



Autónoma
Universidad Autónoma del Perú

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

TESIS

IMPLEMENTACIÓN DE UN MODELO BASADO EN ALGORITMOS DE MACHINE
LEARNING PARA EL ANÁLISIS PREDICTIVO DE DIAGNÓSTICOS CLÍNICOS EN
REDLAB PERÚ S.A.C.

PARA OBTENER EL TÍTULO DE

INGENIERO DE SISTEMAS

AUTORES

ANDERSON ALFREDO FUENTES PINEDO

ORCID: 0000-0002-3897-8464

CARLOS ENMANUEL NAVARRO CARAZA

ORCID: 0000-0002-4863-9685

ASESOR

DR. JULIO ELVIS VALERO CAJAHUANCA

ORCID: 0000-0002-8522-6249

LÍNEA DE INVESTIGACIÓN DE PROGRAMA

DESARROLLO E IMPLEMENTACIÓN DE SISTEMAS INTELIGENTES

LIMA, PERÚ, DICIEMBRE DE 2023



CC BY

<https://creativecommons.org/licenses/by/4.0/>

Esta licencia permite a otros distribuir, mezclar, ajustar y construir a partir de su obra, incluso con fines comerciales, siempre que le sea reconocida la autoría de la creación original. Esta es la licencia más servicial de las ofrecidas. Recomendada para una máxima difusión y utilización de los materiales sujetos a la licencia.

Referencia bibliográfica

Fuentes Pinedo, A., A. & Navarro Caraza, C., E. (2024). *Implementación de un modelo basado en algoritmos de Machine Learning para el análisis predictivo de diagnósticos clínicos en REDLAB Perú S.A.C.* [Tesis de pregrado, Universidad Autónoma del Perú]. Repositorio de la Universidad

HOJA DE METADATOS

Datos del autor	
Nombres y apellidos	Anderson Alfredo Fuentes Pinedo
Tipo de documento de identidad	DNI
Número de documento de identidad	75120338
URL de ORCID	https://orcid.org/0000-0002-3897-8464
Datos del autor	
Nombres y apellidos	Carlos Enmanuel Navarro Caraza
Tipo de documento de identidad	DNI
Número de documento de identidad	77352026
URL de ORCID	https://orcid.org/0000-0002-4863-9685
Datos del asesor	
Nombres y apellidos	Julio Elvis Valero Cajahuanca
Tipo de documento de identidad	DNI
Número de documento de identidad	80543932
URL de ORCID	https://orcid.org/0000-0002-8522-6249
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Wilyam David Torres Meza
Tipo de documento	DNI
Número de documento de identidad	09435088
Secretario del jurado	
Nombres y apellidos	Ana Julieta González García
Tipo de documento	CEX
Número de documento de identidad	003020400
Vocal del jurado	
Nombres y apellidos	Katherine Delgado Paucar
Tipo de documento	DNI
Número de documento de identidad	70884007
Datos de la investigación	

Título de la investigación	Implementación de un modelo basado en algoritmos de Machine Learning para el análisis predictivo de diagnósticos clínicos en REDLAB Perú S.A.C.
Línea de investigación Institucional	Ciencia, Tecnología e Innovación
Línea de investigación del Programa	Desarrollo e implementación de sistemas inteligentes
Línea de acción RSU	Desarrollo e Innovación Social
URL de disciplinas OCDE	https://purl.org/pe-repo/ocde/ford#2.02.04

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS
ACTA DE SUSTENTACIÓN DE TESIS

En la ciudad de Lima, el jurado de sustentación de tesis conformado por: el Dr. WILYAM DAVID TORRES MEZA como presidente, la Mg. ANA JULIETA GONZÁLEZ GARCÍA como secretario y la Mg. KATHERINE DELGADO PAUCAR como vocal, reunidos en acto público para dictaminar la tesis titulada:

IMPLEMENTACIÓN DE UN MODELO BASADO EN ALGORITMOS DE MACHINE LEARNING PARA EL ANÁLISIS PREDICTIVO DE DIAGNÓSTICOS CLÍNICOS EN REDLAB PERÚ S.A.C.

Presentado por el bachiller:

ANDERSON ALFREDO FUENTES PINEDO

Para obtener el **Título Profesional de Ingeniero de Sistemas**; luego de escuchar la sustentación de la misma y resueltas las preguntas del jurado se procedió a la calificación individual, obteniendo el dictamen de **Aprobado-Muy Bueno** con una calificación de **DIECIOCHO (18)**.

