Vail, A., y towardSET Collaboration, “Embryo and uterine influences on ivf outcomes: an analysis of a uk multi-centre cohort”, *Human Reproduction*, vol. 25, no. 11, pp. 2792–2802, 2010.
- [38] Milewski, R., Milewska, A. J., Więsak, T., y Morgan, A., “Comparison of artificial neural networks and logistic regression analysis in pregnancy prediction using the in vitro fertilization treatment”, *Studies in Logic, Grammar and Rhetoric*, vol. 35, no. 1, pp. 39–48, 2013.
- [39] CN Barreto, N., Castro, G. Z., Pereira, R. G., Pereira, F. A., Reis, F. M., Junior, W. M., Cavallo, I. K., y Gomes, K. B., “Predicting in vitro fertilization success in the brazilian public health system: a machine learning approach”, *Medical & Biological Engineering & Computing*, vol. 60, no. 7, pp. 1851–1861, 2022.
- [40] Lundberg, S. M. y Lee, S.-I., “A unified approach to interpreting model predictions”,

Advances in neural information processing systems, vol. 30, 2017.

- [41] Hassan, M. R., Al-Insaf, S., Hossain, M. I., y Kamruzzaman, J., “A machine learning approach for prediction of pregnancy outcome following ivf treatment”, *Neural computing and applications*, vol. 32, no. 7, pp. 2283–2297, 2020.
- [42] Forrest, S., “Genetic algorithms”, *ACM computing surveys (CSUR)*, vol. 28, no. 1, pp. 77–80, 1996.
- [43] Guh, R.-S., Wu, T.-C. J., y Weng, S.-P., “Integrating genetic algorithm and decision tree learning for assistance in predicting in vitro fertilization outcomes”, *Expert Systems with Applications*, vol. 38, no. 4, pp. 4437–4449, 2011.
- [44] Louis, C. M., Handayani, N., Aprilliana, T., Polim, A. A., Boediono, A., y Sini, I., “Genetic algorithm–assisted machine learning for clinical pregnancy prediction in in vitro fertilization”, *AJOG Global Reports*, vol. 3, no. 1, p. 100133, 2023.
- [45] Alahmari, F., “A comparison of resampling techniques for medical data using machine learning”, *Journal of Information & Knowledge Management*, vol. 19, no. 01, p. 2040016, 2020.
- [46] Bach, M., Werner, A., Żywiec, J., y Pluskiewicz, W., “The study of under-and over-sampling methods’ utility in analysis of highly imbalanced data on osteoporosis”, *Information Sciences*, vol. 384, pp. 174–190, 2017.
- [47] Khushi, M., Shaukat, K., Alam, T. M., Hameed, I. A., Uddin, S., Luo, S., Yang, X., y Reyes, M. C., “A comparative performance analysis of data resampling methods on imbalance medical data”, *IEEE Access*, vol. 9, pp. 109960–109975, 2021.
- [48] Schaffer, C., “Selecting a classification method by cross-validation”, *Machine learning*, vol. 13, pp. 135–143, 1993.
- [49] Huang, B., Zheng, S., Ma, B., Yang, Y., Zhang, S., y Jin, L., “Using deep learning to predict the outcome of live birth from more than 10,000 embryo data”, *BMC pregnancy and childbirth*, vol. 22, no. 1, p. 36, 2022.
- [50] Ross, S. M., *Introduction to probability models*. Academic press, 2014.
- [51] Edmondson, M. G., Heaney, C. D., Davis, M. F., y Ramachandran, G., “Application of markov models to predict changes in nasal carriage of staphylococcus aureus among industrial hog operations workers”, *Journal of occupational and environmental hygiene*, vol. 19, no. 3, pp. 145–156, 2022.
- [52] Rao, A. S. S. y Diamond, M. P., “Role of markov modeling approaches to understand the impact of infertility treatments”, *Reproductive Sciences*, vol. 24, no. 11, pp. 1538–1543, 2017.
- [53] Ministerio de Salud, “Modificación del convenio de prestación de servicios de fertilización asistida de alta complejidad entre el Fondo Nacional de Salud y la Facultad de Medicina de la Universidad de Chile”, 2019. Resolución Exenta No. 3137.