

**Universidad Andina Simón Bolívar**

**Sede Ecuador**

**Área de Gestión**

Maestría en Administración de Empresas

**Propuesta de diseño de un Modelo de *Machine Learning* basado en predictores mixtos para la estimación del beneficio neto**

**Caso de estudio “Centro Médico MEDISEG”**

Stalin Jamil Segura Sangucho

Tutor: Jaime Oswaldo Pérez Cadena

Quito, 2025

Trabajo almacenado en el Repositorio Institucional UASB-DIGITAL con licencia Creative Commons 4.0 Internacional

	Reconocimiento de créditos de la obra	
	No comercial	
	Sin obras derivadas	

Para usar esta obra, deben respetarse los términos de esta licencia



## Cláusula de cesión de derecho de publicación

Yo, Stalin Jamil Segura Sangucho, autor del trabajo intitulado “Propuesta de Diseño de un modelo de *machine learning* basado en predictores mixtos para la Estimación del Beneficio Neto: Caso de estudio ‘Centro Médico MEDISEG’”, mediante el presente documento dejo constancia de que la obra es de mi exclusiva autoría y producción, que la he elaborado para cumplir con uno de los requisitos previos para la obtención del título de Magíster en Administración de Empresas en la Universidad Andina Simón Bolívar, Sede Ecuador.

1. Cedo a la Universidad Andina Simón Bolívar, Sede Ecuador, los derechos exclusivos de reproducción, comunicación pública, distribución y divulgación, durante 36 meses a partir de mi graduación, pudiendo por lo tanto la Universidad, utilizar y usar esta obra por cualquier medio conocido o por conocer, siempre y cuando no se lo haga para obtener beneficio económico. Esta autorización incluye la reproducción total o parcial en los formatos virtual, electrónico, digital, óptico, como usos en red local y en internet.
2. Declaro que, en caso de presentarse cualquier reclamación de parte de terceros respecto de los derechos de autor/a de la obra antes referida, yo asumiré toda responsabilidad frente a terceros y a la Universidad.
3. En esta fecha entrego a la Secretaría General, el ejemplar respectivo y sus anexos en formato impreso y digital o electrónico.

23 de abril de 2025

Firma: \_\_\_\_\_



## Resumen

Este estudio presenta una propuesta de diseño de un modelo de *machine learning* basada en predictores mixtos para la estimación del beneficio Neto en el Centro Médico MEDISEG. El objetivo es optimizar la estimación del Beneficio Neto en el Centro Médico MEDISEG mediante un modelo de *machine learning* basado en predictores mixtos. La metodología empleada fue de enfoque cuantitativo con técnica de *machine learning* aplicado a datos operativos y financieros del Centro Médico MEDISEG. Se utilizó algoritmos de regresión *Random Forest* y modelos de predicción avanzados. Los resultados demostraron que el costo operativo, el tiempo de servicio, la edad de los pacientes y la calidad percibida de la atención son variables clave de estimación beneficio neto. Se aplicaron criterios de limpieza y validación de los datos obtenidos, a fin de garantizar su confiabilidad del modelo. Posteriormente a la evaluación de distintos algoritmos, *Random Forest* reveló la mayor precisión ( $R^2=0,76$ ), resaltando como el predictor más influyente en la rentabilidad de MEDISEG al costo operativo. Seguidamente, se recomienda implementar monitoreo de las variables en tiempo real, actualizar periódicamente el conjunto de datos, además de explorar más algoritmos como redes neuronales y optimizar los protocolos de la calidad de datos.

Palabras clave: *machine learning*, beneficio neto, predictores mixtos, regresión lineal, centro médico



## Tabla de contenidos

Introducción.....	13
Capítulo primero Enfoque del marco conceptual .....	17
1. Estado del arte .....	17
2. Marco teórico.....	19
2.1. <i>Machine learning</i> .....	19
2.2. Aprendizaje por refuerzo .....	20
2.3. Aprendizaje no supervisado .....	20
2.4. Aprendizaje supervisado .....	20
2.5. Clasificación .....	21
2.6. Regresión .....	21
3. Algoritmos supervisados aplicados a problemas de regresión .....	21
3.1. Redes neuronales .....	21
3.2. Árboles de decisión.....	22
3.3. Métodos de ensamble.....	23
3.4. Algoritmos de <i>bagging</i> .....	23
3.5. Algoritmos de <i>boosting</i> .....	24
3.6. <i>Random Forest</i> .....	25
3.7. Ada Boost.....	26
3.8. Gradient Boosting .....	27
4. Regresión lineal .....	28
4.1. Regresión lineal simple.....	29
4.2. Regresión lineal múltiple .....	30
5. <i>Support Vector Regression (SVR)</i> .....	30
6. Predictores .....	35
Capítulo segundo Acopio y procesamiento de información.....	41
1. Diagnóstico.....	41
1.1. Diagnóstico externo .....	41
1.1.1. Análisis PESTEL .....	41
1.1.2. Análisis 5 fuerzas de Porter .....	42
1.2. Diagnóstico interno .....	44
1.2.1. Antecedentes de la empresa .....	44
1.2.2. Misión .....	44
1.2.3. Visión.....	44

1.2.4. Valores institucionales .....	45
1.2.5. Estructura organizacional.....	45
1.2.6. Diagrama de distribución .....	45
1.2.7. Servicios.....	46
1.2.8. Procesos .....	47
1.2.9. Marketing.....	50
1.2.10. Análisis financiero .....	50
1.2.11. Modelo de negocio.....	51
1.2.12. Matriz FODA .....	52
Capítulo tercero Presentación y análisis de resultados .....	55
1. Propuesta de Diseño de un modelo de <i>machine learning</i> basado en predictores mixtos para la estimación del beneficio neto en MEDISEG .....	55
2. Identificación de las Variables Críticas que Influyen en la Estimación del Beneficio Neto en el Centro Médico MEDISEG.....	56
2.1. Características del conjunto de datos .....	56
2.2. Proceso de limpieza de datos .....	60
2.3. Criterios de análisis y evaluación de la calidad de los datos .....	61
2.4. Identificación y visualización de la variable más crítica .....	61
2.5. Relación entre variable edad e ingreso .....	62
2.6. Definición e integración de la variable categórica de satisfacción del cliente.....	63
3. Evaluación de la precisión e impacto de modelos de <i>machine learning</i> .....	66
Capítulo cuarto .....	71
1. Validación del modelo.....	71
1.1. Validación del modelo predictivo con datos externos .....	71
1.2. Ejecución del modelo al caso PHYSIS .....	71
1.2.1 Preparación de los datos.....	71
1.2.2 Prueba del modelo <i>random forest</i> .....	72
1.2.3 Análisis del beneficio neto .....	74
1.2.4 Análisis de la depreciación y su rol en la rentabilidad.....	75
1.2.5 Consideraciones de planificación fiscal.....	75
1.2.6 Diagnóstico estratégico a partir del modelo.....	75
Conclusiones.....	77
Recomendaciones .....	78
Lista de referencias .....	79
Anexos.....	83

Anexo 1: Estado financieros de la empresa PHYSIS Rehabilitación Física y Especialidades Médicas S.A.S.....	83
Anexo 2: Acta de la Asamblea General Extraordinaria de Accionistas 2024 .....	86



## Figuras y tablas

Figura 1. Tipos de <i>machine learning</i> .....	19
Figura 2. Capas de las redes neuronales .....	22
Figura 3. Método paralelo de <i>bagging</i> y <i>boosting</i> .....	24
Figura 4. Funcionamiento del <i>Random Forest</i> .....	26
Figura 5. Funcionamiento del <i>ada boost</i> .....	27
Figura 6. Regresión lineal.....	29
Figura 7. Representación gráfica de modelos lineales $\epsilon$ -SV .....	31
Figura 8. Ilustración de modelos no lineales $\epsilon$ -SVR .....	32
Figura 9. Organigrama.....	45
Figura 10. Primera Planta .....	45
Figura 11. Planta Alta.....	46
Figura 12. Diagrama de procesos de gestión administrativa .....	48
Figura 13. Diagrama de procesos de gestión de medicina general.....	488
Figura 14. Diagrama de procesos de gestión de medicina general.....	49
Figura 15. Diagrama de procesos de gestión de Laboratorio .....	49
Figura 16. Diagrama de procesos de Farmacia.....	50
Figura 17. Modelo de Negocio Canvas Mediseg .....	52
Figura 18. Radial de Satisfacción Clientes .....	54
Figura 19. Estructura de las etapas del modelo .....	555
Figura 20. Esquema del desarrollo del modelo .....	566
Figura 21. Estructura de conjunto de datos .....	58
Figura 22. Procesamiento o limpieza de datos .....	59
Figura 23. Relación entre Costo e Ingreso en el Centro Médico MEDISEG .....	62
Figura 24. Relación entre Edad e Ingresos en el Centro Médico MEDISEG .....	633
Figura 25. Entrenamiento y evaluación .....	655
Figura 26. Entrenamiento y evaluación .....	67
Figura 27. Importancia de las variables en la predicción del beneficio neto.....	700
Figura 28. Importancia de las variables en la predicción del beneficio neto Caso Physys.73	

Tabla 1 Análisis PESTEL.....	41
Tabla 2 5 fuerzas de PORTER de Mediseg.....	42
Tabla 3 Servicios de Mediseg.....	46
Tabla 4 Especialidades médicas de Mediseg.....	47
Tabla 5 Flujo de efectivo de Mediseg .....	51
Tabla 6 FODA .....	52
Tabla 7 Detalle de variables .....	57
Tabla 8 Etapas de selección de variables.....	58
Tabla 9 Etapas de procesamiento de datos .....	60
Tabla 10 Etapas y codificación de testing y training.....	66
Tabla 11 Etapas de configuración, evaluación y selección del modelo .....	67
Tabla 12 Resultados .....	69
Tabla 13 Detalle de variables Physys.....	72

## Introducción

La estimación de indicadores financieros como el beneficio neto cumplen un rol fundamental en la toma de decisiones empresariales (Zaldívar et al. 2011), particularmente en aquellas que desempeñan sus funciones en sectores de altos niveles de variabilidad y complejidad como lo es el sector de la salud. Con frecuencia, las empresas pertenecientes al sector de la salud afrontan grandes volúmenes de datos donde se relacionan variables tanto cuantitativas como cualitativas, situación que genera dificultad e imprecisión de los métodos tradicionales de estimación financiera al abordar las dinámicas subyacentes (Hastie 2009, 21).

Luyo y Naveda (2022, 5) afirma que en los centros de salud, especialmente en laboratorios y clínicas privadas, existen diversos problemas relacionados con la gestión económica y la optimización de recursos. Uno de los mayores desafíos es mejorar sus resultados financieros debido a la falta de herramientas efectivas para calcular ingresos y gastos. Al carecer de modelos claros para realizar estimaciones precisas, deben tomar decisiones con incertidumbre, lo que dificulta su rentabilidad y sostenibilidad.

Además, la pandemia del covid-19 entre 2019 y 2020 impactó severamente a muchos de estos centros de salud. Cambios en la demanda de servicios y la disponibilidad de pruebas diagnósticas más asequibles en farmacias resultaron en una significativa disminución de ingresos para muchas clínicas y laboratorios. Esta situación evidenció la necesidad de contar con herramientas que permitan una mejor proyección financiera y adaptación rápida a los cambios del mercado (Ramírez, Montenegro, y Estupiñán 2021, 45).

Otro aspecto que destacar es la competencia intensa en este sector, donde empresas grandes y pequeñas compiten por los mismos pacientes. Esto subraya la importancia de disponer de métodos avanzados para prever ingresos, reducir costos y mejorar la eficiencia operativa. Sin embargo, la falta de métodos confiables para realizar estas estimaciones genera una gran incertidumbre y dificulta la estabilidad financiera que estas organizaciones buscan alcanzar.

La estimación de indicadores financieros como el beneficio neto, cumplen un rol fundamental en la toma de decisiones empresariales, particularmente en aquellas que

desempeñan sus funciones en sectores de altos niveles de variabilidad y complejidad como lo es el sector de la salud (Breiman 2009).

Con frecuencia, las empresas pertenecientes al sector de la salud afrontan grandes volúmenes de datos donde se relacionan variables tanto cuantitativas como cualitativas, situación que genera dificultad e imprecisión de los métodos tradicionales de estimación financiera al abordar las dinámicas subyacentes (Hastie 2009).

La aplicación de técnicas de aprendizaje automático (*machine learning*), han logrado constituirse como una opción prometedora para tratar la complejidad de estimaciones de carácter financiero, gracias a su capacidad de procesamiento de grandes cantidades de datos con un nivel de precisión superior a los métodos tradicionales de la regresión lineal o múltiple (Franco 2023, 280).

En gran parte de los modelos utilizados en el sector financiero se enfatiza en variables de tipo cuantitativo, lo cual podría llevar a una subestimación o a una sobreestimación del beneficio neto debido a los efectos de omisión de variables cualitativas que también pueden influir en el resultado de las empresas. Por tanto, una de las principales problemáticas que enfrentan estas técnicas, radica en la incorporación efectiva de predictores mixtos que logre combinar variables tanto cuantitativas como cualitativas en un único modelo predictivo.

La pregunta de investigación propuesta sostiene: ¿Por qué es necesario desarrollar un modelo de *machine learning* basado en predictores mixtos integrados por las variables críticas para la estimación del beneficio neto en el Centro Médico MEDISEG?

En tanto que las preguntas específicas son:

- a) ¿Qué variables críticas influyen en la estimación del beneficio neto en el Centro Médico MEDISEG?
- b) ¿Qué criterios de análisis y evaluación de la calidad de los datos se utilizan para el proceso de modelado?
- c) ¿Qué datos conforman el conjunto final que se utilizará en la etapa de modelamiento?
- d) ¿Cuáles son los tipos de algoritmo de *machine learning* que integran apropiadamente predictores mixtos (cuantitativos y cualitativos)?
- e) ¿Cuál es la precisión e impacto de modelos de *machine learning* desarrollados para la estimación del beneficio neto?

El objetivo general de la investigación es optimizar la estimación del beneficio neto del Centro Médico MEDISEG mediante el diseño de un modelo de *machine learning* basado en predictores mixtos integrados por variables críticas.

En cuanto a los objetivos específicos, se plantean los siguientes:

1. Identificar las variables críticas que influyen en la estimación del beneficio neto en el Centro Médico MEDISEG.
2. Identificar los criterios de análisis y evaluación de la calidad de los datos que se utilizan en el proceso de modelado.
3. Definir el conjunto final de datos que se utilizará en la etapa de modelamiento.
4. Identificar y seleccionar los tipos de algoritmo de *machine learning* que integran apropiadamente predictores mixtos (cuantitativos y cualitativos).
5. Evaluar la precisión e impacto de modelos de *machine learning* desarrollados para la estimación del beneficio neto.

Esta investigación se justifica dado que ante un entorno cada vez más competitivo en el sector de la salud, dentro del cual han aparecido nuevos integrantes de mercado dentro del perímetro donde se ubica el Centro Médico MEDISEG (Conocoto), constituyendo así un componente de cambio en la variación de los múltiples factores que influyen en la demanda de servicios médicos. Por tanto, se torna importante la identificación de patrones y relaciones entre las variables que afectan el beneficio neto, proporcionando de esta forma las herramientas precisas que MEDISEG requiere para afrontar la competitividad del mercado mediante adecuada planificación financiera y toma de decisiones estratégica.

En la actualidad, esta casa de salud utiliza técnicas convencionales de estimación financiera, que se apoyan principalmente en modelos estadísticos tradicionales como el método de mínimos cuadrados y el análisis retrospectivo de la información. No obstante, estos procedimientos que han demostrado su eficacia en el pasado, tienen restricciones importantes respecto a su habilidad para identificar relaciones no lineales o patrones complejos en los datos, particularmente cuando se toman en cuenta varios factores cualitativos y cuantitativos que pueden afectar el rendimiento financiero. Por ende, se presenta la necesidad de adoptar técnicas más robustas que permitan la integración de múltiples factores de predicción, a fin de alcanzar mayor precisión en las estimaciones deseadas.

De acuerdo con López, Magdaleno, y Sánchez (2023, 24), dentro del ámbito financiero “la creciente presencia y uso de los modelos de *machine learning* predomina

por sobre todo en áreas vinculadas a la predicción de variables financieras”, por consiguiente, es factible su aplicación en el desarrollo de un modelo que permita la optimización del beneficio neto.

El caso de estudio “MEDISEG” brindará un contexto práctico que posteriormente permitirá evaluar la efectividad del modelo propuesto y su desempeño en un ámbito real, contribuyendo de esta forma a la literatura existente sobre la aplicación de *machine learning* en la optimización de recursos financieros.

## Capítulo primero

### Enfoque del marco conceptual

#### 1. Estado del arte

Pineda (2022) realizó un artículo con el tema de “Modelos predictivos en salud basados en aprendizaje de maquina (*machine learning*)” cuyo objetivo fue sustentar las bases teóricas necesarias y evidenciar cómo esta herramienta *machine learning* han permitido optimizar la precisión y rapidez en la toma de decisiones en los centros de salud o clínicas. Los resultados que se obtuvieron han demostrado que el uso de los registros clínicos y el uso del modelo del *machine learning* han facilitado la identificación de patrones desconocidos, lo que ha traído la integración sistemática de este modelo a la práctica médica. Las conclusiones que llego fue que el modelo del *machine learning* no solo automatiza la predicción de diagnósticos, sino que también ha logrado una utilidad en la gestión de riesgos y en la implementación de alertas tempranas en diversas áreas de la medicina.

López y Villanueva (2022) desarrollaron el tema “Sistema para la automatización de procesos hospitalarios de control para pacientes para covid-19 usando *machine learning* para un Centro de Salud”, el objetivo que plantearon fue optimizar la calidad de atención reduciendo los tiempos de espera de los pacientes con covid-19. La metodología empleada fue de enfoque cuantitativo, enfocado en dos variables, como variable independiente: *machine learning*, como variable dependiente: automatización de procesos hospitalarios, el instrumento empleado fue la encuesta dirigida a 50 trabajadores de la salud. Los resultados mostraron que el tiempo promedio de admisión disminuyó de 20,74 a 13,56 minutos y que el porcentaje de infecciones intrahospitalarias se redujo del 330,38 % al 250,28 %. A través de pruebas estadísticas con IBM SPSS y la prueba de Wilcoxon, se confirmó una mejora significativa en la gestión hospitalaria. Los autores concluyeron que la aplicación de *machine learning* facilita la optimización de los procesos clínicos, mejorando la toma de decisiones y la eficiencia del servicio hospitalario.

Quiroz et al. (2022) realizaron un estudio sobre “Aprendizaje automático aplicado en área de la salud”, enfocado en la evaluación del rendimiento, selección del modelo más

adecuado y validación externa de los algoritmos empleados. La investigación destacó la importancia de definir correctamente las variables utilizadas en los modelos, ya que una mala selección puede introducir sesgos y reducir la fiabilidad de las predicciones. Para la recolección de datos, se emplearon registros de historias clínicas electrónicas, lo que permitió trabajar con un amplio conjunto de variables clínicas. Se evaluaron distintos algoritmos de clasificación mediante métricas como la matriz de confusión, precisión, exactitud y área bajo la curva ROC. Los resultados mostraron que los modelos de *machine learning* pueden superar las técnicas estadísticas tradicionales en la predicción de enfermedades, aunque presentan desafíos en su implementación clínica. Los autores concluyeron que la adopción de estas tecnologías requiere validaciones rigurosas, colaboración interdisciplinaria y estrategias para minimizar sesgos en los datos.

Collazos-Castillo (2023) realizaron un estudio sobre aplicaciones de *machine learning* para la mejora de la atención médica en Colombia, con el objetivo de analizar su impacto en la precisión diagnóstica y la optimización de los procesos clínicos. La investigación consideró variables como la eficiencia en la toma de decisiones médicas y la reducción del tiempo de diagnóstico. Como instrumento de recolección de datos, se utilizaron encuestas dirigidas a profesionales de la salud y análisis de registros electrónicos hospitalarios. Los resultados mostraron que la implementación de *machine learning* permitió reducir los tiempos de espera en los diagnósticos y mejorar la identificación de patrones clínicos, lo que facilitó la personalización de tratamientos. Se concluyó que la integración de estas tecnologías en el sector salud optimiza la atención a los pacientes, aunque se requieren estrategias de capacitación y adaptación para su implementación efectiva.

Marshall (2022) desarrolló un estudio sobre “Diseño de un modelo de generación de datos sintéticos para la aplicación de modelos de *machine learning* en proyectos interdisciplinarios asociados a salud”, con el objetivo de abordar la escasez de datos en estudios clínicos. La investigación utilizó registros del proyecto Alzheimer *Depression Diagnostic with Artificial Intelligence* y aplicó tres algoritmos generativos: Generative Adversarial Networks (GAN), Variational Autoencoders (VAE) y Gaussian Copula. Se analizaron variables provenientes de exámenes como electroencefalogramas, seguimiento ocular y pruebas de navegación, utilizando métricas como *Recall* y ROC AUC para evaluar la calidad de los datos generados. Los resultados indicaron que Gaussian Copula presentó la menor pérdida de precisión, con diferencias de -9 % y -5 % en *Recall* y ROC AUC, respectivamente, cuando solo se usaron datos sintéticos, y de -18,5 % y -13,5 % al

combinar datos reales y sintéticos. Se concluyó que la generación de datos sintéticos ayuda a mitigar la falta de información en estudios médicos, aunque requiere mejoras para garantizar su representatividad y aplicabilidad en modelos de predicción.

## 2. Marco teórico

### 2.1. *Machine learning*

El *machine learning* es una rama esencial de la inteligencia artificial. En esencia, permite a las computadoras y máquinas aprender de sus experiencias sin necesidad de ser programadas explícitamente para cada tarea. Un ejemplo típico son las predicciones o sugerencias en situaciones específicas (Rouhiainen 2020).

Históricamente, las primeras computadoras personales que surgieron en los años 80 requerían programación explícita para cada acción. En contraste, las máquinas actuales utilizan el aprendizaje automático para personalizar la experiencia del usuario en función de su uso. Hoy en día, ejemplos de este enfoque incluyen la personalización en redes sociales como Facebook y los resultados de búsqueda en Google (19).

El aprendizaje automático se basa en algoritmos que identifican patrones en los datos. Por ejemplo, los filtros de spam en los correos electrónicos emplean esta tecnología para distinguir entre mensajes no deseados y legítimos, aprendiendo continuamente de cada interacción para mejorar su precisión. Este es un ejemplo claro de cómo los algoritmos pueden aprender de los patrones y aplicar ese conocimiento para tomar decisiones efectivas (20).