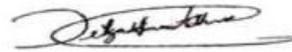
En fe de lo cual firman los miembros del jurado, el 11 de julio del 2024.



PRESIDENTE
Dr. WILYAM DAVID
TORRES MEZA



SECRETARIO
Mg. ANA JULIETA
GONZÁLEZ GARCÍA



VOCAL
Mg. KATHERINE DELGADO
PAUCAR

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS
ACTA DE SUSTENTACIÓN DE TESIS

En la ciudad de Lima, el jurado de sustentación de tesis conformado por: el Dr. WILYAM DAVID TORRES MEZA como presidente, la Mg. ANA JULIETA GONZÁLEZ GARCÍA como secretario y la Mg. KATHERINE DELGADO PAUCAR como vocal, reunidos en acto público para dictaminar la tesis titulada:

IMPLEMENTACIÓN DE UN MODELO BASADO EN ALGORITMOS DE MACHINE LEARNING PARA EL ANÁLISIS PREDICTIVO DE DIAGNÓSTICOS CLÍNICOS EN REDLAB PERÚ S.A.C.

Presentado por el bachiller:

CARLOS ENMANUEL NAVARRO CARAZA

Para obtener el **Título Profesional de Ingeniero de Sistemas**; luego de escuchar la sustentación de la misma y resueltas las preguntas del jurado se procedió a la calificación individual, obteniendo el dictamen de **Aprobado-Muy Bueno** con una calificación de **DIECIOCHO (18)**.

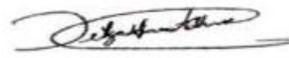
En fe de lo cual firman los miembros del jurado, el 11 de julio del 2024.



PRESIDENTE
Dr. WILYAM DAVID
TORRES MEZA



SECRETARIO
Mg. ANA JULIETA
GONZÁLEZ GARCÍA



VOCAL
Mg. KATHERINE DELGADO
PAUCAR

ACTA DE APROBACIÓN DE ORIGINALIDAD

Yo Julio Elvis Valero Cajahuanca, docente de la Facultad de Ingeniería y Arquitectura de la Escuela Profesional de Ingeniería de Sistemas de la Universidad Autónoma del Perú, en mi condición de asesor de la tesis titulada:

IMPLEMENTACIÓN DE UN MODELO BASADO EN ALGORITMOS DE MACHINE LEARNING PARA EL ANÁLISIS PREDICTIVO DE DIAGNÓSTICOS CLÍNICOS EN REDLAB PERÚ S.A.C.

De los bachilleres Anderson Alfredo Fuentes Pinedo y Carlos Enmanuel Navarro Caraza, certifico que la tesis tiene un índice de similitud de 20% verificable en el reporte de similitud del software Turnitin que se adjunta.

El suscrito revisó y analizó dicho reporte a lo que concluyó que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad Autónoma del Perú.

Lima, 2 de Diciembre de 2024



Julio Elvis Valero Cajahuanca
DNI 80543932

DEDICATORIA

Con gratitud infinita a mi familia, quienes han sido mi apoyo incondicional en cada paso de este camino. Dedico este logro a todas las personas que, con su amor y guía, han formado parte de mi crecimiento y éxito.

Anderson Alfredo Fuentes Pinedo

A mi familia y seres queridos, por su amor, apoyo incondicional y confianza en mí. Este logro no habría sido posible sin ustedes, quienes siempre han sido mi mayor inspiración y fortaleza.

Carlos Enmanuel Navarro Caraza

AGRADECIMIENTOS

Agradecer a todas las personas y organizaciones que, de una u otra forma, han contribuido al desarrollo de esta tesis es un acto de reconocimiento hacia quienes han sido parte fundamental de este proceso. En primer lugar, a Redlab Perú S.A.C., por su confianza, orientación y las oportunidades brindadas que fueron clave para la culminación de este proyecto. A nuestros compañeros de estudios y amigos, quienes, con su apoyo, solidaridad y camaradería en largas jornadas de esfuerzo, hicieron que los momentos difíciles fueran más llevaderos y enriquecedores. Asimismo, extendemos nuestro agradecimiento a nuestros mentores y docentes, cuyas enseñanzas y guía constante fueron una inspiración que nos ayudó a superar cada desafío y a consolidar este trabajo. Finalmente, a todas las personas que, directa o indirectamente, con su motivación, paciencia y confianza, han contribuido a que este logro sea posible, nuestro más sincero reconocimiento, ya que este resultado es reflejo de un esfuerzo compartido.

ÍNDICE

DEDICATORIA	2
AGRADECIMIENTOS	3
LISTA DE TABLAS	5
LISTA DE FIGURAS	7
RESUMEN	10
ABSTRACT	11
CAPÍTULO I: INTRODUCCIÓN	12
CAPÍTULO II: METODOLOGÍA	39
2.1. Tipo y diseño de investigación	40
2.2. Población, muestra y muestreo.....	41
2.3. Hipótesis.....	44
2.4. Variables y operacionalización.....	44
2.5. Técnicas e instrumentos de recolección de datos, validez y confiabilidad..	48
2.6. Procedimientos	51
2.7. Análisis de datos.....	57
2.8. Aspectos éticos.....	58
CAPÍTULO III: RESULTADOS	60
CAPÍTULO IV: DISCUSIÓN	114
CAPÍTULO V: CONCLUSIONES	117
CAPÍTULO VI: RECOMENDACIONES	120
REFERENCIAS	
ANEXOS	

LISTA DE TABLAS

Tabla 1	Fases de CRISP-ML(Q)
Tabla 2	Conceptualización del diseño experimental
Tabla 3	Criterios de elección de la muestra
Tabla 4	Definiciones para la variable independiente
Tabla 5	Clasificación de la variable independiente
Tabla 6	Definición para la variable dependiente
Tabla 7	Operacionalización para la variable dependiente
Tabla 8	Técnicas e instrumentos de recolección de información
Tabla 9	Técnicas e instrumentos de la investigación documental
Tabla 10	Validez de los instrumentos de expertos
Tabla 11	Correlaciones para el Indicador KPI 4: Tiempo de procesamiento de data
Tabla 12	Estadísticas de fiabilidad
Tabla 13	Estadísticas de total de elemento
Tabla 14	Pruebas médicas de laboratorio del dataset estudiado
Tabla 15	Etapas del análisis de datos en SPSS
Tabla 16	Pruebas de normalidad KPI 1
Tabla 17	Pruebas de normalidad KPI 2
Tabla 18	Pruebas de normalidad KPI 3
Tabla 19	Pruebas de normalidad KPI 4
Tabla 20	Pruebas de normalidad KPI 5
Tabla 21	Pruebas de normalidad KPI 6
Tabla 22	Análisis de la tasa de exactitud
Tabla 23	Análisis de la tasa de precisión
Tabla 24	Análisis de la tasa de sensibilidad
Tabla 25	Análisis del tiempo de procesamiento de datos
Tabla 26	Análisis del tiempo de predicción
Tabla 27	Análisis del nivel de confianza en la toma de decisiones
Tabla 28	Medias de los KPIs
Tabla 29	Valores pre-prueba del KPI 1
Tabla 30	Valores post-prueba del KPI 1
Tabla 31	Resultados de las pruebas de normalidad: KPI 1
Tabla 32	Prueba de Wilcoxon KPI 1

Tabla 33	Valores pre-prueba del KPI 2
Tabla 34	Valores post-prueba del KPI 2
Tabla 35	Resultados de las pruebas de normalidad: KPI 2
Tabla 36	Prueba de Wilcoxon KPI 2
Tabla 37	Valores pre-prueba del KPI 3
Tabla 38	Valores post-prueba del KPI 3
Tabla 39	Resultados de las pruebas de normalidad: KPI 3
Tabla 40	Prueba de Wilcoxon KPI 3
Tabla 41	Valores pre-prueba del KPI 4
Tabla 42	Valores post-prueba del KPI 4
Tabla 43	Resultados de las pruebas de normalidad: KPI 4
Tabla 44	Prueba de Wilcoxon KPI 4
Tabla 45	Valores pre-prueba del KPI 5
Tabla 46	Valores post-prueba del KPI 5
Tabla 47	Resultados de las pruebas de normalidad KPI 5
Tabla 48	Prueba de Wilcoxon KPI 5
Tabla 49	Valores pre-prueba del KPI 6 (Cuestionario a 16 personas, 10 preguntas)
Tabla 50	Valores post-prueba del KPI 6 (Cuestionario a 16 personas, 10 preguntas)
Tabla 51	Resultados de las pruebas de normalidad KPI 6
Tabla 52	Prueba de Wilcoxon KPI 6
Tabla 53	Pruebas médicas del dataset
Tabla 54	Métricas obtenidas del modelo - Logistic Regression
Tabla 55	Métricas obtenidas del modelo – Random Forest
Tabla 56	Métricas obtenidas del modelo – Gradient Boosting
Tabla 57	Métricas obtenidas del modelo – Decision Tree
Tabla 58	Métricas del análisis comparativo de modelos de Machine Learning