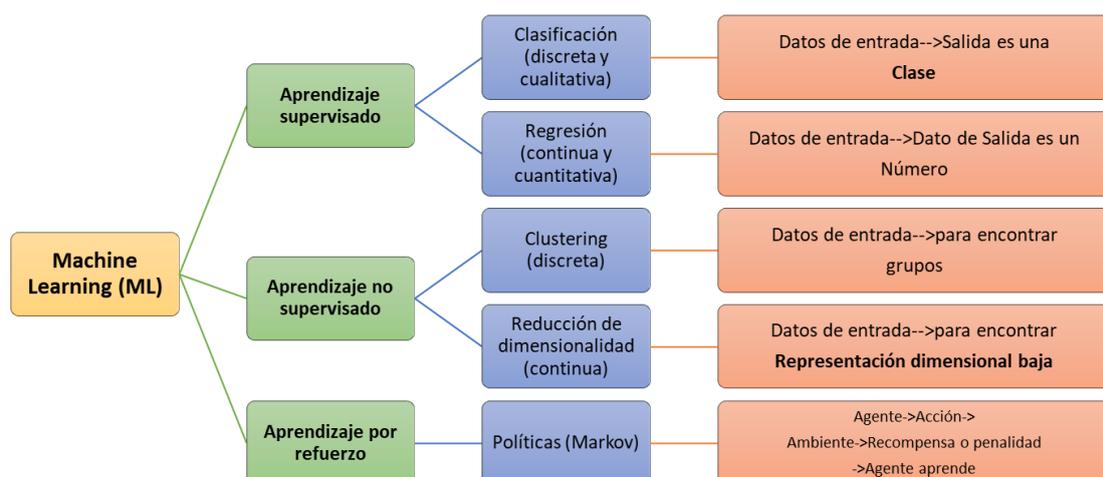


Figura 1. Tipos de *machine learning*  
Fuente: adaptado de Rouhiainen (2020).

Por tanto, el *machine learning* es un área de la informática en la que las computadoras o máquinas tienen la capacidad de aprender sin ser explícitamente programadas para ello. Normalmente, un resultado del aprendizaje automático son predicciones o sugerencias en situaciones específicas. Este aprendizaje utiliza diversos algoritmos para identificar patrones en los datos y aplicar el conocimiento adquirido para tomar decisiones.

El *machine learning* se divide en tres categorías principales: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo (Rouhiainen 2020). A continuación, se definirán brevemente cada uno de estos tipos, con un énfasis especial en el aprendizaje supervisado, que será el enfoque utilizado en este estudio.

## **2.2. Aprendizaje por refuerzo**

El aprendizaje por refuerzo es un enfoque donde los algoritmos aprenden a partir de la experiencia, y se les motiva mediante refuerzos positivos cada vez que aciertan (Rouhiainen 2020). Este tipo de aprendizaje se basa en la psicología conductual, que establece que el aprendizaje ocurre cuando una acción específica es seguida de una recompensa, lo que hace más probable que se repita dicha acción (Ossa y Jaramillo 2021, 14).

## **2.3. Aprendizaje no supervisado**

El aprendizaje no supervisado se caracteriza por no utilizar datos etiquetados o preorganizados que indiquen cómo debe categorizarse la nueva información. En su lugar, los algoritmos deben descubrir la manera de clasificar los datos por sí mismos, generalmente agrupándolos según similitudes o patrones que identifican (Rouhiainen 2020, 21). En problemas de esta naturaleza, solo se puede describir la estructura de los datos para intentar encontrar algún tipo de organización que facilite el análisis. Estos algoritmos funcionan sin necesidad de entrenamiento previo, es decir, comienzan a operar tan pronto como reciben los datos.

## **2.4. Aprendizaje supervisado**

El aprendizaje supervisado utiliza un conjunto de datos que incluye tanto la información de entrada como la de salida. A partir de estos datos, el modelo aprende a producir los resultados deseados para la variable de salida. El algoritmo se ajusta continuamente para minimizar el error hasta alcanzar una precisión aceptable. Es una

buena práctica, antes de usar modelos de pronóstico supervisados, dividir el conjunto de datos en dos subconjuntos: entrenamiento y prueba. Generalmente, se utiliza el 70 % de los datos como conjunto de entrenamiento y el 30 % restante como conjunto de prueba. Este método ayuda a evitar el sobreajuste, un problema común en los modelos de predicción (Alvarez 2023, 26).

El aprendizaje supervisado puede clasificarse en dos tipos de problemas: clasificación y regresión (IBM 2022, párr. 1).

## **2.5. Clasificación**

Los problemas de clasificación buscan entrenar un algoritmo para categorizar datos en una variable discreta a partir de los patrones identificados en los datos. Este proceso implica reconocer entidades específicas dentro de un conjunto de datos e intentar determinar cómo deben ser etiquetadas o clasificadas. Entre los algoritmos que se utilizan para resolver problemas de clasificación se encuentran la regresión logística, k-vecinos más cercanos, máquinas de soporte vectorial, árboles de decisión, y bosques aleatorios, entre otros (IBM 2022, párr. 2).

## **2.6. Regresión**

Por otro lado, la regresión es una técnica que entrena un algoritmo para predecir una salida numérica continua. Este método se utiliza comúnmente en problemas de minería de datos con el objetivo de pronosticar una variable numérica continua. Ejemplos de métodos de regresión incluyen la regresión lineal simple y multivariable, k vecinos más cercanos, máquinas de soporte vectorial, árboles de decisión y redes neuronales, entre otros (párr. 2).

# **3. Algoritmos supervisados aplicados a problemas de regresión**

## **3.1. Redes neuronales**

Las redes neuronales son un tipo de modelo de *machine learning* que busca imitar la estructura y el funcionamiento del cerebro humano. Están conformadas por múltiples nodos interconectados llamados *neuronas*, cada una llevando a cabo operaciones matemáticas en los datos de entrada y transmitiéndolos a la siguiente capa. A medida que los datos atraviesan la red, los pesos de las conexiones entre las neuronas se ajustan basándose en el error de predicción. Este proceso entrena a las redes neuronales para

aprender y representar relaciones no lineales en los datos, haciéndolas herramientas altamente efectivas para abordar problemas complejos. Además, su capacidad para aprender automáticamente características relevantes de los datos las hace muy versátiles para una amplia variedad de aplicaciones, como el reconocimiento de imágenes, el procesamiento de lenguaje natural y las predicciones en series temporales. Sin embargo, es importante considerar que el rendimiento de una red neuronal puede depender de hiperparámetros específicos, como el número de capas, las neuronas por capa y la tasa de aprendizaje. La selección adecuada de estos hiperparámetros es fundamental para lograr un rendimiento óptimo del modelo (Amat 2021).

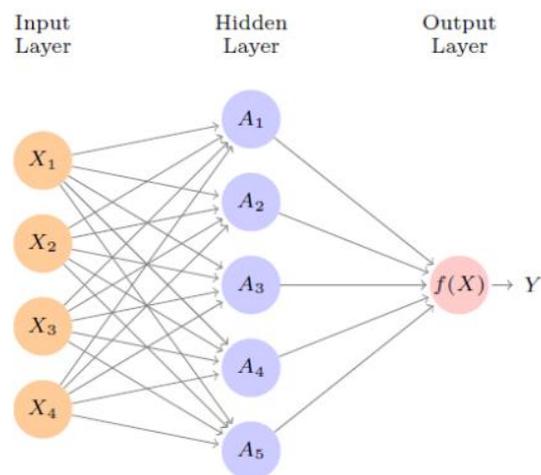


Figura 2. Capas de las redes neuronales  
Fuente: (Blanco y Rubriche 2024).

### 3.2. Árboles de decisión

Los árboles de decisión son algoritmos de *machine learning* supervisado muy potentes y versátiles, utilizados tanto para tareas de regresión como de clasificación con conjuntos de datos complejos. Esta técnica aprende reglas de decisión basadas en características para predecir valores de respuesta. A diferencia de los modelos lineales, los árboles de decisión pueden capturar relaciones no lineales entre las características (variables  $x$ ) y la variable objetivo ( $Y$ ) mediante reglas de decisión basadas en las variables  $x$ . Si se tiene  $Y$  como variable respuesta y las variables predictoras  $x_1, x_2, \dots, x_p$ , el objetivo es determinar una relación entre  $x$  y  $Y$  para poder estimar  $Y$  en función de  $x$ .

Un árbol de regresión tiene una estructura jerárquica básica que incluye un nodo raíz, ramas, nodos internos y nodos hoja. El objetivo es crear un sistema de múltiples

opciones o decisiones bifurcadas en forma de árbol, de modo que al seguir cada rama se obtenga una predicción final para los individuos que cumplen con las características especificadas en cada bifurcación (Gil-Martínez 2018, 23).

Los árboles de decisión tienden a ajustarse demasiado a los datos de entrenamiento, lo cual significa que pueden ser muy precisos al predecir resultados en estos datos, pero esta precisión puede no aplicarse a nuevos datos. Por ello, existen algoritmos de *machine learning* que combinan respuestas generadas por múltiples árboles y, mediante métodos de ensamblaje, producen una única predicción que es mucho más precisa y resuelve problemas de sobreajuste mencionados anteriormente (24).

### 3.3. Métodos de ensamble

Cuando se enfrenta un problema de aprendizaje supervisado, ya sea para clasificación o regresión, lo habitual es entrenar un algoritmo y tratar de optimizar su desempeño en esa tarea. Sin embargo, en la práctica, a veces los algoritmos arrojan resultados insatisfactorios; estos algoritmos son comúnmente denominados como algoritmos débiles. La idea detrás de los métodos de ensamble es seleccionar varios modelos débiles y combinar sus salidas para generar un modelo mucho más sólido. Imagínese hacer una pregunta compleja a miles de personas al azar y luego combinar sus respuestas. A menudo, la respuesta combinada resulta ser mejor que la de un experto único. Esto se conoce como *sabiduría de las masas*. De manera similar, si se combinan las predicciones de diferentes modelos, a menudo se obtendrán mejores resultados que usando solo el mejor predictor individual. Este es el principio fundamental detrás de todos los métodos de ensamble (Géron 2022). A continuación, se detallarán los métodos de bagging y boosting, los cuales se desarrollarán en este proyecto con un enfoque en algoritmos basados principalmente en árboles de decisión.

### 3.4. Algoritmos de *bagging*

El *bagging* es una técnica de ensamblaje usada para reducir la variabilidad en datos ruidosos. Consiste en entrenar varios modelos de manera independiente y luego combinar sus resultados. En este método, se selecciona aleatoriamente una muestra de datos del conjunto de entrenamiento con reemplazo, lo que significa que algunos datos pueden ser seleccionados varias veces. Luego, múltiples modelos débiles se entrenan independientemente utilizando estas muestras de datos y, dependiendo de la tarea

(clasificación o regresión), se combinan las predicciones mediante un promedio o una mayoría para obtener una estimación más precisa (Alvarez 2023, 28).

Los métodos de *bagging* son técnicas donde diversos algoritmos simples se usan en paralelo y luego sus resultados se combinan en una única respuesta. El objetivo principal de los métodos en paralelo es aprovechar la independencia entre los algoritmos simples, ya que el error puede reducirse significativamente al promediar las salidas de dichos modelos simples. Es similar a resolver un problema con varias personas independientes entre sí, y aceptar la solución seleccionada por la mayoría de ellas (De Alba 2020, párr. 4).

### 3.5. Algoritmos de *boosting*

El *boosting* es una técnica de aprendizaje secuencial donde el algoritmo comienza entrenando un modelo con el conjunto completo de datos de entrenamiento. Posteriormente, se crean modelos adicionales ajustando los errores residuales del modelo inicial. Esto permite enfocarse en las observaciones que fueron mal clasificadas por el modelo previo. Una vez construida la secuencia de modelos, las predicciones se ponderan según la precisión de cada modelo y se combinan para obtener una estimación final más precisa.

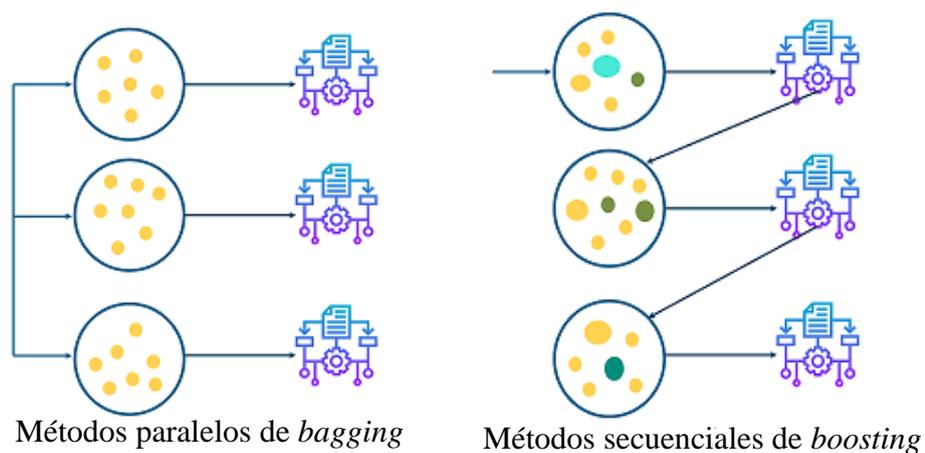


Figura 3. Método paralelo de *bagging* y *boosting*  
Fuente: (De Alba 2020).

De esta manera, el *boosting* mejora el rendimiento del modelo al centrarse en los errores y en las áreas que necesitan mejora. El objetivo principal de los métodos secuenciales es beneficiarse de la relación entre los modelos simples, también conocidos como “modelos débiles”. Este enfoque permite que el rendimiento general se incremente, ya que los modelos posteriores otorgan mayor importancia a los errores cometidos por

los modelos previos, creando un modelo final más robusto y preciso. Es similar a resolver un problema aprovechando el conocimiento de errores anteriores para evitar cometerlos nuevamente (De Alba 2020, párr. 7).

### 3.6. *Random Forest*

*Random Forest* es un método de ensamblaje basado en el algoritmo *bagging* que opera generando múltiples árboles de decisión durante la fase de entrenamiento para luego utilizarlos en la predicción de la variable de salida ( $Y$ ). Es uno de los algoritmos más utilizados debido a su precisión, simplicidad y flexibilidad. El término *bosque* se utiliza porque el algoritmo genera un conjunto de árboles de decisión, cuyos datos se fusionan para generar mejores predicciones. Mientras que un árbol de decisión por sí solo tiene un resultado limitado, el bosque asegura un resultado más preciso al agregar aleatoriedad al modelo, encontrando la mejor característica entre un subconjunto aleatorio de características.

El procedimiento para crear un *Random Forest* es el siguiente (Hernández 2021):

- Crear un nuevo conjunto de entrenamiento mediante la técnica bootstrap, que al igual que el conjunto de entrenamiento original, tendrá la misma cantidad de observaciones  $n$ . Con esta técnica, se toman muestras con reemplazo, permitiendo que algunas observaciones aparezcan múltiples veces y otras no aparezcan en absoluto en el nuevo conjunto.
- Construir un árbol de regresión utilizando un subconjunto de variables independientes predictoras en cada partición del nuevo conjunto de entrenamiento.
- Repetir los pasos 1 y 2  $B$  veces (usualmente  $B = 500$  o  $B = 1000$ ). Por esta forma, se tendrán  $B$  árboles de decisión que se utilizarán para crear predicciones de  $Y$ .

Para predecir la variable  $Y$ , se utilizan los  $B$  árboles creados y se obtienen las predicciones  $\hat{y}^1, \hat{y}^2, \dots, \hat{y}^B$  generadas por dichos árboles. Finalmente, se unifican las  $B$  predicciones mediante un promedio ponderado para obtener una predicción final.

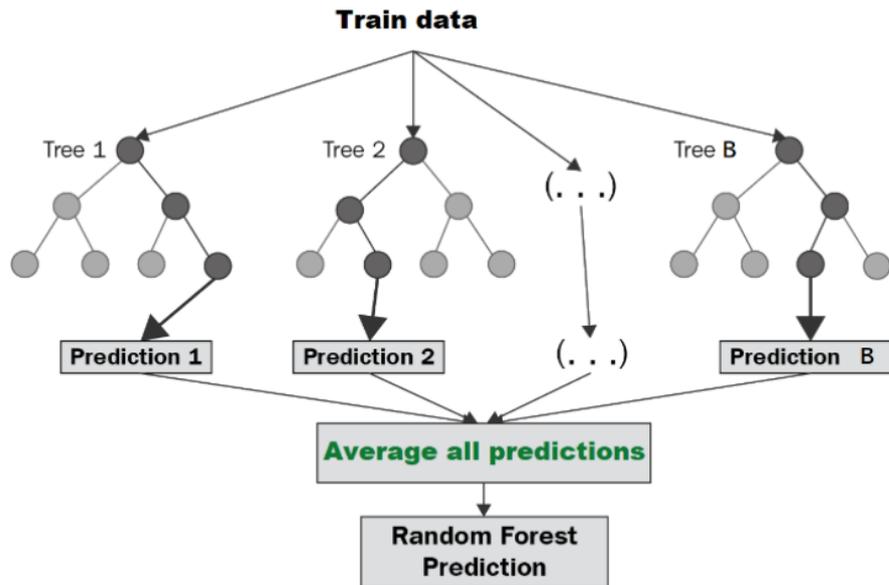


Figura 4. Funcionamiento del *Random Forest*  
Fuente: (Hernández 2021).

### 3.7. Ada Boost

Este método de ensamble utiliza la técnica de *boosting* y consiste en crear varios predictores sencillos en secuencia. El segundo predictor corrige lo que el primero no ajustó correctamente, la tercera mejora lo que el segundo no pudo, y así sucesivamente. Su objetivo principal es combinar varios predictores débiles mediante una combinación lineal ponderada para formar un único predictor fuerte.

En cada iteración del proceso, es decir, cada vez que se genera un nuevo predictor débil, se le asigna un peso correspondiente. Esta asignación se realiza de manera que los predictores con menor error de predicción reciben un mayor peso.

A continuación, se detalla el funcionamiento interno de un modelo Ada Boost (Gil-Martínez 2018, 27):

Supongamos que tenemos un conjunto de entrenamiento con  $n$  observaciones y  $M$  modelos débiles. Inicialmente, el peso asignado a cada observación  $w_i$  es:

$$w_i = 1/n, i = 1, 2, \dots, n.$$

El procedimiento es el siguiente:

1. Se entrena el primer modelo débil con las  $n$  observaciones y se calcula el error de predicción  $err_j$ ,  $j = 1, 2, \dots, M$  cometido por el modelo.

2. El peso  $\alpha_k$ ,  $k = 1, 2, \dots, M$  se asigna al modelo débil según sus aciertos: a mayor aciertos, mayor peso.

3. Se actualizan los pesos  $w_i$  de las observaciones; se disminuye el peso de las bien clasificadas y se aumenta el de las mal clasificadas.

4. Se repiten los pasos 1, 2 y 3 hasta obtener el ensamble final.

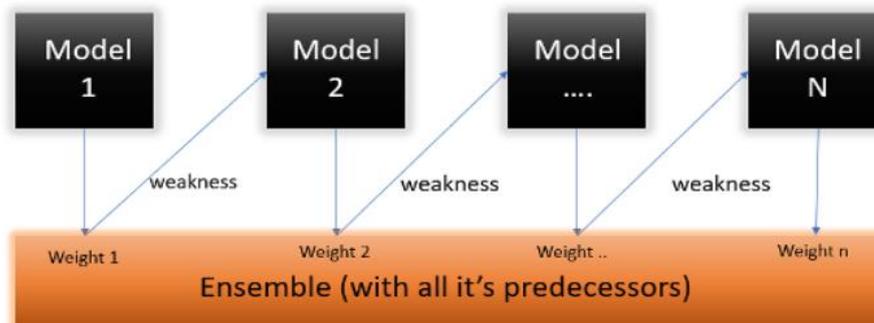


Figura 5. Funcionamiento del ada boost  
Fuente: (Saini 2021).

### 3.8. Gradient Boosting

El *gradient boosting* es un algoritmo de aprendizaje supervisado que emplea la técnica de *boosting* para crear una secuencia de varios predictores, donde cada nuevo predictor se enfoca en corregir los errores del predictor anterior, de manera similar al *ada boost*. No obstante, la diferencia principal radica en que el *gradient boosting* usa como primer modelo predictor la media de la variable ( $Y$ ) para hacer sus predicciones. Desde el segundo predictor en adelante, ajusta los errores del modelo anterior, de tal forma que cada modelo subsiguiente se enfoca en corregir los errores no abordados previamente.

A diferencia del *ada boost*, el *gradient boosting* no asigna un peso independiente  $w_i$  a cada observación del conjunto de entrenamiento. En su lugar, utiliza una función de coste diferenciable  $L(y_i, f(x))$ , cuyo objetivo es minimizar su gradiente o derivada parcial, obtenida a partir de los residuos ( $r_iM$ ) en un proceso iterativo. El gradiente determina la dirección en la que se deben ajustar los parámetros de los predictores débiles para reducir el error de predicción en las siguientes iteraciones. En este método, las predicciones generadas por el predictor débil  $M$  intentan acercarse al gradiente negativo de la función de coste. Cabe señalar que se puede utilizar cualquier función de coste diferenciable que permita calcular derivadas en cualquier dirección (Gil-Martínez 2018,23).

#### 4. Regresión lineal

Se utiliza para predecir el valor de una variable en función del valor de otra. La variable que queremos predecir se llama variable dependiente, y la que usamos para predecir se llama variable independiente. La regresión estima los coeficientes en una ecuación lineal, combinando varias variables independientes para predecir de manera más precisa la variable dependiente. Se ajusta a una línea o superficie que minimiza las discrepancias entre los valores obtenidos. Además, su realización es sencilla y proporciona una fórmula matemática fácil de interpretar.

Según IBM (2018, 34), para realizar un análisis de regresión correcto se deben considerar las siguientes hipótesis:

- Para la variable: Número total de casos válidos, desviación estándar y media.
- Para el modelo: Coeficientes de regresión, matriz, correlaciones, múltiple R,  $R^2$ ,  $R^2$  ajustado, error, varianza, valores previstos y residuales. También se deben considerar los intervalos de confianza al 95 % para cada coeficiente de regresión, factor de inflación, prueba de Durbin-Watson, medidas de distancia y los intervalos de predicción.
- Gráficos: Diagramas de dispersión, gráficos parciales y de probabilidad.
- Datos: Las variables deben ser cuantitativas.

De acuerdo al argumento de Harrington (2012, 67), existen diferentes tipos de regresión, como la lineal, la logística, y los vectores de soporte, entre otros. Explica que en la regresión lineal se pueden sumar las variables de entrada y multiplicarlas por una constante de salida. Se busca construir un modelo lineal obteniendo una línea recta. Se trata de una línea recta que divide los puntos de datos en un diagrama de dispersión, con el objetivo de minimizar la distancia total entre la línea de regresión y los puntos del diagrama, asegurando que cada punto esté lo más cerca posible de la línea de regresión (Theobald 2017).

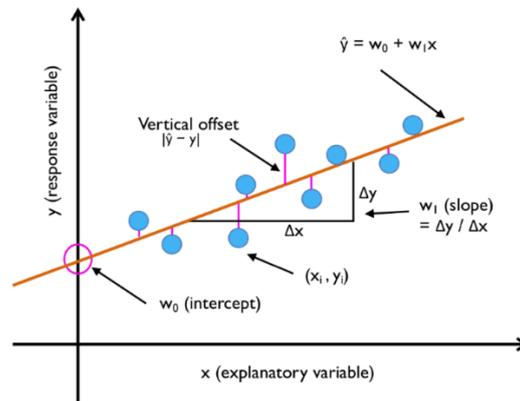


Figura 6. Regresión lineal  
Fuente: (wyj 2023).