LISTA DE FIGURAS

- Figura 1 Porcentaje mercado global de tecnología médica de 2010 a 2022
- Figura 2 Gasto en salud alrededor del mundo
- Figura 3 Especialidades médicas clave en el desarrollo de algoritmos ML
- Figura 4 Arquitectura de modelo de predicción de riesgo para cirugías
- Figura 5 Representación del modelo Decision Tree
- Figura 6 Arquitectura de Bosque aleatorio
- Figura 7 Estimación de Riesgo de Diabetes usando regresión logística
- Figura 8 Ejemplo del uso Gradient Boosting
- Figura 9 Accuracy
- Figura 10 Precisión
- Figura 11 Sensibilidad y especificidad
- Figura 12 Entrenamiento de un modelo ML
- Figura 13 Clasificación de Machine Learning
- Figura 14 Histograma pre KP1: Tasa de exactitud
- Figura 15 Histograma post KPI 1: Tasa de exactitud
- Figura 16 Histograma pre KP2: Tasa de precisión
- Figura 17 Histograma post KP2: Tasa de precisión
- Figura 18 Histograma pre KP3: Tasa de sensibilidad
- Figura 19 Histograma post KP3: Tasa de sensibilidad
- Figura 20 Histograma pre KP4: Tiempo de procesamiento de datos
- Figura 21 Histograma post KP4: Tiempo de procesamiento de datos
- Figura 22 Histograma pre KP5: Tiempo de predicción
- Figura 23 Histograma post KP5: Tiempo de predicción
- Figura 24 Histograma pre KP6: Nivel de confianza en la toma de decisiones
- Figura 25 Histograma post KP6: Nivel de confianza en la toma de decisiones
- Figura 26 Análisis exploratorio de datos de pruebas médicas
- Figura 27 Histogramas de frecuencia de pruebas médicas por periodo del día
- Figura 28 Histogramas y KDE de las pruebas médicas
- Figura 29 Integración de datos en dataset consolidado
- Figura 30 Incorporación de datos
- Figura 31 Transformación de valores en el dataset
- Figura 32 Preparación de campo nuevo en el dataset de resultados clínicos

- Figura 33 Incorporación de nueva característica en el dataset en formato JSON
- Figura 34 Importación de librerías de Machine Learning – Logistic Regression
- Figura 35 Importación y conversión de datos JSON - Logistic Regression
- Figura 36 Codificación de etiquetas - Logistic Regression
- Figura 37 Codificación one-hot - Logistic Regression
- Figura 38 Estandarización - Logistic Regression
- Figura 39 Concatenación de datos - Logistic Regression
- Figura 40 División de datos de entrenamiento y prueba - Logistic Regression
- Figura 41 Evaluación del modelo con métricas de rendimiento - Logistic Regression
- Figura 42 Gráfico de barras de las métricas obtenidas - Logistic Regression
- Figura 43 Función para predicción desde archivos Pickle ML - Logistic Regression
- Figura 44 Predicciones realizadas en pruebas médicas - Logistic Regression
- Figura 45 Importación de librerías de Machine Learning – Random Forest
- Figura 46 Importación y conversión de datos JSON – Random Forest
- Figura 47 Codificación de etiquetas - Random Forest
- Figura 48 Codificación one-hot, estandarización y concatenación - Random Forest
- Figura 49 División de datos de entrenamiento y prueba - Random Forest
- Figura 50 Evaluación del modelo con métricas de rendimiento - Random Forest
- Figura 51 Gráfico de barras de las métricas obtenidas - Random Forest
- Figura 52 Función para predicción desde archivos Pickle ML - Random Forest
- Figura 53 Predicciones realizadas en pruebas médicas - Random Forest
- Figura 54 Importación de librerías de Machine Learning – Gradient Boosting
- Figura 55 Importación y conversión de datos JSON – Gradient Boosting
- Figura 56 Codificación de etiquetas - Gradient Boosting
- Figura 57 Codificación one-hot, estandarización y concatenación - Gradient Boosting
- Figura 58 División de datos de entrenamiento y prueba - Gradient Boosting
- Figura 59 Evaluación del modelo con métricas de rendimiento - Gradient Boosting
- Figura 60 Gráfico de barras de las métricas obtenidas – Gradient Boosting
- Figura 61 Función para predicción desde archivos Pickle ML – Gradient Boosting
- Figura 62 Predicciones realizadas en pruebas médicas – Gradient Boosting
- Figura 63 Importación de librerías de Machine Learning – Decision Tree
- Figura 64 Importación y conversión de datos JSON – Decision Tree
- Figura 65 Codificación de etiquetas – Decision Tree
- Figura 66 Codificación one-hot, estandarización y concatenación – Decision Tree