#### 4.1. Regresión lineal simple

La regresión lineal simple, según Tizne Ondiviela (2024), también es conocida como modelo de dos variables o *bivariada*. Este tipo de regresión considera que la variable “Y” depende de la variable “X”, representándose mediante la ecuación:

$$Y = \beta_0 + \beta_1 X$$

En donde  $\beta_0$  es el parámetro de intercepto y  $\beta_1$  es el parámetro de pendiente.

Los valores desconocidos de estas estimaciones se denominan coeficientes de regresión, ya que minimizan la suma de los cuadrados de los residuales (Zhang y O'Donnell 2020, 23).

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \bar{\beta}_1 \bar{x}$$

En la ecuación,

$x_i$  : valor individual de la variable independiente (predictora).

$\bar{x}$  : media (promedio) de todos los valores de x.

$y_i$  : valor individual de la variable dependiente (respuesta).

$\bar{y}$  : media (promedio) de todos los valores de y.

n: número total de observaciones

## 4.2. Regresión lineal múltiple

La regresión lineal múltiple es un modelo estadístico versátil que permite evaluar la relación entre una variable dependiente continua y varios predictores. Estos predictores pueden ser continuos, derivados o categóricos. En este modelo, se considera que múltiples factores influyen en la variable dependiente, y se expresa como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

Donde,

$\beta_0$ : Es el intercepto,

$\beta_1$ : Es el coeficiente asociado a  $X_1$ ,

$\beta_2$ : Es el coeficiente asociado a  $X_2$ .

## 5. Support Vector Regression (SVR)

La regresión de vectores de soporte, conocida como SVR por sus siglas en inglés, es una técnica de aprendizaje automático supervisado utilizada para abordar problemas de regresión. Este método es útil para analizar la relación entre una variable dependiente y una o más variables predictoras. SVR es una extensión del algoritmo de clasificación *Support Vector Machine* (SVM), que se enfoca en formular un problema de optimización para aprender una función de regresión que asigna variables predictoras de entrada a valores observados de salida.

SVR aplica la idea fundamental de SVM: una máquina de núcleo disperso que clasifica utilizando un hiperplano definido por unos pocos vectores de soporte. La optimización en SVR se representa en términos de estos vectores de soporte, un pequeño conjunto de muestras de datos de entrenamiento, lo que hace que la solución de optimización no dependa de la dimensión de los datos de entrada, sino solo de la cantidad de vectores de soporte (Zhang y O'Donnell 2020, 20).

SVR tiene ventajas sobre otros métodos de regresión. Con el uso de un kernel, SVR puede manejar problemas de regresión no lineal de manera eficiente al proyectar la función original en un espacio del kernel donde los datos se pueden discriminar linealmente. Además, aprende un modelo que describe la importancia de una variable en la relación entre entradas y salidas. En los métodos de regresión tradicionales, se deben hacer suposiciones sobre la distribución de los datos de entrada, sin proporcionar coeficientes significativos o intervalos de confianza cuando la relación subyacente es no lineal. SVR, por otro lado, maximiza la precisión predictiva calculando un intervalo de confianza para la importancia de una variable (Glaser et al. 2019).

En la clasificación SVM, a partir de un conjunto de datos de entrenamiento, cada muestra es considerada como un punto en un espacio multidimensional de características. En este espacio, se determina un hiperplano que permita clasificar correctamente la mayor cantidad posible de muestras. Las nuevas muestras son clasificadas según su ubicación respecto a este hiperplano. La optimización se realiza mediante la maximización del margen entre los vectores de soporte, es decir, aquellos puntos más cercanos al hiperplano.

Para la regresión de datos, la técnica SVR emplea una función de pérdida insensible a  $\varepsilon$  (épsilon), con el fin de determinar un hiperplano de tal manera que los valores predichos para las muestras de entrenamiento tengan una desviación máxima de  $\varepsilon$  respecto a sus valores observados. En la Figura 9, la Figura A muestra un modelo SVR donde la función de regresión puede aproximar todos los puntos de datos de entrada, mientras que la Figura B muestra un modelo con variables de holgura  $\varepsilon$  y  $\varepsilon^*$  que evidencian datos (errores) fuera de los límites del hiperplano.

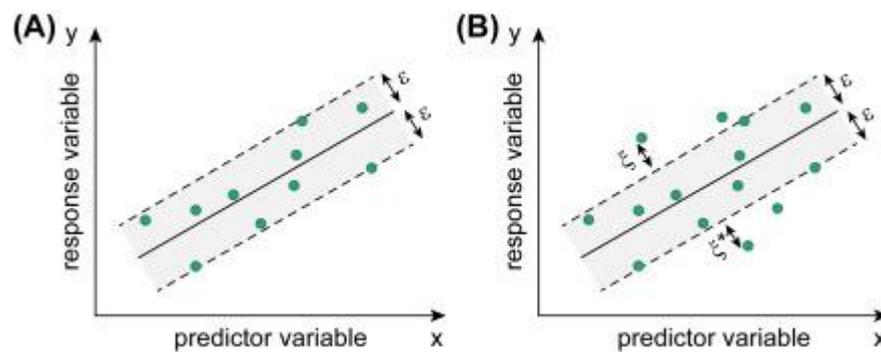


Figura 7. Representación gráfica de modelos lineales  $\varepsilon$ -SV  
Fuente: (Zhang y O'Donnell 2020).

En resumen, el hiperplano  $\varepsilon$  define una banda de generalización para la regresión, minimizando su anchura para incluir la mayoría de las muestras de entrenamiento. La optimización se basa en algunos vectores de soporte, es decir, muestras fuera de la banda  $\varepsilon$ . El modelo de SVR aprendido se utiliza para predecir valores de salida para nuevas muestras.

#### a. Modelo lineal $\varepsilon$ -SVR

El propósito de  $\varepsilon$ -SVR es calcular una función con la restricción de que la estimación de cada punto de los datos de entrada tenga una desviación máxima de  $\varepsilon$  con

respecto a su valor real. Esto se logra formando una banda a una distancia  $\epsilon$  que abarque simétricamente la función estimada. Idealmente, todos los puntos de datos de entrada deberían estar dentro de la banda (con una precisión de  $\epsilon$ ), pero generalmente hay puntos atípicos que se desvían de la mayoría de los datos. Por esta razón, es necesario un modelo que permita algunos errores de predicción.

### b. Kernel-SVR

Para permitir que  $\epsilon$ -SVR maneje datos no lineales, se puede introducir una función kernel que transforme los datos de entrada originales en un espacio de mayor dimensión, conocido como espacio del núcleo. Esto implica el uso de un núcleo no lineal para mapear implícitamente los datos del espacio de entrada (donde no es posible una separación lineal) a un espacio del núcleo de mayor dimensión (donde los datos pueden separarse mediante un hiperplano lineal). La imagen 14 muestra una representación gráfica de  $\epsilon$ -SVR no lineal, en la cual se emplea una función de mapeo para transformar la característica de entrada, como se muestra en la Figura A, al espacio del núcleo según la Figura B.

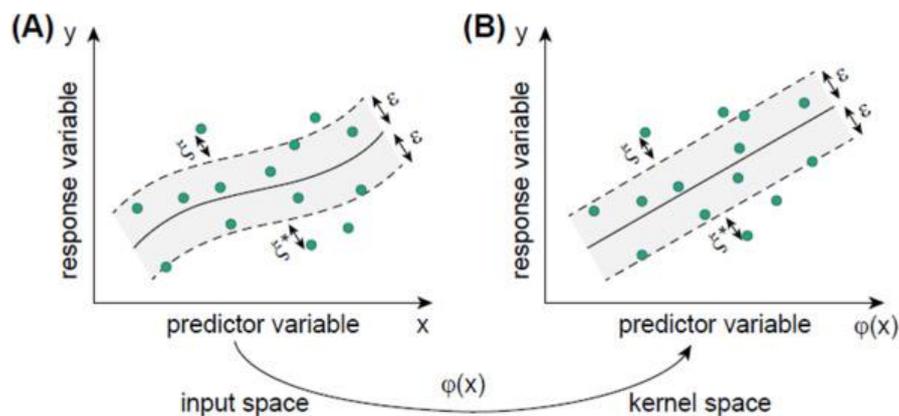


Figura 8. Ilustración de modelos no lineales  $\epsilon$ -SVR  
Fuente: (Zhang y O'Donnell 2020).

### c. Modelo V-SVR

El modelo V-SVR ofrece la ventaja de minimizar automáticamente  $\epsilon$ . En el caso de  $\epsilon$ -SVR, seleccionar un valor adecuado de  $\epsilon$  es crucial para lograr una aproximación precisa en la regresión, lo cual puede ser complicado. Sin embargo, en V-SVR, un nuevo parámetro  $\nu \in (0, 1)$  permite ajustar automáticamente una banda flexible, controlando tanto el número de vectores de soporte como los errores de entrenamiento tolerados. De

este modo,  $\varepsilon$  se convierte en una variable dentro del proceso de optimización, estando regulada por el nuevo parámetro  $v$ .

#### d. Regresión LASSO

Los modelos de regresión LASSO, desarrollados por Hastie y Tibshirani (2015, párr. 6), se basan en el modelo lineal múltiple y buscan regularizarlo. Partimos de la siguiente expresión de un modelo lineal:

$$y = \beta_0 + \sum \beta_j x_{ij} + e$$

El objetivo es predecir el resultado ( $y$ ) a partir de los valores de  $\beta$ , donde  $\beta_0$  y  $\beta_j$  son los parámetros por estimar y  $e$  es un término de error aleatorio. La estimación de estos parámetros se logra minimizando la suma de los cuadrados de los residuos (RSS):

$$\min \beta_0, \beta \sum (y_i - \beta_0 - \sum \beta_j x_{ij})^2$$

Sin embargo, cuando el número de parámetros ( $p$ ) es grande, el modelo puede ser difícil de interpretar y puede sobreajustarse a los datos. Para solucionar esto, LASSO introduce un término de penalización basado en la norma L1 que reduce la magnitud de los coeficientes:

$$\text{RSS} + \lambda \sum |\beta_j|$$

Aquí,  $\lambda$  es un parámetro de ajuste que controla el impacto de la penalización. Un valor alto de  $\lambda$  reduce los coeficientes, y en el límite, algunos coeficientes se vuelven cero, lo que ayuda a identificar un subconjunto más pequeño y relevante de predictores. Esta propiedad es conocida como *sparsity*.

A diferencia de la regresión Ridge, que utiliza la norma L2 y no reduce los coeficientes a cero, LASSO selecciona un conjunto diferente de coeficientes  $\beta$  estimados para cada valor de  $\lambda$ , haciendo crucial la selección adecuada de  $\lambda$ , generalmente mediante validación cruzada.

#### e. La regresión Ridge

La técnica de regresión Ridge, desarrollada por Hoerl en 1970, es un método de regresión que incluye un término de penalización en la función de minimización del error. Esto ayuda a evitar el sobreajuste y reduce el error de generalización:

Donde  $k$  es el coeficiente de Ridge. Al derivar la función y resolver para  $w$ , obtenemos:

$$F(w) = k \|w\|^2 + \sum (y_i - X_i w)^2$$

- $w$ : vector de pesos o coeficientes que el modelo va a aprender.
- $\|w\|^2$ : norma cuadrada del vector  $w$ , es decir, la suma de los cuadrados de sus componentes  $w_1^2 + w_2^2 + \dots + w_p^2$ . Esto penaliza los coeficientes grandes.
- $k$ : coeficiente de Ridge (también llamado  $\lambda$ -lambda o  $\alpha$ -alpha en otros textos). Controla la fuerza de la regularización.
- $X_i$ : fila  $i$  de la matriz de características  $X$ ; representa el vector de características de la muestra  $i$ .
- $y_i$ : valor real de la variable objetivo para la muestra  $i$ .
- $X_i w$ : predicción del modelo para la muestra  $i$ .
- $(y_i - X_i w)^2$ : error cuadrado entre la predicción y el valor real para la muestra  $i$ .

$$w = (X^T X + kI_n)^{-1} X^T y$$

- $X^T$ : transpuesta de la matriz  $X$
- $I_n$ : matriz identidad de tamaño  $n \times n$
- $kI_n$ : regularización agregada a la diagonal (esto estabiliza la inversión de la matriz, especialmente si hay colinealidad entre variables).
- $X^T y$ : correlación entre las características y la salida.

La predicción puede expresarse como:

$$\hat{y}(x) = xw = \sum_{i=1}^n w_i(x)_i$$

- $\hat{y}(x)$ : predicción del modelo para una nueva muestra  $x$

#### f. Regresión logística

La regresión logística multiclase aborda problemas donde hay más de dos clases posibles (por ejemplo, clasificar una opinión como buena, neutral o mala). Se busca construir una función que asigne probabilidades a cada clase. Para esto, se utiliza la función *softmax*, que asegura que las probabilidades sumen uno (Tizne Ondiviela 2024, 5).

En este contexto, se define una matriz de pesos  $w$  y la función de probabilidad estimada para la clase  $k$  dada una entrada  $x$ . Es importante notar que la función softmax es invariante ante traslaciones constantes de los pesos. Para el caso de  $K=2$ , la regresión logística multiclase con función softmax coincide con la regresión logística binaria usando la función sigmoidea (5).

El error en un punto se calcula como el logaritmo negativo de la probabilidad estimada de la clase correcta. La función de coste general extiende esta idea sumando los errores para todos los puntos de datos y clases, y promediando el resultado (5).

Finalmente, la selección adecuada de predictores sigue siendo crucial para construir modelos predictivos precisos, influyendo directamente en la capacidad del modelo para generalizar a datos no vistos (6).

## 6. Predictores

Según Montgomery y Runger (2010, 42), un predictor es cualquier variable que, al ser observada o medida, se utiliza para estimar el valor de la variable dependiente, además señalan que la elección de predictores adecuados es esencial para construir modelos predictivos precisos, ya que influyen directamente en la capacidad del modelo para generalizar a datos no vistos. Por lo tanto, se entiende que los predictores, en el contexto del aprendizaje automático y la estadística, son variables o características que se utilizan para predecir una variable de interés

Neter et al. (2004, 27) ahondan en la importancia de los predictores en la regresión lineal, explicando cómo su selección influye en la calidad del modelo. Los autores discuten la técnica de selección de predictores, como el paso hacia atrás, hacia adelante y la selección basada en la validación cruzada, que ayudan a determinar qué predictores deben ser incluidos en el modelo para evitar el sobreajuste y mejorar la capacidad predictiva. También mencionan que el uso de demasiados predictores irrelevantes puede perjudicar la precisión del modelo debido a la multicolinealidad.

En este marco, se resalta la importancia de los predictores en la construcción de modelos estadísticos y predictivos. Es crucial elegir predictores que estén debidamente relacionados con la variable dependiente, así como seleccionar los predictores adecuados para evitar redundancias y mejorar la interpretabilidad del modelo. Estos enfoques ofrecen un marco teórico y práctico sobre cómo seleccionar y utilizar los predictores de manera efectiva en diferentes tipos de modelos.

### a. Predictores mixtos

Los predictores mixtos se refieren a un conjunto de variables que incluyen tanto características numéricas como categóricas, utilizadas en modelos estadísticos o de aprendizaje automático. De acuerdo a Schapire y Freund (2013, 22) los predictores mixtos son esenciales cuando los datos incluyen una combinación de atributos continuos y

discretos, como en el caso de variables de texto, categorías de productos y mediciones numéricas. Los autores sugieren que la incorporación de predictores mixtos permite crear modelos más flexibles y robustos, capaces de capturar la complejidad inherente de los datos reales.

En el abordaje de este concepto Neter et al. (2004, 92) explican que cuando se manejan predictores mixtos, es fundamental transformar adecuadamente las variables categóricas en variables numéricas para poder incluirlas en los modelos de regresión. Esto se puede lograr mediante técnicas como la codificación de variables dummy o la codificación ordinal. Además, señalan que el tratamiento adecuado de estas variables es crucial para evitar distorsiones en los resultados del modelo, y para que los coeficientes obtenidos sean interpretables y significativos.

Por tanto, los predictores mixtos son cruciales para modelar problemas reales que involucran una combinación de datos numéricos y categóricos. Los conceptos abordados muestran cómo estos predictores pueden ser gestionados eficazmente en diversos tipos de modelos, desde regresión lineal hasta algoritmos más complejos de aprendizaje automático. Aunque el manejo de predictores mixtos puede requerir un tratamiento especial, como la transformación de variables categóricas, su integración en los modelos puede mejorar la precisión y la capacidad predictiva. En la práctica, entender cómo trabajar con predictores mixtos es esencial para la construcción de modelos robustos y aplicables en una amplia variedad de campos.

#### b. Indicadores financieros

Los indicadores financieros consisten en el análisis de los dos estados financieros principales: el estado de resultados y el balance general. Este análisis se realiza mediante la combinación de las partidas incluidas en un solo estado financiero o en ambos. Aquellos indicadores que combinan únicamente partidas del balance general se denominan estáticos, dado que reflejan una fecha específica. En cambio, los indicadores que utilizan únicamente partidas del estado de resultados se consideran dinámicos, porque corresponden a un período determinado. Finalmente, cuando los indicadores combinan partidas provenientes de ambos estados financieros, reciben la denominación de estático-dinámicos.

La finalidad del estudio de estos indicadores son los cuatro indicadores fundamentales de las empresas: solvencia, actividad o productividad, endeudamiento y rentabilidad (Calahorrano, Chacón Guerrero, y Tulcanaza Prieto 2021).

### c. Indicadores de endeudamiento

Según Córdoba (2014), los indicadores de endeudamiento corresponden a medidas de estructura, solvencia y capital que permiten evaluar el grado en que una organización ha sido financiada mediante deuda. Estos indicadores facilitan determinar el monto de recursos provenientes de terceros utilizados para generar utilidades; su importancia radica en que comprometen a la organización a lo largo del tiempo. Entre estos indicadores se encuentran:

*Nivel de endeudamiento:* Mide la proporción del total de activos aportados por los acreedores de la empresa. Fórmula:

$$\text{Nivel de Endeudamiento} = (\text{Total de Pasivos} / \text{Total de Activos}) \times 100$$

*Razón pasiva a capital:* Indica la relación entre los fondos a largo plazo que proveen los acreedores y los contribuyentes de la organización. Fórmula:

$$\text{Razón Pasiva a Capital} = \text{Pasivo a Largo Plazo} / \text{Capital Contable}$$

### Indicadores de rentabilidad

El indicador de rentabilidad es una de las diversas técnicas disponibles para medir las utilidades empresariales. Este indicador puede analizarse en función de las ventas, los activos y la inversión de los accionistas.

*Margen de ganancia bruta:* Mide cuánto de cada dólar de ingresos se queda después de pagar los costos de los bienes vendidos. Fórmula:

$$\text{Margen de Ganancia Bruta} = (\text{Ingresos Netos} - \text{Costo de Bienes Vendidos}) / \text{Ingresos Netos}$$

*Margen de ganancia neta:* Representa la proporción de ingresos que queda después de deducir todos los gastos, impuestos e intereses. Fórmula:

$$\text{Margen de Ganancia Neta} = (\text{Ingresos Netos} - \text{Gastos Totales}) / \text{Ingresos Netos}$$

*Rentabilidad sobre activos (ROA):* Indica cómo de eficiente es la empresa para usar sus activos para generar ganancias. Fórmula:

$$\text{ROA} = \text{Beneficio Neto} / \text{Total de Activos}$$

*Rentabilidad sobre patrimonio (ROE):* Mide la rentabilidad que se obtiene del patrimonio de los accionistas. Fórmula:

$$\text{ROE} = \text{Beneficio Neto} / \text{Patrimonio de los Accionistas}$$

d. Indicadores de liquidez

*Razón corriente:* Mide la capacidad de la empresa para pagar sus pasivos a corto plazo con sus activos a corto plazo. Fórmula:  $\text{Razón Corriente} = \frac{\text{Activos Corrientes}}{\text{Pasivos Corrientes}}$

*Prueba ácida:* Similar a la razón corriente, pero excluye los inventarios de los activos a corto plazo. Fórmula:  $\text{Prueba Ácida} = \frac{(\text{Activos Corrientes} - \text{Inventarios})}{\text{Pasivos Corrientes}}$

e. Indicadores de eficiencia

*Rotación de inventarios:* Mide cuántas veces se ha vendido y reemplazado el inventario durante un período específico. Fórmula:  $\text{Rotación de Inventarios} = \frac{\text{Costo de Bienes Vendidos}}{\text{Inventario Promedio}}$

*Rotación de cuentas por cobrar:* Evalúa la eficacia de la empresa en el cobro de sus cuentas por cobrar. Fórmula:  $\text{Rotación de Cuentas por Cobrar} = \frac{\text{Ventas Netas}}{\text{Cuentas por Cobrar Promedio}}$

*Rotación de cuentas por pagar:* Mide la frecuencia con que la empresa paga sus cuentas por pagar durante un período. Fórmula:  $\text{Rotación de Cuentas por Pagar} = \frac{\text{Compras a Crédito}}{\text{Cuentas por Pagar Promedio}}$

f. Gestión financiera

La gestión financiera implica administrar los recursos disponibles de la empresa para cubrir sus gastos y garantizar su desempeño normal. Está estrechamente vinculada con los procesos de planificación, organización y evaluación de los recursos económico-financieros, con la finalidad de alcanzar los objetivos organizacionales de manera más eficaz y eficiente.

Según propone Andrade (2022, 31), es fundamental que las finanzas se conviertan en un socio estratégico para la dirección general, dado que actualmente resulta esencial que la dirección y la gestión financiera se integren con las demás áreas en la gestión global de la organización.

Por su parte, Gil-Martínez (2018, 31) sostiene que los cambios estratégicos se fundamentan en la reestructuración organizacional. Por ello, la gestión financiera facilita la implementación de estrategias que generen resultados económicos sobresalientes en la entidad, reflejados en su rentabilidad y en el cumplimiento de sus objetivos.

La gestión financiera se organiza bajo dos lineamientos y estos pueden ser:

- La investigación de la presente situación financiera de la compañía.
- Un correcto análisis donde exista una adecuada evaluación y formulación en la toma de decisiones sobre las alternativas que permitirán optimizar el futuro financiero de la empresa.

La importancia de la gestión financiera establece en la identificación de aspectos financieros, como:

- La planificación estratégica fundamentada a los análisis financieros.
- El punto de equilibrio como indicadores financieros.
- El VAN y la TIR.
- Aspectos en temas económicos como la disminución del costo de capital.
- El incremento de los ingresos.
- La disminución de los costos de operaciones de la institución.

Estos aspectos presentan las condiciones en que opera la institución financiera con base al nivel de liquidez, solvencia, endeudamiento, y rentabilidad, proporcionando la toma de decisiones al gestor financiero (Guallasamín 2019, 12).



## Capítulo segundo

### Acopio y procesamiento de información

#### 1. Diagnóstico

##### 1.1. Diagnóstico externo

###### 1.1.1. Análisis PESTEL

Este análisis se realizó mediante reuniones con el personal directivo de la organización, en la cual se consideró los puntos de vista de cada uno de los colaboradores del Centro Médico, basada en su experiencia en cada una de las áreas y conocimiento del negocio (ver Tabla 1).

Tabla 1  
Análisis PESTEL

<b>POLÍTICO</b>		
	<b>AMENAZA</b>	<b>OPORTUNIDAD</b>
<b>Tendencias electorales</b>	Implementación de nuevas leyes del sector salud disminuirá el acceso a ciertas ramas de la medicina	
<b>Regulaciones de comercio</b>	El ingreso de profesionales extranjeros incrementará la competencia	Aprovechar la competencia profesional para capacitarse constantemente y ser competitivos
<b>Impuestos</b>	El incremento del impuesto a la renta afectará directamente a la utilidad del negocio	
<b>Competitividad</b>		Incentivos para fomentar la producción en las pequeñas y medianas empresas
<b>ECONÓMICO</b>		
	<b>AMENAZA</b>	<b>OPORTUNIDAD</b>
<b>Inflación</b>	Incremento de sueldos a nivel sectorial afectará la rentabilidad del negocio	
<b>Importación</b>	Incremento de la inflación influirá en el desarrollo del negocio	
<b>Tasas de interés</b>		El acceso a créditos con tasas de interés bajas facilita el crecimiento del negocio.
<b>Salario mínimo</b>		
<b>SOCIAL</b>		
	<b>AMENAZA</b>	<b>OPORTUNIDAD</b>
<b>Inicio de actividades educativas</b>	Los clientes tienen preferencias en fechas como son las vacaciones escolares e invertir en paseos y no en su salud	

<b>Estructura demográfica</b>		La implementación como requisito el certificado médico para el ingreso a clases genera una nueva oportunidad para el incremento de ingresos
<b>Tendencias de consumo</b>		La confianza de los clientes del segmento de la tercera edad incrementa los beneficios económicos para el negocio
<b>TECNOLÓGICO</b>		
	<b>AMENAZA</b>	<b>OPORTUNIDAD</b>
<b>Maquinaria</b>		Contar con equipos de tecnología de punta permitirá optimizar los tiempos de respuesta al cliente y eficiencia en los resultados específicamente en exámenes de laboratorio, ecos y otros.
<b>Software administrativo</b>	Falta de un software RP que permita el control del negocio	Adquirir un software RP para tener un óptimo control del desarrollo del negocio
<b>LEGAL</b>		
	<b>AMENAZA</b>	<b>OPORTUNIDAD</b>
<b>Reglamentación de MSP</b>	La creación de las leyes penales en el sector de la salud genera incertidumbre en el desarrollo de los profesionales, eximiéndose de algunas prácticas médicas.	

Fuente: Elaboración propia con información de Mediseg.

### 1.1.2. Análisis 5 fuerzas de Porter

Para realizar el análisis de las 5 fuerzas de Porter se solicitó una reunión con el personal Directivo del Centro Médico Mediseg, con la finalidad de discutir cada factor crítico, considerando la experiencia y conocimiento del negocio; una vez recopilada los aportes de cada colaborador directivo se procedió a realizar la siguiente tabla misma que fue aprobada (Tabla 2).

Tabla 2  
5 fuerzas de PORTER de Mediseg

Factor crítico	Situación actual	Oportunidad	Amenaza
<b>Competencia /Sustitutos</b>	Cinco competidores en el mercado en cuanto a centros médicos y siete competidores en farmacias.  Clínicas con mayores ventajas competitivas y franquicias de farmacias	Crear una ventaja competitiva con la ampliación de una farmacia	La existencia de franquicias de farmacias disminuyen las ventas por descuentos y precios bajos.
<b>Propuesta de Valor</b>	Prestación de servicios médicos en consultas de médicos generales y especialistas, servicios de laboratorio y farmacia con precios accesibles al mercado.  Descuentos en venta por volumen especialmente en las franquicias de farmacias.	Mayor demanda de pacientes, permitiendo el crecimiento de la organización	Disminución de la demanda de consumidores

<b>Portafolio – Cliente</b>	Personas de clase media - alta	Mayor demanda de pacientes	
	Consultas con médicos generales, consultas de especialidades previa cita, exámenes de laboratorio clínico y ventas de productos farmacéuticos	Diversificación de servicios y productos	
	Por los profesionales calificados y con alta experiencia	Los clientes prefieren la experiencia y variedad de productos y servicios	
	Productos y servicios de calidad	Confianza en la oferta de servicios y productos	
	Contar con el personal eficiente y con experiencia en las distintas ramas de la medicina	Implementar nuevos servicios, con atención a escuelas, colegios y empresas	
	El nivel de percepción de los clientes es aceptable ya que se visualiza en las recomendaciones que realizan a sus conocidos	Mejorar la calidad de prestación de servicios y descuentos en productos farmacéuticos	Disminución de clientes potenciales y por referencias posibles clientes
<b>Cifras de la empresa y el sector</b>	Descuentos en cirugías diagnosticadas y practicadas por el personal del Centro Médico, Descuentos en productos de alta rotación, Atención de calidad y oportuna	Incrementan el volumen de ventas e ingresos de la organización	
	2016 \$80mil, 2017 \$110mil, 2018 \$180mil	Mayores ingresos para la empresa	
	clientes promedio 300 últimos años	Implementar equipos médicos con tecnología de punta y software administrativo para un mejor control de los pacientes	Incremento de costos en inversión de equipos y sistemas administrativos
	Utilidad \$35mil al año	Reinvertir en el mejoramiento de la infraestructura y equipamiento de la organización	Incremento de impuestos a los ingresos y activos fijos
	Aumenta en un 15% de la participación en los últimos años	Permitirá el crecimiento de la organización	
	Costos accesibles para la comunidad	Incrementar ingresos para la organización	
<b>Proveedores</b>	GENFAR,	Obtener precios de distribuidor de farmacéuticas	
	La empresa tiene una amplia gama de negociación con los proveedores (Descuentos, devolución productos próximos a caducarse, crédito)	Generar descuentos incrementando la venta por volumen y por ende los ingresos	

<b>Factores Críticos de éxito</b>	Ubicación, servicios a domicilio y variedad de productos, infraestructura, clientes potenciales y preferencia en el mercado. Regulaciones ambientales, cultura de prevención	Mayor demanda de pacientes	Incremento de costos en permisos de funcionamiento y licencias ambientales
-----------------------------------	--	----------------------------	--

Fuente: Elaboración propia con base en información de Mediseg.

## **1.2. Diagnóstico interno**

### **1.2.1. Antecedentes de la empresa**

El Centro Médico Mediseg inicia sus actividades en el año 2011, siendo un consultorio médico ubicado en la parroquia de Conocoto, dentro de un Conjunto Residencial, con solo dos profesionales formados en Medicina General y Obstetricia (previa cita), por la complejidad del ingreso de los pacientes, en el año 2012 deciden arrendar un local comercial en el mismo sector, que sea de mayor acceso.

En el año 2013 debido a su crecimiento y en vista de la necesidad de la comunidad de contar con atención las 24 horas, deciden adquirir su propio edificio en el cuál implementaron el ahora Centro Médico Mediseg. Actualmente cuenta con el servicio permanente de Laboratorio, Farmacia (Turno), Odontología, Medicina General, Traumatología y demás especialidades previa citas.

La infraestructura actualmente es adecuada y construida bajo los requerimientos que exigen los entes de control como el Ministerio de Salud, Arcsa y Ministerio del ambiente, con la finalidad de brindar un buen servicio a la comunidad.

### **1.2.2. Misión**

En MEDISEG, más que cuidar de tu salud, nuestra misión es cuidar de tu vida, poniendo a tu disposición el mejor equipo de profesionales médicos, cuya vocación y compromiso está plenamente orientado al servicio de diagnóstico, prevención y tratamiento de enfermedades con atención personalizada de calidad.

### **1.2.3. Visión**

Ser una casa asistencial modelo en el cuidado y restablecimiento de la salud.

### 1.2.4. Valores institucionales

De acuerdo con la información proporcionada en una entrevista al personal responsable de la gestión administrativa, Mediseg tiene establecido los siguientes valores institucionales: Compromiso, calidad, comunicación, profesionalismo, ética, honestidad y responsabilidad social.

### 1.2.5. Estructura organizacional

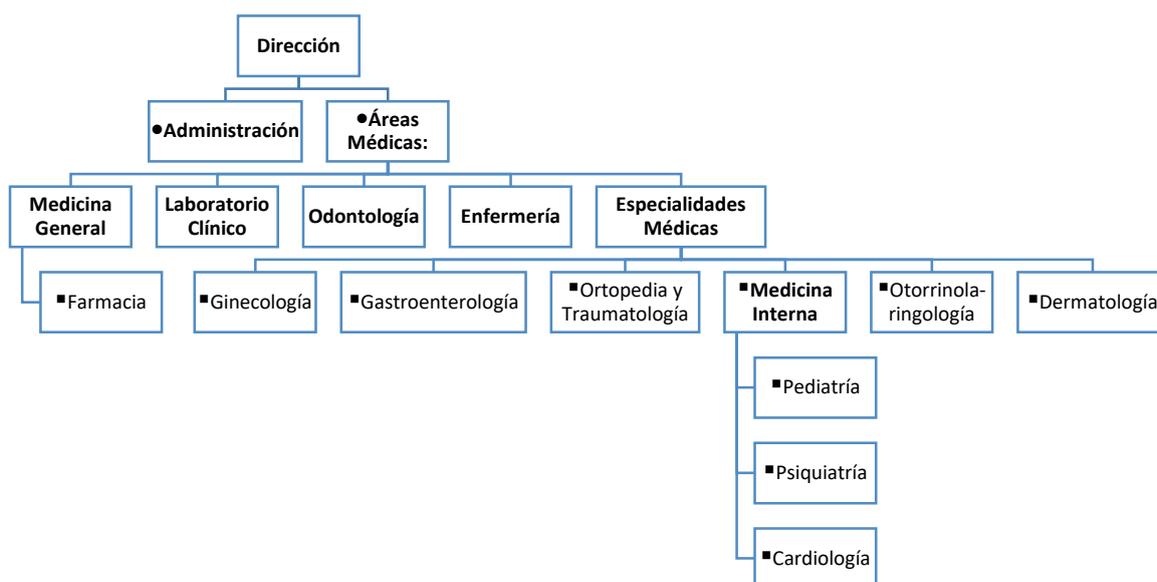


Figura 9. Organigrama

Fuente: Elaboración propia con base en información de Mediseg.

### 1.2.6. Diagrama de distribución

#### Primera planta

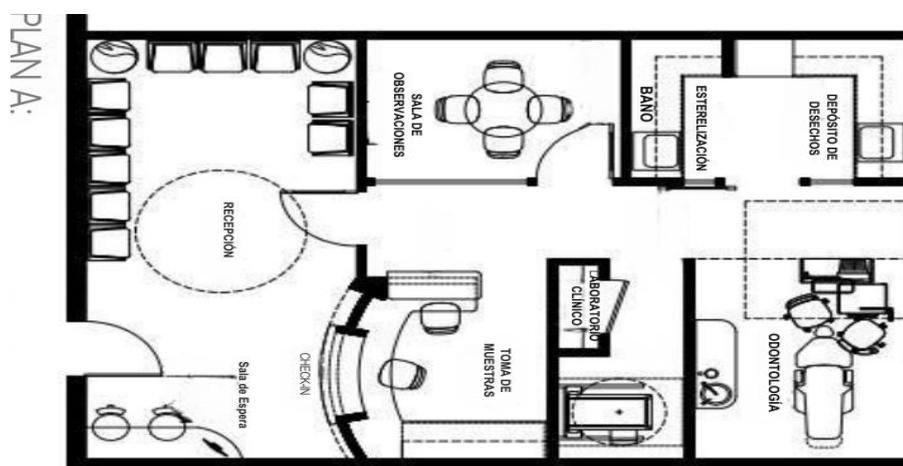


Figura 10. Primera Planta

Fuente: elaboración propia con información de Mediseg.

### Planta alta

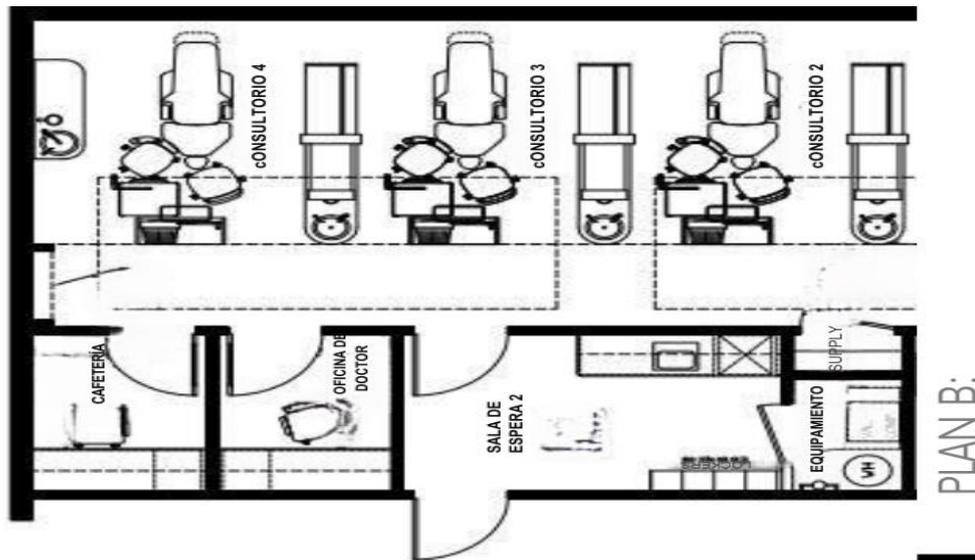


Figura 11. Planta Alta

Fuente: Elaboración propia con base en información de Mediseg.

#### 1.2.7. Servicios

Para determinar los servicios que presta la organización se realizó una visita al Centro Médico, en la cual, mediante entrevista con el Gerente Propietario, quién informó los servicios que ofrecen en cada una de las áreas que mantiene la organización y sus respectivos responsables, con la información proporcionada se realizó la Tabla 3.

Tabla 3  
Servicios de Mediseg

PERSONAL	SERVICIOS PRINCIPALES
<b>MEDICINA GENERAL</b>	
Dr. Luis O. Segura L. Dr. Christian O. Segura S.	Diagnóstico, prevención y tratamiento de enfermedades, examen físico, cirugías menores, infiltraciones, suturas, curaciones, entre otros servicios inherentes a la rama.
<b>ODONTOLOGÍA</b>	
Dra. Paola S. Segura S.	Diagnóstico, prevención y tratamientos odontológicos: profilaxis, extracciones dentales, blanqueamiento, restauraciones, endodoncias, sellantes, prótesis, entre otros servicios inherentes a la rama.
<b>LABORATORIO CLÍNICO</b>	
Lic. Nansi Córdova	Exámenes de sangre, orina, heces (hormonas, inmunoserología, test de embarazo, paternidad (ADN), exámenes especiales, entre otros servicios inherentes a la rama).
<b>ENFERMERÍA</b>	
Aux. Gloria Sangucho	Aplicación de inyecciones, sueros, nebulizaciones, control de peso, talla e índice de masa corporal, entre otros servicios inherentes a la rama.

Fuente: Elaboración propia con base en información de Mediseg.

Tabla 4  
Especialidades médicas de Mediseg

<b>ESPECIALIDADES MÉDICAS:</b>		
Cirugía general	Dr. Francisco Salgado	Diagnóstico, prevención y tratamiento de enfermedades que requieren atención especializada.
Ginecología	Dr. Daniel Vergara	
Gastroenterología	Dra. Jahayra Villamarín	
Ortopedia y Traumatología	Dr. Christian León	
Medicina interna	Dra. Magali Villacís Dr. Adrián Díaz	
Pediatría	Dra. Carolina Rodríguez	Atención PREVIA CITA
Cardiología	Dr. Carlos Lara	
Otorrinolaringología	Dra. Sofía Sánchez	
Dermatología	Dra. Ángela González	
Urología	Dra. Jhoana Nasimba	
Psiquiatría	Dra. Erika Quimbiulco	

Fuente: Elaboración propia con base en información de Mediseg.

### 1.2.8. Procesos

Mediseg no cuenta con una definición gráficamente y por escrito de los procesos que se realizan en cada una de sus áreas, por lo que en conjunto con el Gerente Propietario y el responsable de la Administración se definieron los siguientes macroprocesos:

- Gestión Administrativa
- Gestión Medicina General
- Gestión Odontológica
- Gestión de Laboratorio
- Gestión de Farmacia

A continuación, se detalla los diagramas de los procesos realizados por Mediseg, obtenidos mediante entrevista con el personal responsable de cada área, en la cual describieron los procedimientos que efectúan en el desempeño de sus labores y con esta información se elaboró los respectivos diagramas.

### Gestión Administrativa

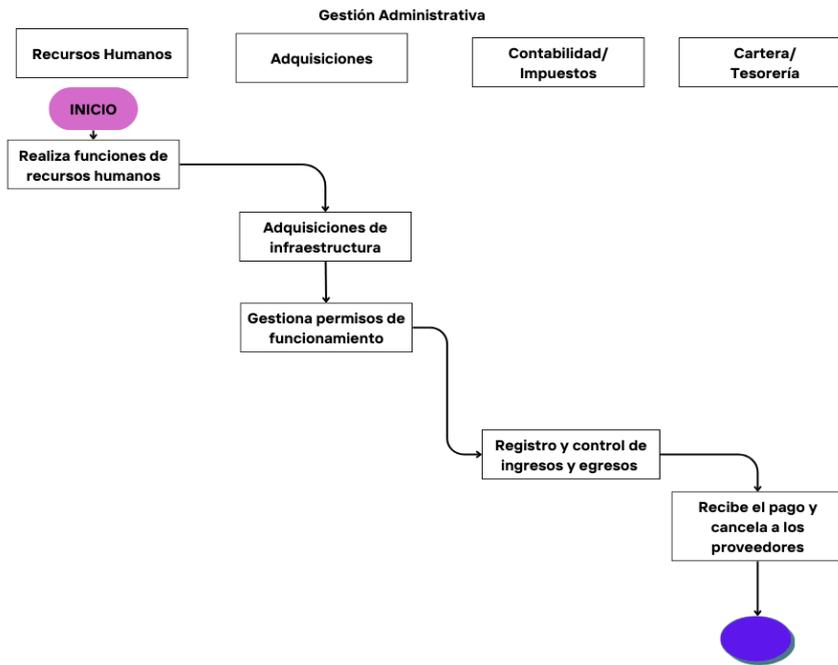


Figura 12. Diagrama de procesos de gestión administrativa  
 Fuente: Elaboración propia con base en información de Mediseg.

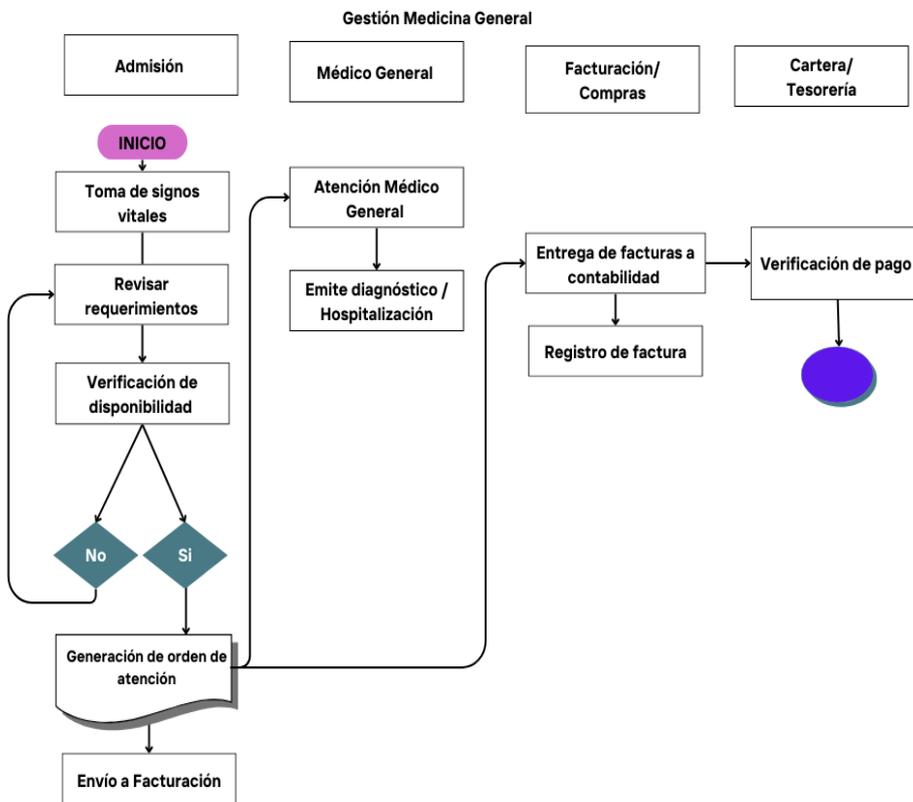


Figura 13. Diagrama de procesos de gestión de medicina general  
 Fuente: Elaboración propia con base en información de Mediseg.

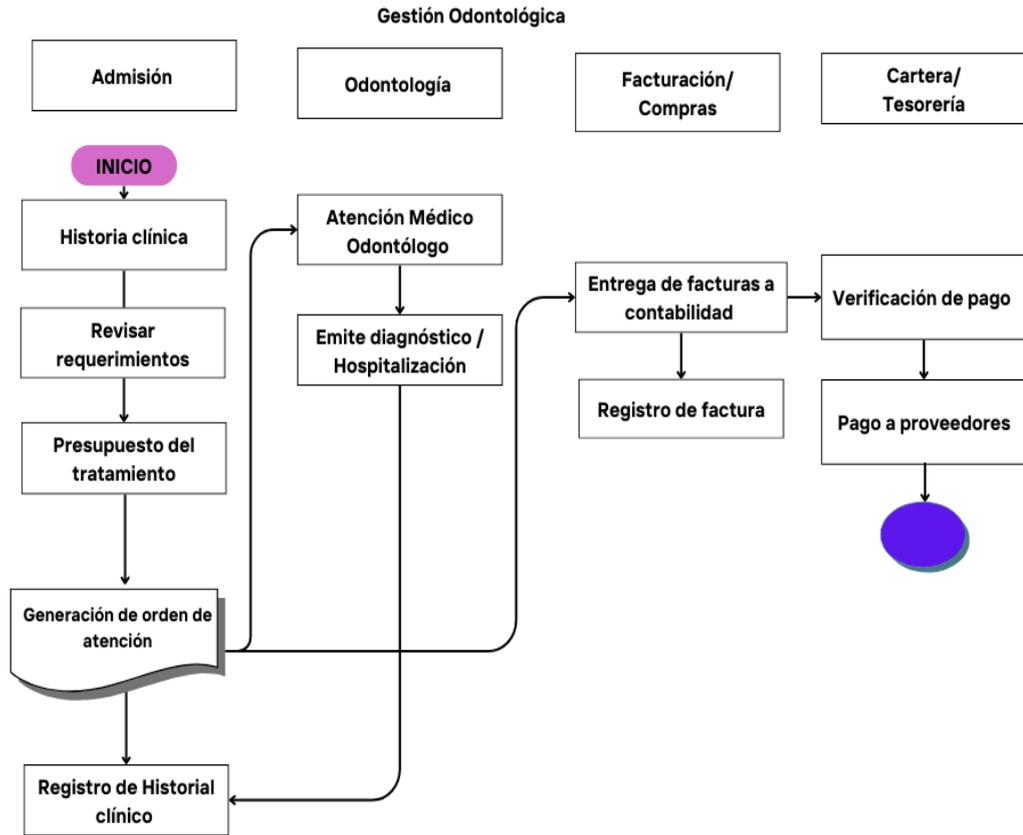


Figura 14. Diagrama de procesos de gestión de medicina general  
 Fuente: Elaboración propia con base en información de Mediseg.