- Figura 67 División de datos de entrenamiento y prueba – Decision Tree
- Figura 68 Evaluación del modelo con métricas de rendimiento – Decision Tree
- Figura 69 Gráfico de barras de las métricas obtenidas – Decision Tree
- Figura 70 Métricas obtenidas del modelo – Decision Tree
- Figura 71 Predicciones realizadas en pruebas médicas – Decision Tree
- Figura 72 Aplicación de validación cruzada sobre el modelo Decision Tree
- Figura 73 Evaluación del modelo Decision Tree con el conjunto de prueba
- Figura 74 Obtención de métricas y elementos por prueba médica – Decision Tree
- Figura 75 Resultados de métricas de rendimiento de determinadas pruebas médicas
- Figura 76 Verificación de sobreajuste mediante curvas de aprendizaje
- Figura 77 Curvas de aprendizaje del Modelo Decision Tree
- Figura 78 Entorno Django donde se consume el Modelo ML
- Figura 79 Hosting donde sube la API Django
- Figura 80 Informe de laboratorio de paciente con predicción de resultado clínico
- Figura 81 Predicción de diagnóstico/resultados de pruebas médicas
- Figura 82 Predicción de diagnóstico/resultados de pruebas médicas desde archivo

**IMPLEMENTACIÓN DE UN MODELO BASADO EN ALGORITMOS DE MACHINE
LEARNING PARA MEJORAR EL ANÁLISIS PREDICTIVO DE DIAGNÓSTICOS
CLÍNICOS EN REDLAB PERÚ S.A.C**

**ANDERSON ALFREDO FUENTES PINEDO
CARLOS ENMANUEL NAVARRO CARAZA**

UNIVERSIDAD AUTÓNOMA DEL PERÚ

RESUMEN

La presente investigación evaluó la aplicación de algoritmos de Machine Learning en el laboratorio clínico RedLab Perú S.A.C., con el objetivo de optimizar la precisión y la rapidez en los diagnósticos clínicos. Se utilizaron algoritmos supervisados como regresión logística y árboles de decisión, aplicados a datos recolectados y preprocesados del sistema interno del laboratorio, los cuales fueron evaluados mediante métricas como precisión, sensibilidad y especificidad. Los resultados demostraron que los árboles de decisión destacaron entre los modelos empleados, logrando niveles de precisión entre el 98% y el 99%, reduciendo significativamente los tiempos de análisis y mejorando la calidad de los diagnósticos clínicos. Además, la implementación de estas herramientas tecnológicas permitió optimizar los flujos de trabajo, evidenciando su potencial para transformar los procesos de diagnóstico en el sector clínico-laboratorial. Se concluyó que la integración de Machine Learning en RedLab Perú S.A.C. representa una estrategia eficaz para potenciar la calidad y efectividad de los resultados clínicos, beneficiando directamente a la comunidad de pacientes y posicionando al laboratorio como líder en la adopción de tecnologías avanzadas en sus servicios.