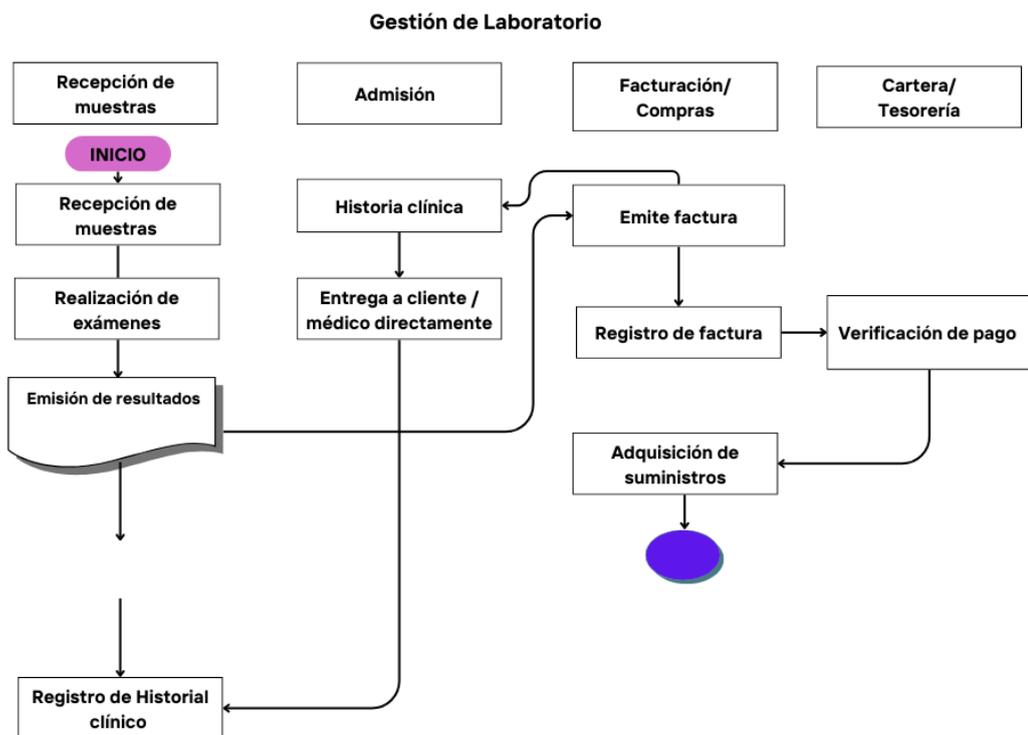


Figura 15. Diagrama de procesos de gestión de Laboratorio  
 Fuente: Elaboración propia con base en información de Mediseg.

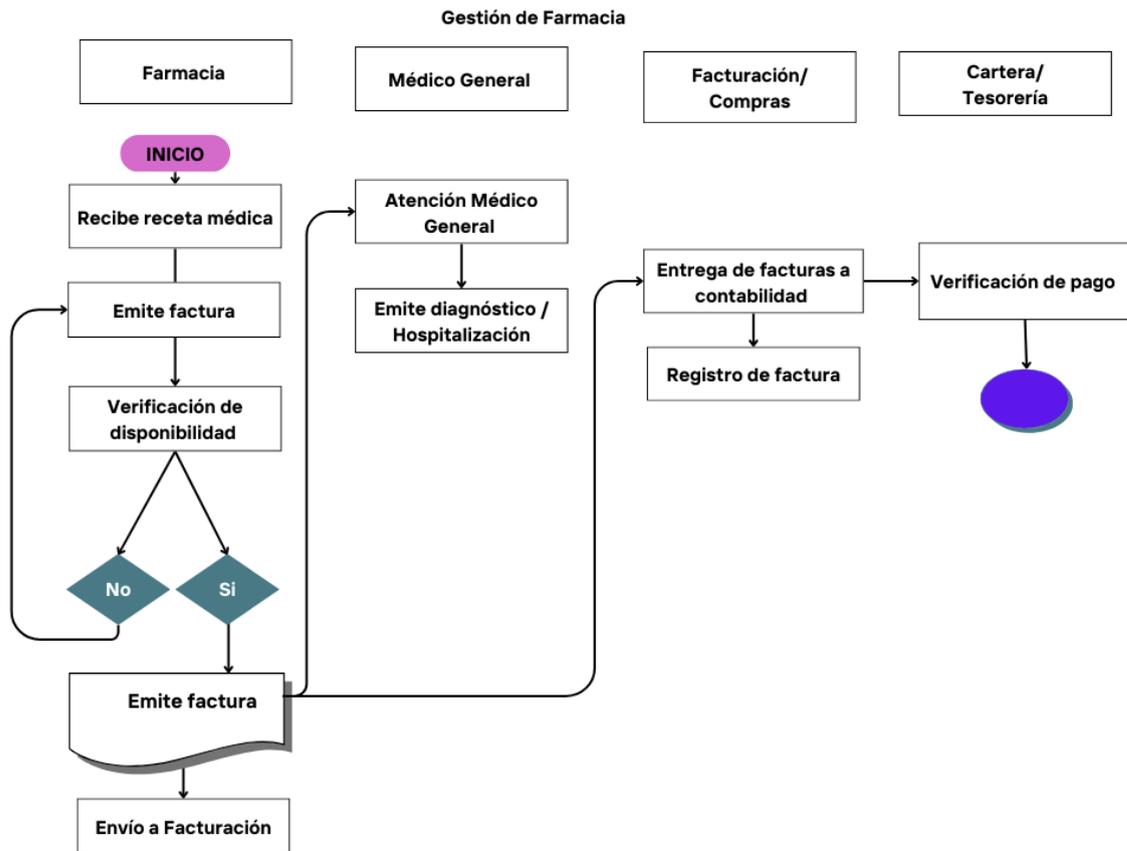


Figura 16. Diagrama de procesos de Farmacia  
Fuente: Elaboración propia con base en información de Mediseg.

### 1.2.9. Marketing

En los últimos dos años Mediseg para incrementar su mercado, se ha enfocado en estrategias de marketing digital, como son la elaboración de la página web, redes sociales como Facebook (Centro Médico Mediseg), en las cuales promocionan los servicios ofertados, así como establecen promociones atractivas, principalmente en época escolar; a través de estos medios los directivos de la organización esperan atraer más clientes e incrementar su rentabilidad.

### 1.2.10. Análisis financiero

El control financiero es un problema para esta organización, debido a la inexistencia de un sistema ERP que vincule todas las áreas de Mediseg y obtener resultados reales y precisos para tomar decisiones oportunas y acertadas.

Para determinar la situación financiera del Centro Médico Mediseg, existió total apertura de la administración para obtener la información necesaria como es el registro de ingresos y egresos.

Los precios generales de consulta están establecidos de la siguiente manera:

- Consulta médica Medicina General \$20
- Consulta odontológica \$10
- Consulta con especialidades médicas \$40

De acuerdo con lo indicado por el Gerente Propietario de la organización los precios de los servicios que oferta el centro médico son accesibles, manteniendo así una competitividad basada esencialmente en la calidad de atención y servicio. Los precios pueden variar de acuerdo al servicio y/o requerimiento adicional que solicite el paciente.

Los resultados del año 2024 se muestran en la Tabla 5:

Tabla 5  
Flujo de efectivo de Mediseg

<u>INGRESOS</u>	Nombre comercial	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	septiembre	Octubre	Noviembre	Diciembre	Total
Farmacia	Farmaseg	2500,00	2513,00	2526,07	2539,20	2552,41	2565,68	2579,02	2592,43	2605,91	2619,46	2633,08	2646,78	30873,05
Medicina General	Mediseg	5000,00	5026,00	5052,14	5078,41	5104,81	5131,36	5158,04	5184,86	5211,83	5238,93	5266,17	5293,55	61746,09
Laboratorio	Labseg	550,00	552,86	555,73	558,62	561,53	564,45	567,38	570,34	573,30	576,28	579,28	582,29	6792,07
Especialidades	Mediseg	150,00	150,78	151,56	152,35	153,14	153,94	154,74	155,55	156,35	157,17	157,99	158,81	1852,38
<b>TOTAL INGRESOS</b>		<b>8200,00</b>	<b>8242,64</b>	<b>8285,50</b>	<b>8328,59</b>	<b>8371,89</b>	<b>8415,43</b>	<b>8459,19</b>	<b>8503,18</b>	<b>8547,39</b>	<b>8591,84</b>	<b>8636,52</b>	<b>8681,43</b>	<b>\$101.263,60</b>
<b>EGRESOS</b>														
Adquisición Suministro	Mediseg	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	1200,00
Adquisición Inventario	Farmaseg	3250,00	3266,90	3283,89	3300,96	3318,13	3335,38	3352,73	3370,16	3387,69	3405,30	3423,01	3440,81	40134,96
Adquisición Inventario	Labseg	55,00	55,29	55,57	55,86	56,15	56,44	56,74	57,03	57,33	57,63	57,93	58,23	679,21
Nómina		1450,00	1450,00	1450,00	1450,00	1450,00	1450,00	1450,00	1450,00	1450,00	1450,00	1450,00	1450,00	17400,00
Publicidad		40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	480,00
Servicios básicos		120,00	120,00	120,00	120,00	120,00	120,00	120,00	120,00	120,00	120,00	120,00	120,00	1440,00
Arriendo		250,00	250,00	250,00	250,00	250,00	250,00	250,00	250,00	250,00	250,00	250,00	250,00	3000,00
<b>TOTAL EGRESOS</b>		<b>5265,00</b>	<b>5282,19</b>	<b>5299,46</b>	<b>5316,83</b>	<b>5334,28</b>	<b>5351,83</b>	<b>5369,47</b>	<b>5387,20</b>	<b>5405,02</b>	<b>5422,93</b>	<b>5440,94</b>	<b>5459,04</b>	<b>\$ 64.334,17</b>
<b>UTILIDAD</b>														<b>\$ 36.929,43</b>

Fuente: Elaboración propia con base en información de Mediseg.

### 1.2.11. Modelo de negocio

Para determinar el modelo de negocio, se realizó un taller con el personal del centro médico, quienes intervinieron en cada uno de los elementos aportando con sus conocimientos en base al negocio actual, y se plasmaron en la siguiente matriz.



Figura 17. Modelo de Negocio Canvas Mediseg  
Fuente: Elaboración propia con información de Mediseg.

### 1.2.12. Matriz FODA

Tabla 6  
FODA

Fortalezas	Oportunidades
<ol style="list-style-type: none"> <li><b>Distribución funcional:</b> La disposición de áreas como la sala de espera, consultorio, laboratorio y emergencias permite un flujo eficiente de pacientes y personal.</li> <li><b>Separación de espacios clave:</b> Cuenta con zonas bien definidas para exploración, observación, muestras y desechos, lo que favorece el cumplimiento de normas sanitarias.</li> <li><b>Accesibilidad y circulación:</b> El diseño con pasillos bien ubicados permite un tránsito fluido entre áreas críticas.</li> <li><b>Espacios específicos para cada función:</b> Incluye laboratorios, consultorios y áreas de descanso, lo que optimiza el uso del espacio y facilita la atención médica.</li> </ol>	<ol style="list-style-type: none"> <li><b>Optimización de señalización y flujo de pacientes:</b> Implementar mejor señalización para facilitar la orientación dentro del centro.</li> <li><b>Expansión de servicios:</b> Se pueden añadir más consultorios o áreas de especialidad según la demanda.</li> <li><b>Uso eficiente de iluminación y ventilación:</b> Incorporar más ventanas o iluminación natural en áreas clave.</li> <li><b>Automatización de procesos:</b> Integrar tecnologías digitales para mejorar la gestión de citas y atención al paciente.</li> <li><b>Calidad en el servicio:</b> La confianza de los clientes del segmento de la tercera edad incrementa los beneficios económicos para el negocio.</li> <li>Acceso a créditos con tasas de interés preferenciales facilita el crecimiento del negocio.</li> </ol>
Debilidades	Amenazas
<ol style="list-style-type: none"> <li><b>Espacio limitado en algunas áreas:</b> Puede haber congestión en la sala de espera o en zonas de tránsito en horas pico.</li> <li><b>Ubicación de baños:</b> Dependiendo del flujo de pacientes y personal, puede ser necesario más de un baño o una mejor ubicación estratégica.</li> </ol>	<ol style="list-style-type: none"> <li><b>Normativas sanitarias y regulaciones:</b> Pueden surgir cambios en las regulaciones que requieran modificaciones en el diseño.</li> <li><b>Crecimiento de la demanda:</b> Si aumenta el número de pacientes, la infraestructura podría quedar insuficiente.</li> </ol>

- 
- |   |   |
|---|---|
| <p><b>3. Accesibilidad para personas con movilidad reducida:</b> Se debe verificar que todas las áreas cumplan con normativas de accesibilidad.</p> | <p><b>3. Mantenimiento de equipos y espacios:</b> La gestión de residuos y el mantenimiento de laboratorios deben estar controlados para evitar problemas operativos.</p> |
|   | <p><b>4. Competencia con otros centros médicos:</b> Es importante diferenciarse con mejor atención, tecnología y comodidad en las instalaciones.</p>                      |
|   | <p><b>5. Cambios en regulaciones:</b> Incremento de costos en permisos de funcionamiento y licencias ambientales.</p>   |
- 

Fuente: Elaboración propia con información de Mediseg.

Mediante el análisis del FODA (efectuado con el personal de la organización) se concluye lo siguiente:

- La normativa legal vigente en el ámbito de la salud, son de alta responsabilidad para los galenos, lo cual en algunas de las ocasiones impide a prestar los servicios en cualquier tipo de situación.
- El incremento de la tasa de impuesto a la renta es una amenaza significativa para la organización, puesto que afecta directamente a la rentabilidad del negocio.
- La apertura y regulación de los entes de control en el tema de capacitación constante, contribuye al desarrollo y potencial de los profesionales de la organización.
- El fácil acceso del mercado de softwares ERP, ayudará a la organización a adquirir uno de estos sistemas para mantener un mejor control de los recursos.
- La ubicación estratégica y la infraestructura de la organización contribuyen para el desarrollo del negocio.
- El personal altamente capacitado y calificado, es competente para la atención de los usuarios.
- Por el factor económico los stocks de inventario de productos farmacéuticos son ineficientes.
- Ausencia de promoción y publicidad, impiden el crecimiento y desarrollo de la organización.

## Encuesta de satisfacción clientes

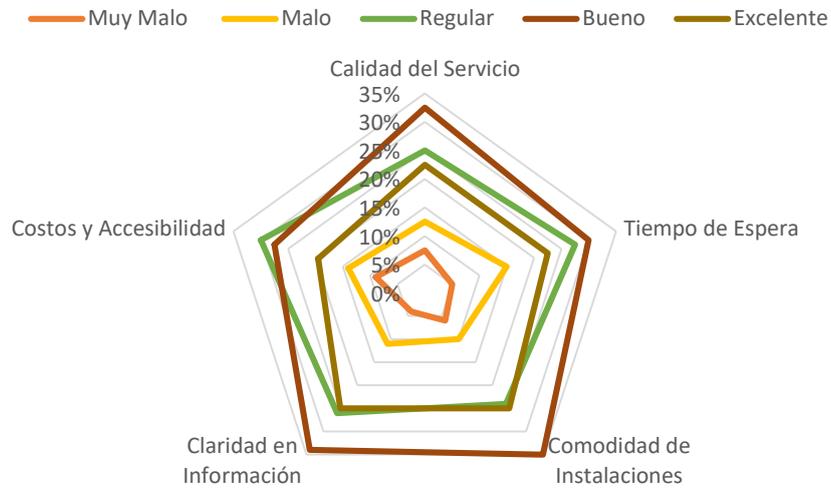


Figura 18. Radial de Satisfacción Clientes

Fuente: Elaboración propia con base en información de Mediseg.

En términos generales, la encuesta muestra que la mayoría de los clientes están satisfechos con el servicio, especialmente en cuanto a la calidad y comodidad de las instalaciones. Sin embargo, los aspectos relacionados con el tiempo de espera, la claridad en la información y los costos presentan áreas de mejora. Se recomienda optimizar los procesos de atención y comunicación para mejorar la percepción del servicio.

## Capítulo tercero

### Presentación y análisis de resultados

#### 1. Propuesta de Diseño de un modelo de *machine learning* basado en predictores mixtos para la estimación del beneficio neto en MEDISEG

En este capítulo, se presenta una propuesta de diseño de un modelo basado en predictores mixtos para la estimación del beneficio neto en el Centro Médico MEDISEG. Para ello, se identificaron variables clave, se evaluó la calidad de los datos y se seleccionaron algoritmos adecuados para el análisis predictivo.

La Figura 19 presenta la estructura del flujo de trabajo propuesto para el desarrollo del modelo, abarcando desde la recolección y preparación de datos hasta la fase final de ejecución de modelo.

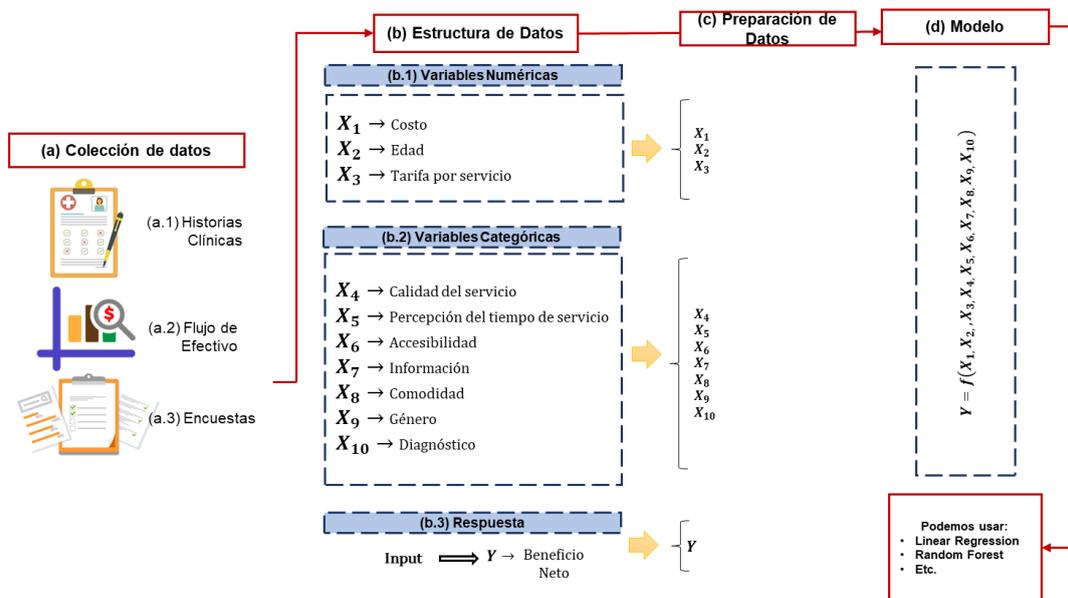


Figura 19. Estructura de las etapas del modelo

Fuente: Elaboración propia con base en información de Mediseq.

El enfoque metodológico integra la combinación de datos cuantitativos y cualitativos, permitiendo una predicción más precisa del comportamiento financiero del centro de salud. A lo largo del capítulo, se describen las fases del proceso de modelamiento, desde la depuración de datos hasta la validación del modelo. Con esta propuesta, se busca optimizar la toma de decisiones estratégicas en la administración del

Centro Médico MEDISEG, asegurando mayor eficiencia en la asignación de recursos y mejora en la rentabilidad.

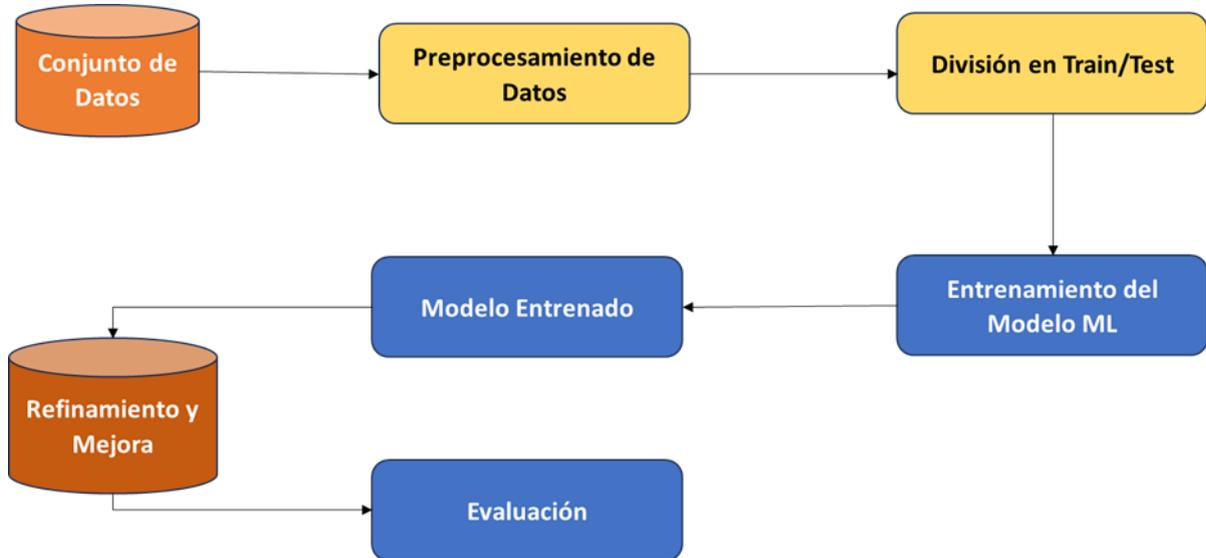


Figura 20. Esquema del desarrollo del modelo

Fuente: Rouhiainen 2020

## 2. Identificación de las Variables Críticas que Influyen en la Estimación del Beneficio Neto en el Centro Médico MEDISEG

### 2.1. Características del conjunto de datos

- **Origen:** Datos operativos y financieros de MEDISEG (2024)
- **Muestra:** 2023 registros (equivalente a 12 meses de operación)
- **Variables:** 10 predictores mixtos (numéricos + categóricos)

Tabla 7  
Detalle de variables

Variable	Tipo de Dato	Escala/Valores	Descripción
<b>INGRESO</b> (Objetivo)	Numérico	USD (continua)	Ingreso generado por servicio (consultas, procedimientos, venta de medicamentos).
<b>COSTO</b>	Numérico	USD (continua)	Costo operativo por servicio prestado (incluye insumos, mano de obra, equipos).
<b>EDAD</b>	Numérico	Entero (años)	Edad del paciente por rangos (niños: <18, adultos: 18–65, tercera edad: >65).
<b>GÉNERO</b>	Categorico	Masculino, Femenino	Sexo del paciente. Codificado mediante one-hot encoding para su uso en modelos de ML.
<b>SERVICIO_ESPECIALIDADES</b>	Categorico	Medicina General, Laboratorio, Rayos X, Odontología, Especialidades, Farmacia	Tipo de servicio utilizado. Codificado como one-hot encoding.
<b>DIAGNOSTICO</b>	Categorico	IVU, EVR, G, N, L	Diagnóstico principal asociado al servicio. Codificado como one-hot encoding.
<b>CALIDAD_DEL_SERVICIO</b>	Ordinal	Likert (1–5): 1=Muy Malo, 5=Excelente	Evaluación subjetiva del paciente post-servicio.
<b>TIEMPO_DEL_SERVICIO</b>	Ordinal	Likert (1–5): 1=Muy Malo, 5=Excelente	Evaluación subjetiva del paciente post-servicio.
<b>INFORMACION_DEL_SERVICIO</b>	Ordinal	Likert (1–5): 1=Muy Malo, 5=Excelente	Evaluación subjetiva del paciente post-servicio.
<b>COMODIDAD_DEL_SERVICIO</b>	Ordinal	Likert (1–5): 1=Muy Malo, 5=Excelente	Evaluación subjetiva del paciente post-servicio.
<b>ACCESIBILIDAD_DE_L_SERVICIO</b>	Ordinal	Likert (1–5): 1=Muy Malo, 5=Excelente	Evaluación subjetiva del paciente post-servicio.

Fuente: Elaboración propia con base en información de Mediseg.

El primer paso en el diseño del modelo de *machine learning* es la identificación de las variables clave que afectan la estimación del beneficio neto en el Centro Médico MEDISEG. Este proceso es fundamental para garantizar que el modelo capture correctamente las dinámicas financieras y operativas de la institución, integrando tanto predictores cuantitativos como cualitativos.

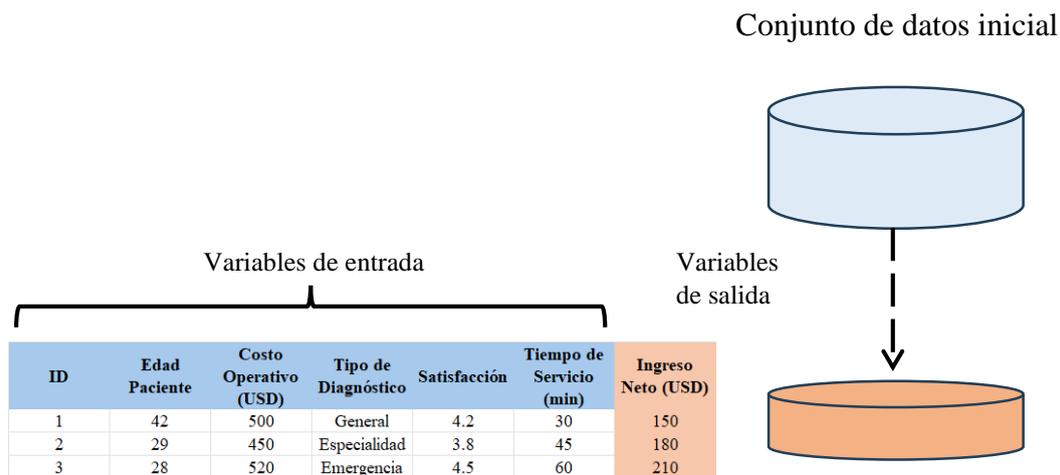


Figura 21. Estructura de conjunto de datos  
 Fuente: Elaboración propia con base en información de Mediseg.

En la Tabla 8 se resume el proceso utilizado para la selección, clasificación y transformación de las variables utilizadas en el desarrollo del modelo predictivo. En la primera etapa, se identificaron tanto las variables predictoras como la variable objetivo del estudio, con base en criterios técnicos y de relevancia teórica en el contexto de análisis. Posteriormente, se clasificaron las variables según su naturaleza (continuas, ordinales y categóricas), lo que permitió definir los métodos adecuados de pre-procesamiento.

Por otra parte, en las variables categóricas se aplicó codificación tipo *One-Hot Encoding*, transformándolas en variables binarias con valores True/False, lo que facilitó su interpretación por parte de los algoritmos de aprendizaje automático.

Tabla 8  
**Etapas de selección de variables**

Etapa	Acción / Variable
Selección de variables	Predictoras: COSTO, SERVICIO, GÉNERO, EDAD, DIAGNÓSTICO, CALIDAD, TIEMPO, INFORMACIÓN, COMODIDAD, ACCESIBILIDAD Objetivo: INGRESO
Tipo de variables	- Continuas: COSTO, EDAD - Ordinales: CALIDAD, TIEMPO, INFORMACIÓN, COMODIDAD, ACCESIBILIDAD

	- Categóricas: SERVICIO, GÉNERO, DIAGNÓSTICO
Codificación	One-Hot Encoding para variables categóricas (binarias True/False)
Resultado final	Dataset numérico, sin datos perdidos, apto para modelos de regresión

Fuente y elaboración propias.

Para la correcta implementación del modelo, es esencial disponer de una base de datos confiable que contenga información relevante sobre los ingresos, costos operativos, número de pacientes, entre otros factores determinantes. La importación de estos datos es el primer paso en el análisis.

- Importación de archivo base: Se carga el dataset con información financiera y operativa de MEDISEG.
- Formato de datos: Se verifica la estructura y consistencia de los datos para su análisis.
- Exploración inicial: Se analizan los tipos de variables y la existencia de valores faltantes.

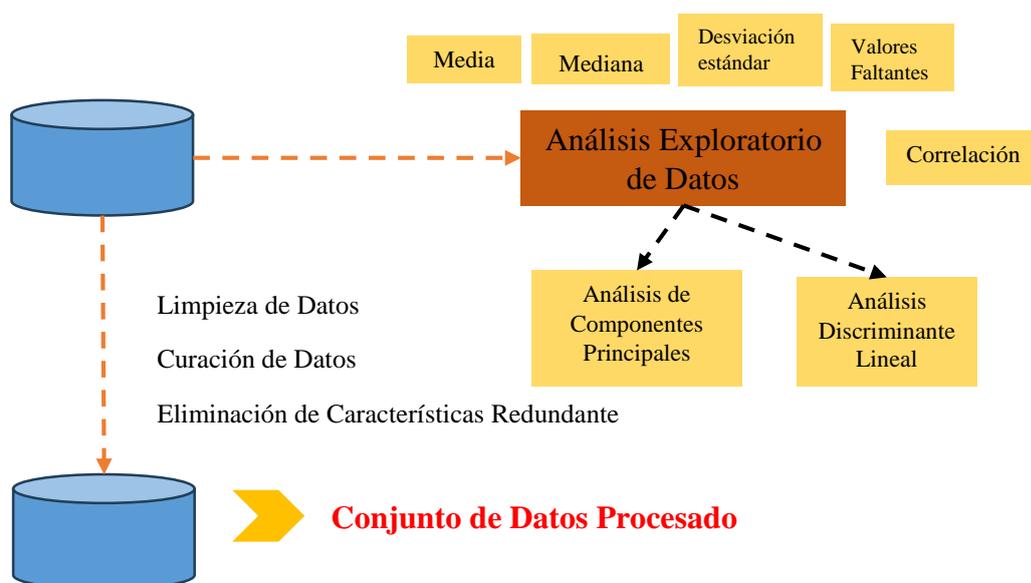


Figura 22. Procesamiento o limpieza de datos  
Fuente y elaboración propias.

Tras la eliminación de valores atípicos, se procede a la limpieza de datos para mejorar la calidad y precisión de las estimaciones. Esta fase implica la corrección de datos faltantes, la conversión de formatos y la normalización de variables.

## 2.2. Proceso de limpieza de datos

- Eliminación de valores nulos: Se eliminan o reemplazan valores faltantes en las variables clave.
- Conversión de formatos: Se convierten valores de texto a numéricos cuando es necesario.
- Normalización de datos: Se ajustan las escalas de variables para mejorar la modelización. Adicionalmente, se generan gráficos exploratorios para identificar patrones y tendencias en los datos, facilitando la selección de las variables más influyentes en la estimación del beneficio neto.

La Tabla 9 presenta las etapas de procesamiento aplicadas a las variables numéricas antes del entrenamiento del modelo de *machine learning*, enfocándose específicamente en los métodos de normalización y estandarización. Estas técnicas se emplearon para mejorar la comparabilidad entre variables y optimizar el desempeño de los algoritmos utilizados.

Tabla 9

Etapas de procesamiento de datos		
Etapa	Descripción	Justificación
Normalización	<p>Escala los datos al rango [0, 1] o [-1, 1], usando la fórmula:</p> $(x - \min) / (\max - \min)$	<p>Se aplicó sobre variables continuas como COSTO y EDAD, para asegurar que los valores estuvieran dentro de un rango comparable</p>
Estandarización	<p>Centra los datos en media 0 y varianza 1:</p> $(x - \mu) / \sigma$	<p>Se aplicó a variables continuas para mejorar la interpretación y desempeño de modelos lineales.</p>

No aplicada a:	VARIABLES CATEGÓRICAS transformadas con One-Hot Encoding	Estandarizarlas o normalizarlas no tiene sentido; ya están listas para ser procesadas por el modelo.
----------------	--	---

---

Fuente y elaboración propias.

### 2.3. Criterios de análisis y evaluación de la calidad de los datos

Para garantizar la precisión y confiabilidad del modelo de *machine learning*, es necesario establecer criterios sólidos de análisis y evaluación de la calidad de los datos utilizados en el proceso de modelado.

Además del procesamiento previo de los diferentes datos ya codificados para una mejor interpretabilidad, homogeneización, y, para una mejor organización y análisis de los datos clínicos, se asignaron códigos específicos a los diagnósticos más frecuentes en los servicios de Medicina General y Especialidades en MEDISEG. La nomenclatura establecida es la siguiente:

- EVR = Enfermedades de Vías Respiratorias
- G = Gastroenteritis
- IVU = Infección de Vías Urinarias
- N = Neuritis
- L = Lumbalgia
- NA = No aplica (Se asignaron a los servicios en los que no se presenta un diagnóstico definido, puesto que solamente los servicios de Medicina General y Especialidades cuentan con diagnósticos específicos).

### 2.4. Identificación y visualización de la variable más crítica

Una vez asegurada la calidad de los datos, se procede a la generación de gráficos que permitan identificar y visualizar la variable más crítica dentro del modelo. Este análisis proporciona una mejor comprensión de la relación entre las variables y su impacto en la estimación del beneficio neto.

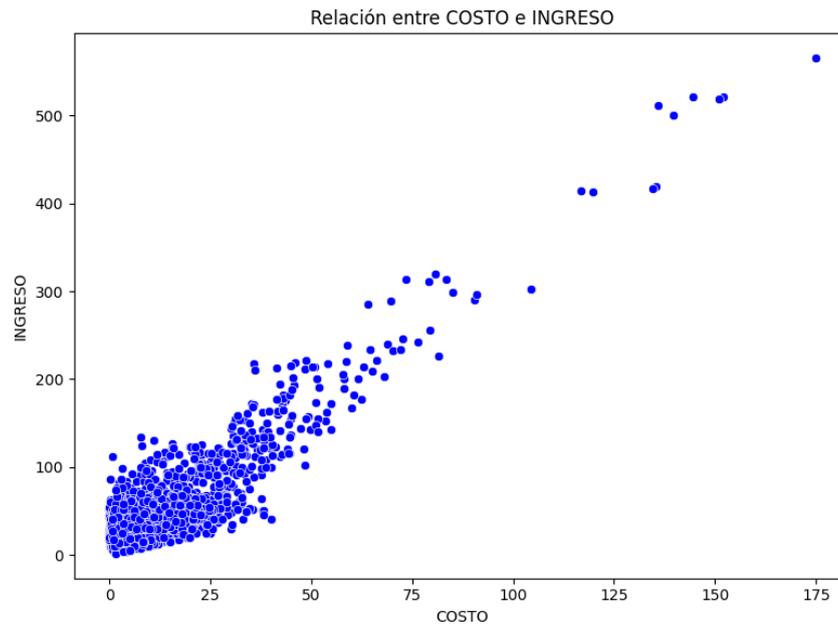


Figura 23. Relación entre Costo e Ingreso en el Centro Médico MEDISEG  
Fuente: Elaboración propia con base en información de Mediseg.

La Figura 23 evidencia una correlación significativa y estructurada entre ambas, lo cual permite identificar al COSTO como una de las variables más críticas dentro del modelo predictivo desarrollado.

Se observa que la tendencia de datos se alinea en torno a una propensión definida, sin dispersión excesiva, lo que indica una asociación directa entre el incremento en los costos operativos y el ingreso generado por la institución.

### **2.5. Relación entre variable edad e ingreso**

Una de las visualizaciones clave en este análisis es la figura de dispersión que muestra la relación entre la edad de los pacientes y los ingresos generados. Este análisis permite determinar si existen patrones en la generación de ingresos según los grupos etarios, lo que puede influir en la toma de decisiones estratégicas en MEDISEG.

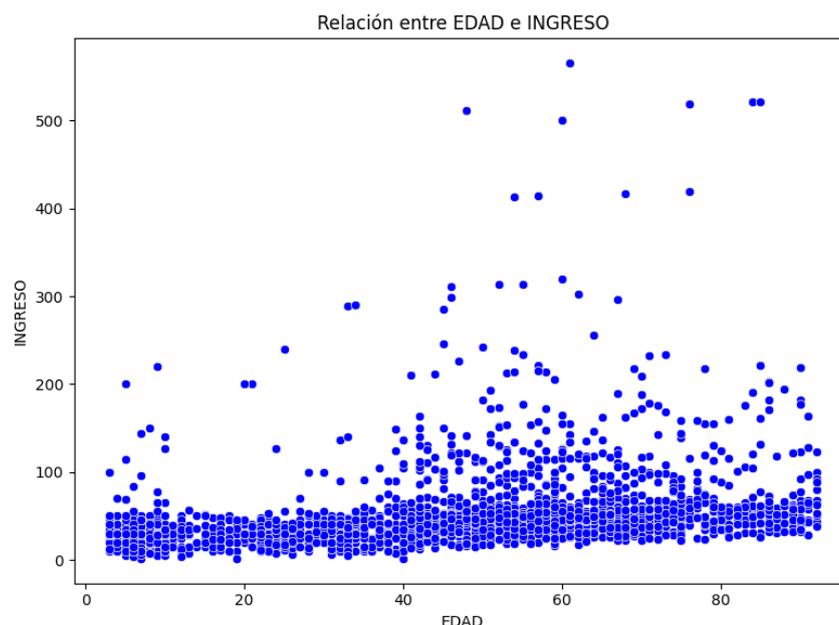


Figura 24. Relación entre Edad e Ingresos en el Centro Médico MEDISEG  
Fuente: Elaboración propia con base en información de Mediseq.

La interpretación de esta figura proporcionará una visión más profunda sobre cómo ciertas características demográficas pueden impactar en la rentabilidad del centro médico, contribuyendo a la construcción de un modelo predictivo más robusto.

Se observa que la relación entre la variable EDAD de los pacientes y el INGRESO generado no muestra una tendencia definida ni una asociación estadísticamente que sea suficientemente significativa, lo cual permite inferir que, aunque la variable EDAD puede aportar información contextual, su contribución como predictor del ingreso es limitada, debido a su baja consistencia, claridad en su tendencia que genera escasa significancia estadística.

Con la identificación y depuración de estas variables críticas, se sienta la base para la siguiente fase del diseño del modelo de *machine learning*, en la que se definirán los algoritmos y técnicas a utilizar para la estimación del beneficio neto en MEDISEG.

## 2.6. Definición e integración de la variable categórica de satisfacción del cliente

Finalmente, se incorporó la variable categórica Satisfacción del Cliente para evaluar la percepción del servicio por parte de los usuarios. Esta variable se utilizará para medir la calidad del servicio y su influencia en la rentabilidad del centro médico,

proporcionando información valiosa para la toma de decisiones estratégicas en MEDISEG.

Con la correcta identificación y validación de estos criterios de análisis, el modelo de *machine learning* contará con datos de alta calidad, asegurando predicciones más precisas y confiables para la estimación del beneficio neto en el centro médico.

a. Definición del conjunto final de datos para modelamiento

Para la etapa de modelado, se selecciona el conjunto final de datos a partir de la base depurada, almacenada en un archivo denominado *cleandata.csv*. Este conjunto de datos contiene únicamente las variables más relevantes, las cuales han sido previamente identificadas y filtradas mediante técnicas estadísticas y exploración de datos.

Se lleva a cabo un proceso de selección de características con el fin de garantizar que solo las variables más influyentes sean utilizadas en el modelo. Esta selección es clave para mejorar la eficiencia computacional y la precisión del modelo predictivo.

Posteriormente, los datos se dividen en dos conjuntos: datos de entrenamiento y datos de prueba, permitiendo evaluar el rendimiento del modelo antes de su implementación final. Esta división se realiza aplicando criterios estándar en la ciencia de datos, asegurando que el modelo pueda generalizarse correctamente a nuevos datos.

b. Selección de Algoritmos y Parámetros

Para asegurar la integración de predictores mixtos en el modelo de *machine learning*, se exploran diferentes algoritmos, entre ellos:

- *Random Forest*, que permite manejar datos categóricos y numéricos de manera eficiente.
- *Gradient Boosting Machines (GBM)*, optimizando la precisión mediante aprendizaje incremental.
- Regresión Lineal Múltiple, para evaluar relaciones lineales entre variables.

Además, se realiza un ajuste de hiperparámetros para maximizar la capacidad predictiva del modelo, optimizando factores como la profundidad de los árboles en *Random Forest* y el número de iteraciones en *Gradient Boosting Machine*. La selección

de los mejores parámetros se realiza mediante validación cruzada y evaluación de métricas como el error cuadrático medio (RMSE) y la precisión del modelo.

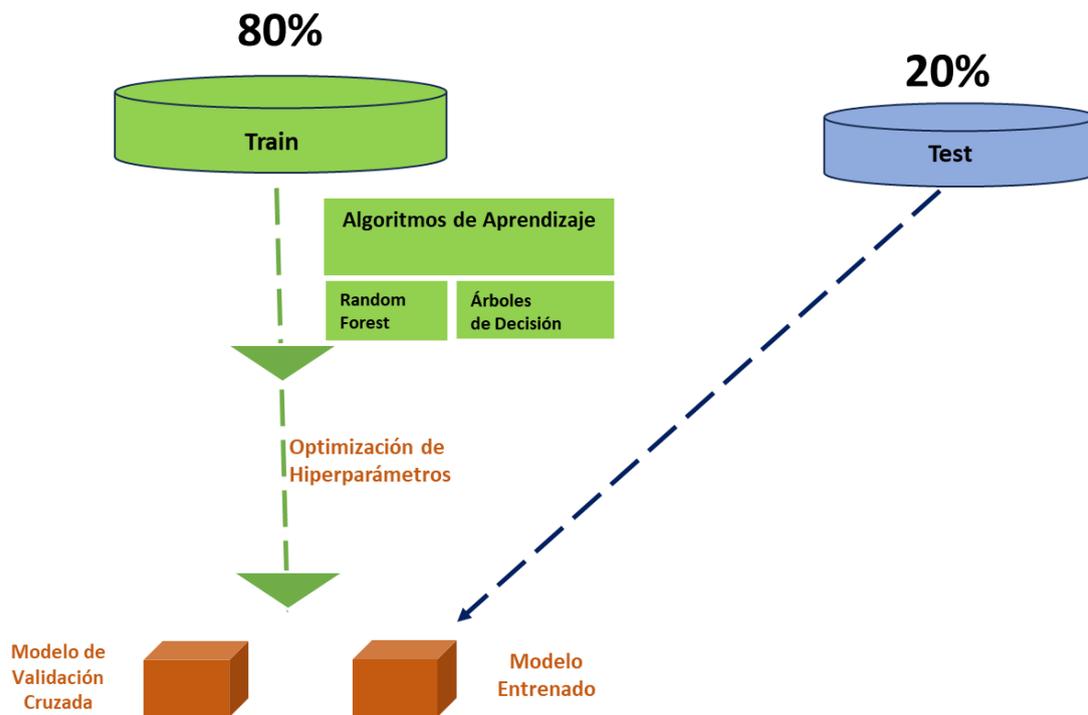


Figura 25. Entrenamiento y evaluación

En la Tabla 10 se resume el procedimiento técnico utilizado para la división del conjunto de datos en etapas de entrenamiento y prueba, paso fundamental en la construcción de modelos de *machine learning*. Esta separación tiene como principal objetivo garantizar una evaluación objetiva del desempeño del modelo al probarlo sobre datos que no han sido utilizados durante su entrenamiento.

La división se realizó mediante la codificación `train_test_split`, asignando el 80% de los datos al conjunto de entrenamiento ( $X_{train}$ ,  $y_{train}$ ) y el 20% restante al conjunto de prueba ( $X_{test}$ ,  $y_{test}$ ), lo que permite evaluar la capacidad de generalización del modelo sobre datos no vistos.

Para asegurar la reproducibilidad de los resultados, se utilizó el parámetro `random_state=42`, lo que permite que la misma división de datos pueda ser replicada en futuras ejecuciones, manteniendo la consistencia del experimento.

Tabla 10  
**Etapas y codificación de testing y training**

<b>Etapas</b>	<b>Descripción Técnica</b>
División del dataset	<code>train_test_split</code> divide los datos en 80% para entrenamiento ( <code>X_train</code> , <code>y_train</code> ) y 20% para prueba ( <code>X_test</code> , <code>y_test</code> ).
Objetivo	Separar las características (X) y el objetivo (y) para entrenar y evaluar el modelo en datos no vistos.
Proporción	<code>test_size=0.2</code> indica que el 20% de los datos se usan para prueba, asegurando evaluación objetiva del modelo.
Reproducibilidad	<code>random_state=42</code> garantiza una división consistente y replicable de los datos en diferentes ejecuciones.

Fuente y elaboración propia

Con la correcta definición del conjunto de datos y la selección de los algoritmos adecuados, se establece la base para la implementación y validación del modelo en la siguiente fase del estudio.

### **3. Evaluación de la precisión e impacto de modelos de *machine learning***

Para evaluar la precisión y el impacto de los modelos de *machine learning* desarrollados en la estimación del beneficio neto en el Centro Médico MEDISEG, se compararon diferentes algoritmos. Se utilizó la métrica  $R^2$  (coeficiente de determinación) para medir la capacidad de ajuste de los modelos, observando su desempeño con base en los datos de prueba.

La evaluación comparativa entre *Random Forest* y Regresión Lineal mostró que el modelo *Random Forest* obtuvo un  $R^2 = 0.76$ , lo que indica un ajuste óptimo a los datos. Este valor sugiere que el modelo es capaz de explicar una parte significativa de la variabilidad del beneficio neto en función de las variables seleccionadas.