Palabras clave: machine learning, predicción, crisp-ml(q), diagnóstico clínico

**IMPLEMENTATION OF A MODEL BASED ON MACHINE LEARNING
ALGORITHMS TO IMPROVE THE PREDICTIVE ANALYSIS OF CLINICAL
DIAGNOSES IN REDLAB PERU S.A.C.**

**ANDERSON ALFREDO FUENTES PINEDO
CARLOS ENMANUEL NAVARRO CARAZA**

UNIVERSIDAD AUTÓNOMA DEL PERÚ

ABSTRACT

This research evaluated the application of Machine Learning algorithms in the clinical laboratory RedLab Perú S.A.C., with the objective of optimizing the accuracy and speed of clinical diagnostics. Supervised algorithms, such as logistic regression and decision trees, were applied to data collected and preprocessed from the laboratory's internal system. These were evaluated using metrics such as accuracy, sensitivity, and specificity. The results showed that decision trees stood out among the models used, achieving accuracy levels between 98% and 99%, significantly reducing analysis times and improving the quality of clinical diagnostics. Furthermore, the implementation of these technological tools optimized workflows, demonstrating their potential to transform diagnostic processes in the clinical-laboratory sector. It was concluded that the integration of Machine Learning in RedLab Perú S.A.C. represents an effective strategy to enhance the quality and efficiency of clinical results, directly benefiting the patient community and positioning the laboratory as a leader in the adoption of advanced technologies in its services.

Keywords: machine learning, prediction, crisp-ml(q), clinical diagnosis

CAPÍTULO I
INTRODUCCIÓN

A medida que el mundo avanza en la digitalización, la salud no permanece inmune a estos cambios. El entrelazamiento progresivo de desarrollos tecnológicos punteros con las disciplinas médicas ha originado una sinergia innovadora en el sector de la salud que ha facilitado un cambio notable en el método en que las clínicas, laboratorios y hospitales optimizan sus procesos operacionales, brindando así un nivel de atención al paciente notablemente mejorado. Esta amalgama de disciplinas ha llevado al nacimiento de soluciones clínicas innovadoras, modificando fundamentalmente la perspectiva y el enfoque de la atención sanitaria.

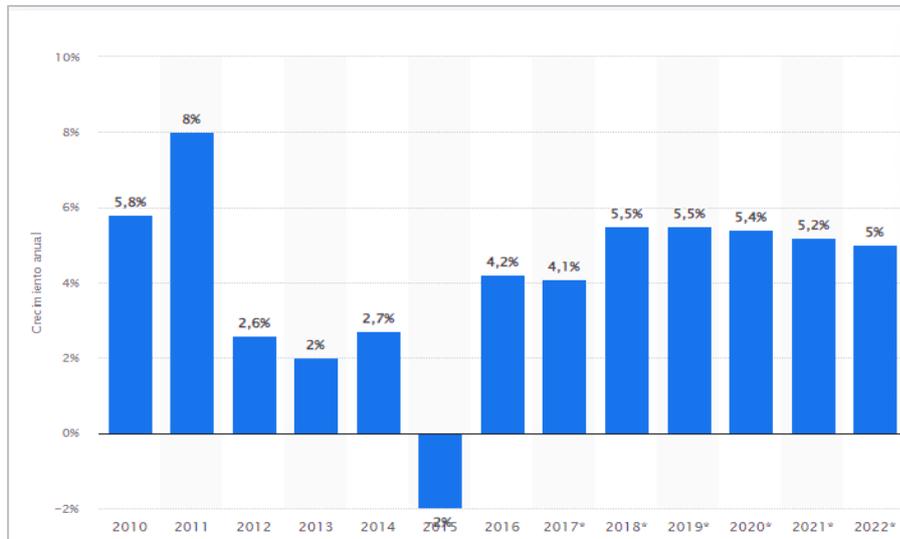
Sin embargo, una de las áreas críticas que aún demanda evolución es el análisis predictivo en diagnósticos clínicos. La rapidez y exactitud en este ámbito pueden marcar la diferencia en el resultado de un tratamiento, haciendo evidente que cada segundo cuenta. En este contexto, y aprovechando el aluvión de datos clínicos que la era digital ha puesto a disposición, las técnicas de Machine Learning, con su inherente capacidad para procesar y deducir patrones de vastos conjuntos de datos, se posicionan como herramientas de alto potencial.

No obstante, este fenómeno no se circunscribe únicamente a Perú. A escala global, el sector sanitario está atravesando una revolución digital, resultando en una acumulación de datos clínicos sin parangón. Desde las naciones más desarrolladas hasta las economías en desarrollo, todos están siendo testigos de cómo tecnologías punteras, como el Machine Learning, están transformando la atención médica. Pero, el mero acceso a esta profusión de datos no basta. El desafío principal radica en la gestión, análisis e interpretación de dicha información. No solo se busca precisión, sino también asegurar que los datos sean relevantes y aplicables en situaciones clínicas concretas. Este cruce entre tecnología y medicina es un caldo de cultivo para

oportunidades innovadoras, pero también es un terreno donde los desafíos no deben ser minimizados.