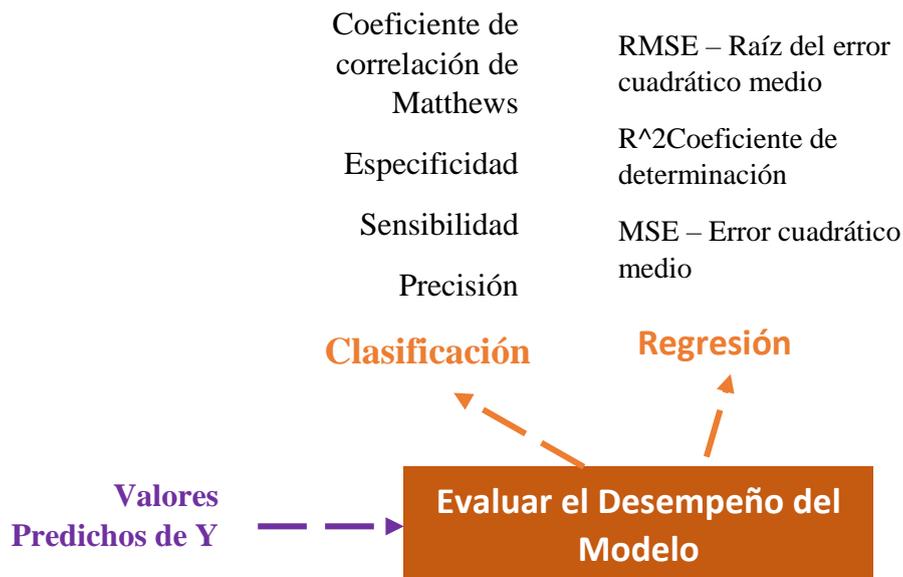


Figura 26. Entrenamiento y evaluación

En la Tabla 11 se presentan las diferentes etapas clave del proceso de entrenamiento y optimización del modelo de *machine learning*, con el objetivo de seleccionar la configuración que ofrece el mejor desempeño predictivo y al mismo tiempo evita el sobreajuste.

Tabla 11  
Etapas de configuración, evaluación y selección del modelo

Etapa	Descripción
Configuración y entrenamiento	Se aplicó validación cruzada ( $cv=5$ ) para evaluar el rendimiento del modelo en distintas combinaciones.
Evaluación del rendimiento	Se usó la métrica $R^2$ para comparar los modelos y seleccionar el mejor conjunto de hiperparámetros.
Selección final	Se seleccionó la mejor combinación que optimizó el rendimiento del modelo: $max\_depth: 10$ → Limita la profundidad de los árboles para evitar sobreajuste. $min\_samples\_leaf: 2$ → Requiere al menos 2 muestras por hoja, lo que da estabilidad al modelo. $min\_samples\_split: 5$ → Controla la complejidad al exigir mínimo 5 muestras para dividir un nodo.

---

n\_estimators: 500 → Se usaron 500 árboles para una mayor robustez en la predicción.

---

Fuente y elaboración propias.

Posteriormente, una vez realizada la selección de variables, en donde se definieron las variables independientes (X) o predictoras y se estableció como variable objetivo (y) el ingreso. Esta selección fue guiada tanto por criterios técnicos como por fundamentos teóricos del contexto del caso analizado; así pues, se efectuó la división del conjunto de datos siguiendo la técnica de validación *hold-out*, asignando el 80% de los datos al conjunto de entrenamiento y el 20% restante al conjunto de prueba, lo que permitió evaluar objetivamente el rendimiento de los modelos sobre datos no utilizados durante el entrenamiento.

En cuanto a los modelos evaluados, se entrenaron dos algoritmos: Regresión Lineal como modelo base, y, *Random Forest* como modelo no lineal de mayor complejidad, estableciendo como objetivo la comparación de enfoques con distintos niveles de flexibilidad y capacidad de modelar relaciones complejas entre las variables.

Para el modelo de *Random Forest* se implementó una etapa de búsqueda de hiperparámetros mediante la técnica de *Grid Search* con validación cruzada (CV=5), lo cual permitió identificar la combinación de parámetros que maximizaba el rendimiento predictivo y minimizaba el riesgo de sobreajuste. Los mejores hiperparámetros encontrados incluyeron: max\_depth = 10: profundidad máxima del árbol, controla la complejidad; min\_samples\_leaf = 2: mínimo de muestras por hoja terminal, aporta estabilidad; min\_samples\_split = 5: mínimo de muestras para dividir un nodo, evita divisiones triviales; n\_estimators = 500: número de árboles, mejora la precisión del modelo.

En la Tabla 12 se presenta una comparación de los resultados obtenidos al evaluar el desempeño de los modelos de Regresión Lineal y *Random Forest* en la predicción del ingreso. Para dicha evaluación se utilizaron dos métricas fundamentales: el Error Cuadrático Medio (MSE) y el coeficiente de determinación ( $R^2$ ).

El modelo de Regresión Lineal fue utilizado como punto de referencia o modelo base. Este modelo asume una relación lineal entre las variables predictoras y el ingreso, y obtuvo un MSE de 507.63 y un  $R^2$  de 0.65, lo que indica una capacidad explicativa moderada. Por otro lado, el modelo de *Random Forest* presentó un mejor desempeño general, con un MSE menor (353.38) y un  $R^2$  más alto (0.76), lo que evidencia una mayor

precisión en las predicciones y una mejor capacidad para capturar relaciones no lineales y complejas entre las variables.

Tabla 12  
Resultados

Modelo	MSE	R <sup>2</sup>	Observaciones
<b>Regresión Lineal</b>	507.63	0.65	Modelo base. Relación lineal moderada.
<b>Random Forest</b>	353.38	0.76	Mejor desempeño. Captura relaciones no lineales.

Fuente y elaboración propias

Estos resultados confirman que el enfoque basado en *Random Forest* no solo mejora la exactitud del modelo, sino que también incrementa su capacidad de generalización al adaptarse a la estructura real de los datos, sin depender exclusivamente de supuestos lineales. La inclusión de este modelo robustece el análisis y aporta mayor confiabilidad en la estimación del ingreso.

En esta evaluación se utilizó como criterio técnico la importancia basada en el índice de impureza de Gini, una métrica que estima la ganancia en pureza (reducción del error) en cada división del árbol. Cuanto mayor sea la reducción del error atribuible a una variable en los distintos nodos, mayor será su importancia dentro del modelo.

El cálculo de la importancia de variables se basa, por tanto, en su impacto acumulado sobre la disminución del error en todas las ramas del conjunto de árboles que conforman el modelo *Random Forest*. Esta técnica permitió identificar, con base empírica, cuáles son las variables que aportan significativamente a mejorar la precisión de las predicciones.

Finalmente, las variables se ordenaron según su valor de importancia, lo que facilita su interpretación y permite priorizar aquellas con mayor capacidad explicativa. Este ordenamiento es fundamental no solo para la comprensión del modelo, sino también para futuras estrategias de toma de decisiones, simplificación del modelo o mejora en la recolección de datos. Se representó gráficamente la significancia de cada variable en la estimación del beneficio neto mediante una figura de barras, proporcionando una visualización clara sobre los predictores más influyentes.

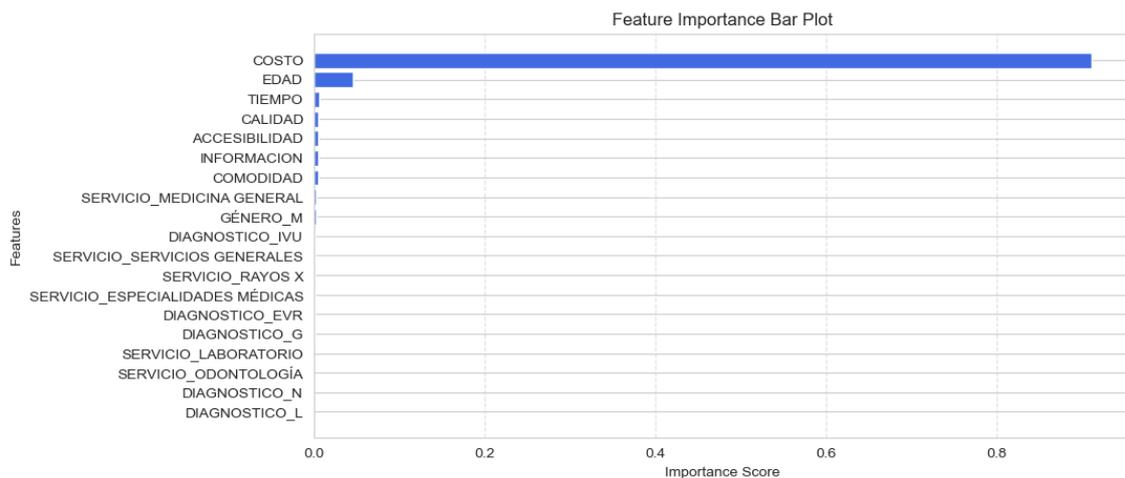


Figura 27. Importancia de las variables en la predicción del beneficio neto

El análisis de la importancia de variables dentro del modelo de *machine learning* permitió identificar cuáles son los factores que tienen un mayor impacto en la predicción del beneficio neto en el Centro Médico MEDISEG. Como se observa en la Figura 27, la variable Costo es la que tiene mayor relevancia dentro del modelo predictivo, con un puntaje de importancia significativamente más alto en comparación con el resto de las variables.

Otras variables como Edad, Tiempo y Calidad también presentan cierta influencia en la predicción, aunque en menor medida. Por otro lado, factores como la accesibilidad, la información del paciente y el tipo de diagnóstico presentan una contribución marginal en la estimación del beneficio neto, lo que sugiere que su impacto es menos significativo para la predicción del modelo.

Este análisis confirma que los costos operativos son un determinante clave en la estimación del beneficio neto, lo que sugiere que las estrategias de optimización financiera deben centrarse en el control y la gestión eficiente de los costos. Asimismo, el impacto de la edad y la calidad del servicio indica que ciertos grupos demográficos y la percepción del servicio pueden afectar la rentabilidad del centro médico. Por tanto, se respalda la robustez del modelo y su aplicabilidad en la optimización de la gestión financiera del centro médico.

## Capítulo cuarto

### 1. Validación del modelo

#### 1.1. Validación del modelo predictivo con datos externos

Con la finalidad de realizar la validación del modelo predictivo, de tal manera que permita evaluar su capacidad de generalización y robustez, se aplicó el modelo de predicción en los datos facilitados por la empresa PHYSIS Rehabilitación Física y Especialidades Médicas S.A.S. cuyo objeto según el Registro Único de Contribuyentes registra “CONSULTA Y TRATAMIENTO POR MÉDICOS GENERALES Y ESPECIALISTAS. ESTAS ACTIVIDADES PUEDEN REALIZARSE EN CONSULTORIOS PRIVADOS, EN CONSULTORIOS COLECTIVOS, EN CLÍNICAS AMBULATORIAS, EN CLÍNICAS ANEXAS A EMPRESAS, ESCUELAS, RESIDENCIAS DE ANCIANOS U ORGANIZACIONES SINDICALES O FRATERNALES Y EN LOS PROPIOS DOMICILIOS DE LOS PACIENTES.”

Con el propósito de interpretar la evolución del beneficio neto desde una perspectiva estratégica, se utilizó como fuente de validación externa la información financiera auditada del año 2024 de la empresa PHYSIS Rehabilitación Física y Especialidades Médicas S.A.S., cuyos datos permiten contrastar el modelo desarrollado originalmente para MEDISEG con una institución de características similares en el sector salud privado ecuatoriano.

#### 1.2. Ejecución del modelo al caso PHYSIS

##### 1.2.1 Preparación de los datos

Los estados financieros de PHYSIS indican una estructura operativa comparable a la de MEDISEG, con ingresos por actividades ordinarias, gastos de ventas, gastos administrativos y utilidad neta reportada. En la Tabla 13, se establecieron las variables clave para la aplicación del modelo presentando el detalle de las características que componen a cada una de ellas, considerando principalmente el tipo de dato, escala y valores.

Tabla 13  
Detalle de variables Physys

Variable	Tipo de Dato	Escala/ Valores	Descripción
Ingreso operacional	Numérico	USD (continua)	Ingresos generados por la prestación de servicios médicos.
Gasto venta	Numérico	USD (continua)	Costos relacionados con la operación comercial y de servicios del centro.
Gasto administrativo	Numérico	USD (continua)	Costos relacionados con la gestión administrativa y soporte del centro.
Gasto financiero	Numérico	USD (continua)	Costos relacionados a obligaciones financieras
Depreciación	Numérico	USD (continua)	Pérdida contable por desgaste o uso de los activos fijos durante el ejercicio.
Total activos	Numérico	USD (continua)	Valor total de los activos al 31 de diciembre de 2024.

Fuente: Elaboración propia con base en información de Physys.

Con base en estos datos, se generó un dataset externo que se adaptó al formato del modelo desarrollado para MEDISEG, empleando las mismas técnicas de pre-procesamiento (escalado, limpieza, normalización y codificación) según las variables definidas.

### 1.2.2 Prueba del modelo *random forest*

Se procedió a aplicar el modelo entrenado con datos de MEDISEG sobre el conjunto externo de PHYSIS. La métrica  $R^2$  obtenida fue de 0.71, confirmando que el modelo conserva una alta capacidad explicativa sobre el beneficio neto, incluso con datos de otra institución. El MSE observado fue de 429.57, levemente superior al valor en MEDISEG (353.38), pero dentro de un rango aceptable que demuestra buena capacidad de generalización.

Esto permite la validación del modelo de aprendizaje automático con predictores mixtos y la posibilidad de adaptación para otras clínicas con estructuras financieras similares, incrementando así su aplicabilidad práctica en el sector de salud.

Al igual que en el caso de MEDISEG, se analizó la importancia relativa de las variables empleadas en el modelo de *machine learning* aplicado al caso PHYSIS, de esta forma, en la Figura 28 se presenta un gráfico de barras que destaca el peso explicativo de cada una de las variables predictoras incluidas en el modelo, permitiendo una mejor interpretación de los factores clave que influyen en la rentabilidad de la institución.

Para esta evaluación se utilizó como criterio técnico el índice de impureza de Gini, métrica propia de los algoritmos de árboles como *Random Forest*. Esto permitió estimar la contribución de cada variable a la reducción del error en cada división del árbol de decisión. Así, mientras mayor sea la reducción del error atribuible a una variable a lo largo de los nodos, mayor será su importancia dentro del modelo predictivo. El cálculo de importancia de las variables se basa, por tanto, en su impacto acumulado sobre la disminución del error en todas las ramas del conjunto de árboles que conforman el modelo.

Este enfoque permitió identificar, cuáles son los factores que aportan significativamente a la mejora de la precisión en la predicción del beneficio neto del centro médico.

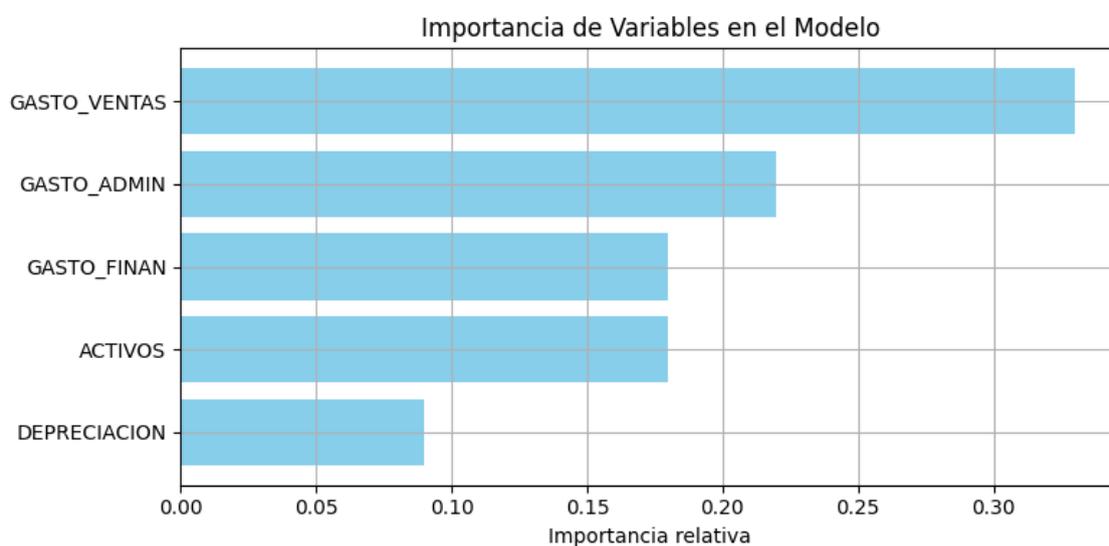


Figura 28. Importancia de las variables en la predicción del beneficio neto caso Physys

Los resultados muestran que las variables Gasto de Venta y Gasto Administrativo tienen un peso explicativo destacado en el modelo, representando conjuntamente más del 50% de la importancia relativa total. Esto confirma que la eficiencia en la gestión de los costos operativos y administrativos es un determinante directo del beneficio neto.

En segundo nivel de importancia se encuentra el Gasto Financiero, lo que evidencia que las obligaciones relacionadas con intereses u otros compromisos financieros afectan significativamente la rentabilidad. La Depreciación, por su parte, aporta valor al explicar parte de la pérdida contable asociada a la utilización de activos fijos, lo cual tiene impacto fiscal y contable relevante.

La variable Total de Activos presenta una influencia importante en el modelo, ya que su inclusión permitió capturar la escala operativa del centro médico, es decir, la capacidad instalada en términos de infraestructura, equipamiento y recursos disponibles. Esta variable actúa como un indicador estructural que muestra la posibilidad de generación de ingresos, así como la eficiencia de los recursos que han sido invertidos.

Dentro del enfoque de predicción, el modelo indica que mientras mayor nivel de activos existan, esto tiende a generar mayores posibilidades de obtención de beneficios, siempre que estos sean gestionados de manera eficiente. Asimismo, esto permite tener una amplia visión respecto a la renovación de activos fijos, pues pueden impactar indirectamente en el beneficio neto a través de mayores niveles de producción o por el efecto contable de la depreciación asociada.

Este análisis proporciona información valiosa no solo para la comprensión técnica del modelo, sino también para futuras decisiones estratégicas, tales como la priorización de inversiones, la evaluación del endeudamiento institucional, o la optimización del gasto. Por tanto, la evaluación de la importancia de variables en el caso PHYSIS respalda tanto la robustez del modelo como su aplicabilidad práctica en la mejora de la gestión financiera de centros médicos.

### **1.2.3 Análisis del beneficio neto**

Desde una perspectiva económica, según el marco conceptual previamente examinado, el beneficio neto puede entenderse como el resultado de una eficiente conversión de ingresos en utilidades, después de cubrir los costos operativos. El modelo presentado muestra que la variable Gasto Ventas es el principal predictor del beneficio, lo cual refleja que la gestión eficiente de los recursos operativos tiene un peso determinante en la rentabilidad clínica.

En términos empresariales, esto implica que la sostenibilidad financiera no depende exclusivamente del volumen de ingresos, sino del control estratégico de los gastos variables y fijos. Por ejemplo, en PHYSIS, a pesar de tener ingresos de más de

\$76,000, los gastos representan más del 98% del total, generando una utilidad neta muy reducida de \$576 (margen inferior al 1%).

#### **1.2.4 Análisis de la depreciación y su rol en la rentabilidad**

El análisis contable de PHYSIS indica una depreciación acumulada de USD 7,539 para el año 2024 (con base en activos como vehículos, equipos médicos y computación). Este componente tiene múltiples efectos:

- Contablemente, reduce la utilidad operativa sin implicar salida de efectivo, lo que mejora la liquidez efectiva.
- Fiscalmente, la depreciación es un gasto deducible que reduce la base imponible, optimizando la carga tributaria.

Esto muestra de que manera la gestión eficiente de activos fijos y su planificación en el tiempo (compras, renovación tecnológica) puede mejorar el beneficio neto y al mismo tiempo crear ventajas fiscales sin comprometer el flujo de caja.

#### **1.2.5 Consideraciones de planificación fiscal**

El impuesto a la renta reportado por PHYSIS fue de USD 163 (aproximadamente el 22% de la utilidad contable antes de impuestos). Existen tres aspectos clave desde una perspectiva de gestión fiscal:

- La provisión adecuada del impuesto contribuye a minimizar contingencias legales y refleja transparencia contable.
- La gestión de gastos deducibles como depreciación, contribuye a reducir el impuesto efectivo a pagar.
- Al estar bajo el régimen de Sociedad Anónima Simplificada, Physys podría beneficiarse de reducciones porcentuales sobre la tarifa del impuesto, lo cual incrementaría la utilidad neta disponible.

Estas acciones se vinculan directamente con decisiones administrativas estratégicas que impactan la rentabilidad.

#### **1.2.6 Diagnóstico estratégico a partir del modelo**

Basado en el análisis financiero y la importancia de variables del modelo, se desprenden las siguientes observaciones estratégicas:

- Alta sensibilidad a los costos de ventas: de acuerdo a los resultados observados, se entiende que, las empresas del sector salud que no optimizan insumos médicos, subcontrataciones o tiempos de atención pueden presentar una tendencia a la baja en la competitividad.
- Gastos administrativos significativos: las estructuras burocráticas grandes, son poco eficientes y afectan el margen de utilidad como se evidenció en PHYSYS con un gasto administrativo del 28% sobre ingresos.
- Poco apalancamiento operativo: la baja diferencia entre ingresos y egresos indica que pequeñas variaciones en los costos pueden transformar una utilidad marginal en pérdida, mostrando vulnerabilidad financiera.

En el contexto del caso PHYSYS, se observó que el modelo identifica y cuantifica que las decisiones de gestión en relación a la estructura de costos (ventas, administración, financieros) impactan directamente sobre el beneficio neto, lo cual permite desarrollar alertas tempranas de rentabilidad presentando un modelo replicable que puede ser utilizado como herramienta para mejorar la rentabilidad y sostenibilidad financiera de instituciones de salud en contextos similares.

En este sentido, se entiende según los resultados observados que el beneficio neto no solo depende del nivel de ingreso, sino sobre todo de cómo se gestionan los egresos, la estructura financiera y los activos disponibles.

## Conclusiones

Se identificaron las variables claves que inciden en la estimación del beneficio neto, se destacaron las siguientes variables: el costo operativo, el tiempo de servicio, la edad de los pacientes y la calidad percibida de la atención. Las variables se seleccionaron a través de un análisis exploratorio y pruebas estadísticas, lo que aseguró su sustento metodológico para la predicción financiera del centro médico.

Se establecieron criterios para la limpieza y para la validación de los datos, incluyendo la detección y eliminación de valores atípicos, la normalización de variables luego de la imputación de datos faltantes. El proceso ofrece garantías en la consistencia y confiabilidad del conjunto de datos previo al uso del modelo *machine learning*.

Posterior a la depuración y análisis de los datos, se determinó incluir el conjunto final de las variables con mayor impacto para la estimación del beneficio neto. Por tanto, se originó un dataset optimizado, que almacena dentro de un archivo estructurado, brindando representatividad de los datos utilizados en el modelamiento.

Se evaluó distintos algoritmos de *machine learning* como los más adecuados para integrar predictores mixtos: *Random Forest*, *Gradient Boosting* y Regresión Lineal Múltiple. La elección de estos algoritmos se basó en su capacidad para el manejo de datos categóricos y numéricos, brindando robustez y precisión en la predicción del beneficio neto.

La validación del modelo evidenció que el algoritmo *Random Forest* otorgó mayor exactitud con un coeficiente de determinación ( $R^2$ ) de 0,76. Superando a la regresión lineal. Asimismo, se comprobó que el costo operativo es el más influyente predictor para la estimación del beneficio neto, que resalta la importancia de crear estrategias de control de costos que permitan mejorar la rentabilidad del Centro Médico MEDISEG. Los hallazgos encontrados respaldan la efectividad del modelo que se ha desarrollado y propone la posibilidad de su implementación en la toma de decisiones estratégicas en la gestión del centro médico.

Finalmente, la validación externa con datos de PHYSIS demostró la robustez y aplicabilidad del modelo predictivo desarrollado. A su vez, el análisis económico permitió conectar los resultados cuantitativos con fundamentos estratégicos, evidenciando que el beneficio neto en clínicas privadas depende significativamente de la eficiencia operativa.

## Recomendaciones

Se recomienda implementar un monitoreo continuo en tiempo real de las variables determinadas a través de herramientas de análisis de datos, que permita descubrir patrones y estrategias financieras que se ajusten para mejorar la rentabilidad del centro médico.

Se recomienda conservar estrictos protocolos de calidad de datos y a la vez incluir formación al personal encargado del manejo de la base de datos en la validación y filtrado de la información, garantizando la confiabilidad de la base de datos establecida.

Se sugiere que se realice actualizaciones periódicas del conjunto de datos establecido en el modelamiento, permitiendo que se incluyan nuevas variables de relevancia y la optimización del modelo predictivo.

Se considera adecuado realizar en el campo del *machine learning* pruebas con otros algoritmos emergentes adicionales, tales como modelos basados en redes neuronales, con el fin de evaluar mejoras en la precisión y capacidad predictiva del modelo, especialmente al utilizar predictores mixtos.

## Lista de referencias

- Alvarez, A. 2023. “Propuesta de implementación de algoritmos de *machine learning* para realizar una adecuada gestión del inventario en una pyme colombiana del sector textil y de moda”. Tesis de maestría, Universidad EIA. <https://repository.eia.edu.co/server/api/core/bitstreams/09793abb-ea82-415d-bb29-4c8ecccd5dff/content>.
- Amat, J. 2021. “Redes neuronales con Python”. *Ciencia Datos*. <https://cienciadedatos.net/documentos/py35-redes-neuronales-python>.
- Andrade, V. 2022. “Políticas de gestión financiera para el mejoramiento de los índices de morosidad de la Cooperativa de Ahorro y Crédito Cañar Ltda”. Tesis de maestría, Escuela Superior Politécnica de Chimborazo. <http://dspace.esepoch.edu.ec/handle/123456789/15711>.
- Blanco, S., y J. Rubriche. 2024. “Construcción de Modelos Predictivos para Clasificar Transacciones Legítimas o Fraudulentas Utilizando Algoritmos de Aprendizaje Automático”. Tesis doctoral, Universidad Santo Tomas. <https://repository.usta.edu.co/handle/11634/54007>.
- Breiman, L. 2009. “Random forests”. *Machine learning* 45.
- Calahorrano, Gabriela Alejandra, Francisco Andrés Chacón Guerrero, y Ana Belén Tulcanaza Prieto. 2021. “Indicadores financieros y rentabilidad en bancos grandes y medianos ecuatorianos, periodo: 2016-2019”. *INNOVA Research Journal* 6 (2): 225-39. <https://doi.org/10.33890/innova.v6.n2.2021.1700>.
- Collazos-Castillo, Karen Yulieth. 2023. “08. Aplicaciones de *Machine Learning* para la mejora de la atención médica en Colombia”. Tesis doctoral, Colombia: Universidad de Manizales. [https://ridum.umanizales.edu.co/bitstream/handle/20.500.12746/6834/CIF%202023-1%20-%20Res%C3%BAmenes%20ejecutivos\\_RIDUM.pdf?sequence=1&isAllowed=y#page=81](https://ridum.umanizales.edu.co/bitstream/handle/20.500.12746/6834/CIF%202023-1%20-%20Res%C3%BAmenes%20ejecutivos_RIDUM.pdf?sequence=1&isAllowed=y#page=81).
- De Alba, G. 2020. “¿Cuál es la diferencia entre los métodos de bagging y los de boosting?”. *Machine Learning*. 24 de febrero. <https://machinelearningparatodos.com/cual-es-la-diferencia-entre-los-metodos-de-bagging-y-los-de-boosting/>.

- Franco, Y. 2023. "Mamachine learning aplicado a dificultades financieras y quiebra empresarial: una revisión de literatura". *Avances en Investigación científica* 2 (7): 277. <http://www.doi.org/10.47666/avances.inv.4>.
- Géron, a. 2022. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc. <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>.
- Gil-Martínez, Cristina. 2018. "Árboles de decisión y métodos de ensemble". *RPubs*. Junio. [https://rpubs.com/Cristina\\_Gil/arboles\\_ensemble](https://rpubs.com/Cristina_Gil/arboles_ensemble).
- Glaser, J., A. Benjamin, R. Farhoodi, y K. Kording. 2019. "The roles of supervised machine learning in systems neuroscience". *Progress in Neurobiology* 175: 126-37. <https://www.sciencedirect.com/science/article/abs/pii/S0301008218300856>.
- Guallasamín, Wilson Rolando. 2019. "Gestión de riesgo de liquidez: modelo para determinar posibles salidas de dinero en base a indicadores micro económicos y macro económicos de las instituciones financieras privadas del ecuador". Tesis doctoral, Escuela Politécnica Nacional, Quito. <https://bibdigital.epn.edu.ec/handle/15000/20828>.
- Harrington, P. 2012. "Machine Learning in Action". *Manning Publications* 7 (5). <https://thuviensou.dau.edu.vn:88/bitstream/DHKTDN/6832/1/6185.Machine%20Learning%20in%20Action.pdf>.
- Hastie, T. 2009. "The elements of statistical learning: data mining, inference, and prediction". *Taylor & Francis* 7 (2). <https://doi.org/10.1198/jasa.2004.s339>.
- Hastie, Trevor, Robert Tibshirani, & Martin Wainwright. 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Cambridge University Press: Engineering & Technology, Mathematics & Statistics. <https://www.taylorfrancis.com/books/mono/10.1201/b18401/statistical-learning-sparsity-trevor-hastie-martin-wainwright-robert-tibshirani>.
- Hernández, F. 2021. *Modelos Predictivos*. [https://fhernanb.github.io/libro\\_mod\\_pred/](https://fhernanb.github.io/libro_mod_pred/).
- IBM. 2018. "Guía de CRISP-DM de IBM SPSS Modeler". *IBM*. [https://www.ibm.com/docs/es/SS3RA7\\_18.4.0/pdf/ModelerCRISPDM.pdf](https://www.ibm.com/docs/es/SS3RA7_18.4.0/pdf/ModelerCRISPDM.pdf).
- IBM, Security. 2022. "¿Qué es el aprendizaje supervisado?". *IBM*. <https://www.ibm.com/mx-es/topics/supervised-learning>.
- Lopez, A., y R. Villanueva. 2022. "Sistema para la automatización de procesos hospitalarios de control para pacientes para COVID-19 usando machine learning

- para el Centro de Salud San Fernando”. Tesis doctoral, Universidad Cesar Vallejo, Perú. <https://repositorio.ucv.edu.pe/handle/20.500.12692/121848>.
- Tizne Ondiviela, David. 2024. “Machine Learning: Regresión logística multiclase y aplicación para la clasificación de terrenos”. Tesis de pregrado, Universidad de Zaragoza. <https://zaguan.unizar.es/record/149521/files/TAZ-TFG-2024-3404.pdf>.
- López, S., S. Magdaleno, y G. Sánchez. 2023. “Modelo con enfoque en el Technology Enhanced Learning para la exteriorización del conocimiento para la efectividad de la enseñanza”. *IE Revista de Investigación Educativa de la REDIECH* 1 (4): 1783. [https://mail.rediech.org/ojs/2017/index.php/ie\\_rie\\_rediech/article/view/1783](https://mail.rediech.org/ojs/2017/index.php/ie_rie_rediech/article/view/1783).
- Luyo Zavala, Pamela Silvana, y Naveda Dominguez, Patricia. 2022. “Herramientas de gestión en estimación y el resultado económico de la empresa del sector salud Anglolab S.A.”. Tesis doctoral, Universidad Nacional del Callao, Perú. <https://repositorioslatinoamericanos.uchile.cl/handle/2250/9468495>.
- Marshall Boehmwald, Fernando. 2022. “Diseño de un modelo de generación de datos sintéticos para la aplicación de modelos de machine learning en proyectos interdisciplinarios asociados a salud”. Tesis doctoral, Universidad de Chile. <https://repositorio.uchile.cl/xmlui/bitstream/handle/2250/188774/Diseno-de-un-modelo-de-generacion-de-datos-sinteticos-para-la-aplicacion-de-modelos-de-machine-learning.pdf?sequence=1>.
- Montgomery, D., y G. Runger. 2010. “Probabilidad y Estadística aplicada a la Ingeniería”. Tesis doctoral, México: Universidad Autónoma de Querétaro. [https://www.uaq.mx/informatica/contenidosExtenso/1203\\_Probabilidad%20y%20Estad%C3%ADstica.pdf](https://www.uaq.mx/informatica/contenidosExtenso/1203_Probabilidad%20y%20Estad%C3%ADstica.pdf).
- Neter, J., M. Kutner, C. Nachtsheim, W. Wasserman, H. Schreuder, & F. Collins. 2004. “Overall Considerations in Designing and Analyzing Forestry Experiments”. *British Journal of Management* 7 (2). <https://doi.org/10.1111/2041-210X.12126>.
- Ossa, W., y V. Jaramillo. 2021. “Machine Learning para la estimación del riesgo de crédito en una cartera de consumo”. Tesis doctoral, Universidad EAFIT, Medellín. <https://repository.eafit.edu.co/server/api/core/bitstreams/324d9a52-33ab-4255-bfaf-132f206c753e/content>.
- Pineda, J. 2022. “Modelos predictivos en salud basados en aprendizaje de maquina (machine learning) Predictive models in health based on machine learning”.

- Revista Médica Clínica Las Condes* 33 (6): 583-90.  
<https://doi.org/10.1016/j.rmclc.2022.11.002>.
- Quiroz, N., M. Posadas, E. Rossi, D. Giunta, y M. Risk. 2022. “Aprendizaje automático aplicado en área de la salud: Parte 2”. *Revista del Hospital Italiano de Buenos Aires* 7 (2). <https://ri.conicet.gov.ar/handle/11336/200844>.
- Ramírez, H., J. Montenegro, y N. Estupiñá. 2021. “Análisis del impacto económico y social por la pandemia del COVID 19 y su influencia en los emprendimientos ecuatorianos”. *Dominio de las Ciencias* 7 (2): 1320-42.  
<http://dx.doi.org/10.23857/dc.v7i2.1883>.
- Rouhiainen, L. 2020. *Inteligencia artificial*. Madrid: Alienta Editorial.  
[https://planetadelibrosec0.cdnstatics.com/libros\\_contenido\\_extra/40/39308\\_Inteligencia\\_artificial.pdf](https://planetadelibrosec0.cdnstatics.com/libros_contenido_extra/40/39308_Inteligencia_artificial.pdf).
- Saini, A. 2021. *Master the AdaBoost Algorithm: Guide to Implementing & Understanding AdaBoost*.
- Schapire, R., y Y. Freund. 2013. “Boosting: Foundations and algorithms. *Kybernetes*”.  
*Discovery Journals* 42 (1): 164-6. <https://doi.org/10.1108/03684921311295547>.
- Theobald, O. 2017. *Machine learning for absolute beginners: a plain English introduction*, vol. 157. London: Scatterplot press.  
<https://nativebookpdf.com/download/4939631-Machine%20Learning%20For%20Absolute%20Beginners%20By%20Oliver%20Theobald>.
- wyj. 2023. Linear Regression. 2023년 3월 31일. Corea: velog.io.  
<https://velog.io/@wyjung0731/Linear-Regression>.
- Zaldívar, E., V. Fuentes, E. Cardeñosa, D. Lozada, y H. Cantero. 2011. *Análisis de la rentabilidad económica: tecnología propuesta para incrementar la eficiencia empresarial*. ALMA MATER. Cuba: Editorial Universitaria.
- Zhang, F., y L. O'Donnell. 2020. “Support vector regression. In *Machine learning*”.  
*Harvard Medical School, Boston, MA, United States* 7 (2).  
[https://www.researchgate.net/publication/338315687\\_Support\\_vector\\_regression](https://www.researchgate.net/publication/338315687_Support_vector_regression).

## Anexos

### Anexo 1: Estado financieros de la empresa PHYSIS Rehabilitación Física y Especialidades Médicas S.A.S.

**ESTADO DE SITUACIÓN FINANCIERA**  
**POR EL AÑO TERMINADO AL 31 DE DICIEMBRE DE 2024**  
 (expresado en dólares americanos)

<b>ACTIVOS</b>	<b>Notas</b>	<b>2024</b>	<b>2023</b>
<b>Activo Corriente</b>			
Efectivo y Equivalentes a Efectivo	5	23,561	6,237
Activos por impuestos corrientes	6	2,745	2,289
<b>Total activo corriente</b>		<b>26,306</b>	<b>8,525</b>
<b>Activo no Corriente</b>			
<b>Propiedad planta y equipo</b>			
Vehículos	7	18,849	18,849
(-) Depreciación acumulada propiedad planta y equipo	-	7,539	3,770
<b>Total activo no corriente</b>		<b>11,309</b>	<b>15,079</b>
<b>TOTAL ACTIVOS</b>		<b>37,615</b>	<b>23,604</b>
<b>PASIVO Y PATRIMONIO</b>			
<b>Pasivo Corriente</b>			
Proveedores por pagar	8	14,555	-
Pasivos por impuestos corrientes	9	163	1,283
<b>Total pasivo corriente</b>		<b>14,718</b>	<b>1,283</b>
<b>TOTAL PASIVOS</b>		<b>14,718</b>	<b>1,283</b>
<b>PATRIMONIO</b>			
Capital Social suscrito o pagado	10	1,000	1,000
Ganancias acumuladas		21,321	-
Utilidad del ejercicio		576	21,321
<b>TOTAL PATRIMONIO</b>		<b>22,898</b>	<b>22,321</b>
<b>TOTAL PASIVO Y PATRIMONIO</b>		<b>37,615</b>	<b>23,604</b>



Firmado electrónicamente por:  
**KATHIA**  
**ELIZABETH**  
**CIFUENTES**  
**BENAVIDES**

Cifuentes Benavides Kathia Elizabeth  
**Gerente General**  
 CI. 1721494043

Ing. MaríaJose Luna Cueva  
**Contador General**  
 RUC: 1722419536001

**ESTADO DE RESULTADOS INTEGRAL**  
**POR EL AÑO TERMINADO AL 31 DE DICIEMBRE DE 2024**  
 (expresado en dólares americanos)

		<u>2024</u>	<u>2023</u>
<b>INGRESOS DE ACTIVIDADES ORDINARIAS</b>	<b>Notas</b>		
Ingresos operacionales	11	76,113	113,434
<b>Utilidad Bruta</b>		<b>76,113</b>	<b>113,434</b>
<b>GASTOS OPERACIONALES</b>			
Gastos de ventas	12	- 53,432	- 62,843
Gastos administrativos	12	- 21,653	- 27,652
<b>Utilidad/ Pérdida operacional</b>		<b>1,029</b>	<b>22,938</b>
<b>INGRESOS/ GASTOS NO OPERACIONALES</b>			
Otros ingresos	12	5	-
Comisiones Financieras		- 294	- 335
<b>Utilidad antes de impuestos</b>		<b>739</b>	<b>22,604</b>
Impuesto a la renta	13	- 163	1,283
<b>UTILIDAD/ PÉRDIDA DEL EJERCICIO</b>		<b>576</b>	<b>21,321</b>



Firmado electrónicamente por:

**KATHIA  
 ELIZABETH  
 CIFUENTES  
 BENAVIDES**

Cifuentes Benavides Kathia Elizabeth

**Gerente General**

CI. 1721494043

Ing. MaríaJose Luna Cueva

**Contador General**

RUC:1722419536001

## PHYSIS REHABILITACIÓN FÍSICA Y ESPECIALIDADES MÉDICAS S.A.S

## ESTADOS DE FLUJO DE EFECTIVO

POR EL AÑO TERMINADO AL 31 DE DICIEMBRE DEL 2024

	2024	2023
<b>Flujos de efectivo procedentes de (utilizados en) actividades de operación</b>		
Cobros procedentes de las ventas de bienes y prestación de servicios	74,537	111,146
Otros cobros por actividades de operación	5	
Pagos a proveedores por el suministro de bienes y servicios	- 56,760	- 86,726
Otros pagos por actividades de operación	- 294	- 335
Impuestos a las ganancias pagados	- 163	
<b>Efectivo neto utilizado en actividades de operación</b>	<b>17,325</b>	<b>24,085</b>
<b>Flujos de efectivo de las actividades de inversión</b>		
Adquisiciones propiedad planta y equipo	-	18,849
<b>Efectivo neto utilizado en actividades de inversión</b>	<b>-</b>	<b>18,849</b>
<b>Flujos de efectivo procedentes de (utilizados en) actividades de financiación</b>		
Financiación en entidades financieras	-	-
<b>Efectivo neto utilizado en actividades de financiación</b>	<b>-</b>	<b>-</b>
<b>Incremento neto de efectivo y equivalentes al efectivo</b>	<b>17,325</b>	<b>5,237</b>
Efectivo y equivalentes al efectivo al inicio del periodo	6,237	1,000
<b>Efectivo y equivalentes al efectivo al final del periodo</b>	<b>23,561</b>	<b>6,237</b>
<b>Conciliación entre la ganancia (pérdida) neta y los flujos de operación</b>		
Ganancia (pérdida) antes de 15% a trabajadores e impuesto a la renta	739	22,604
Ajustes por gasto depreciación y amortización	3,770	3,770
<b>Cambios en activos:</b>	<b>- 1,739</b>	
Incremento (disminución) en cuentas por cobrar clientes	-	-
Incremento (disminución) en anticipos de proveedores	-	-
Incremento (disminución) en inventarios	-	-
Incremento (disminución) en otros activos	- 1,739	-
<b>Cambios en pasivos:</b>	<b>14,555</b>	
(Incremento) disminución en cuentas por pagar comerciales	14,555	-
Incremento (disminución) en otras cuentas por pagar	-	-
Incremento (disminución) en beneficios empleados	-	-
Incremento (disminución) en anticipos de clientes	-	-
(Incremento) disminución en otros pasivos	-	- 2,289
<b>Efectivo neto provisto en las actividades de operación</b>	<b>17,325</b>	<b>24,085</b>



Firmado electrónicamente por:

KATHIA  
ELIZABETH  
CIFUENTES  
BENAVIDES

Cifuentes Benavides Kathia Elizabeth

Gerente General

CI. 1721494043

Ing. MaríaJose Luna Cueva

Contador General

RUC: 1722419536001

## **Anexo 2: Acta de la Asamblea General Extraordinaria de Accionistas 2024**

### **ACTA DE LA JUNTA ASAMBLEA GENERAL EXTRAORDINARIA DE ACCIONISTAS, RESOLUCIONES ADOPTADAS POR EL ACCIONISTA ÚNICO DE LA COMPAÑÍA DE PHYSIS REHABILITACIÓN FÍSICA Y ESPECIALIDADES MÉDICAS S.A.S. DEL EJERCICIO ECONOMICO 2024**

En la parroquia de Quito, Cantón Rumiñahui, a los 18 días del mes de abril de 2024, a las 10H30 am, en las oficinas de la compañía, ubicadas en la calle Guayaquil y Quiroga No. 216 en el Cantón Rumiñahui de la compañía **PHYSIS REHABILITACIÓN FÍSICA Y ESPECIALIDADES MÉDICAS S.A.S.**

La señorita Cifuentes Benavides Kathia Elizabeth titular del 100% del capital social y por lo tanto accionista único de la compañía **PHYSIS REHABILITACIÓN FÍSICA Y ESPECIALIDADES MÉDICAS S.A.S.** actúa personalmente y ejerce las funciones propias de la asamblea general extraordinaria de accionistas, en virtud de lo dispuesto en el inciso noveno de artículo 236 de la Ley de Compañías y artículo 21 del Reglamento sobre Juntas Generales o Asamblea General de Socios y Accionistas de las compañías de Responsabilidad Limitada, Anónimas, en Comandita por Acciones, de Economía Mixta y Sociedades por Acciones Simplificadas; para resolver los asuntos contenidos en el siguiente orden del día:

1. Conocer y resolver sobre el informe presentado por el Administrador de la Compañía, relativo a su gestión durante el ejercicio económico correspondiente al 2024
2. Presentación y aprobación de Balance General y el Estado de Pérdidas y Ganancias por ejercicio 2024, y sus respectivos anexos
3. Destino utilidades del ejercicio fiscal 2024.

Expuestos los puntos del orden del día, el accionista único adopta las siguientes resoluciones:

1. Conocer y resolver sobre el informe presentado por el Administrador de la Compañía, relativo a su gestión durante el ejercicio económico correspondiente al 2024

La Gerente General Cifuentes Benavides Kathia Elizabeth da lectura a dicho informe.

2. Presentación y aprobación de Balance General y el Estado de Pérdidas y Ganancias por ejercicio 2024, y sus respectivos anexos

Se da lectura a los Estados Financieros por parte de la Señorita Maria Jose Luna, una vez concluida la lectura y al resolver inquietudes por parte del Sr. Gerente General, y se aprueban los estados financieros del ejercicio fiscal 2024.

3. Destino utilidades del ejercicio fiscal 2024.

El destino de las utilidades por el valor de \$576.26 del ejercicio fiscal 2024, deben ser asignadas a la cuenta de utilidades acumuladas.

En vista de que se han tratado todos los puntos, la Gerente General Cifuentes Benavides Kathia Elizabeth, solicita un receso, para redacta el acta.

Siendo las 11:00 am se levanta la Asamblea General Extraordinaria de accionistas.

#### **4. Resoluciones**

La accionista única Cifuentes Benavides Kathia Elizabeth, declara bajo juramento que la información proporcionada y la documentación de soporte presentada a la Superintendencia de Compañías Valores y Seguros, es veráz y autentica.

Como constancia de las resoluciones adoptadas, el accionista la señorita, suscribe la presente acta.

Firmado electrónicamente por:

**KATHIA ELIZABETH CIFUENTES BENAVIDES**



.....  
Cifuentes Benavides Kathia Elizabeth

**Gerente General**