Figura 1

Porcentaje mercado global de tecnología médica de 2010 a 2022



Nota. El gráfico ilustra el crecimiento porcentual del mercado global de tecnología médica desde 2010 hasta 2022, destacando las fluctuaciones anuales y las tendencias clave en la expansión de este sector. Fuente: Statista (s.f.).

Dado el panorama actual, las aportaciones de expertos en la materia se han convertido también en luces guía. Por ejemplo, Joo et al. (2020) exploraron los desafíos tecnológicos inherentes al manejo de grandes volúmenes de datos, subrayando la importancia de contar con una infraestructura robusta y sistemas capaces de enfrentar dichos desafíos. En la misma línea, Alahmar y Benlamri (2020) destacaron el esfuerzo significativo requerido para transicionar innumerables registros médicos desde su estado físico original hacia el mundo digital. Esta digitalización no constituye un simple cambio de formato, sino una necesidad crítica para garantizar análisis clínicos de alta calidad y aprovechar plenamente el potencial de los algoritmos de aprendizaje automático, representando un desafío complejo.

La exactitud y confiabilidad de los datos en el ámbito médico también son esenciales. Errores o inconsistencias en el registro de información pueden tener

impactos negativos en la identificación y manejo terapéutico de los pacientes. Estos desafíos se intensifican cuando se agrupan datos de diversas fuentes a nivel global. Inconsistencias recurrentes podrían afectar la calidad de los análisis clínicos, generando dudas sobre el estado real de salud de los pacientes debido a registros inadecuados.

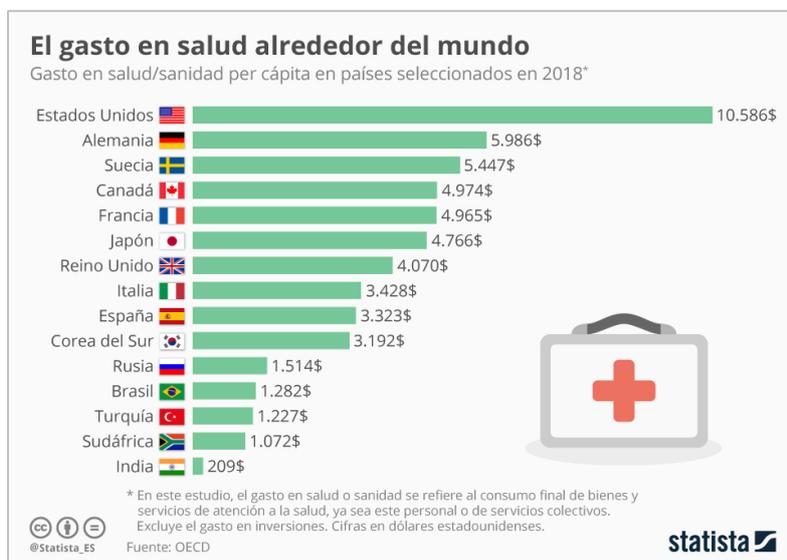
Estos desafíos pueden incrementarse en relación con los datos faltantes. Sin embargo, para abordar estos aspectos, existen diversas técnicas disponibles. Por ejemplo, Ibrahim et al. (2022), en su estudio sobre predicciones hospitalarias basadas en registros electrónicos de salud, destacaron métodos como la interpolación, que permite estimar valores entre dos puntos conocidos, y el escalado, que ajusta valores de diferentes rangos a una escala común, asegurando coherencia y comparabilidad. De manera similar, Montisci et al. (2022) analizaron la aplicación del Big Data en la cirugía cardíaca. En su trabajo, señalaron que la preocupación por la calidad de la información puede surgir debido a que la inteligencia computacional avanzada y el análisis de grandes conjuntos de datos aún no se han consolidado plenamente como herramientas de investigación rigurosa. A pesar del notable potencial de estas tecnologías para mejorar las prácticas y los resultados quirúrgicos, especialmente en el ámbito cardíaco, enfrentan desafíos significativos en cuanto a la recopilación, almacenamiento y análisis adecuado de los datos.

Un mal análisis puede deberse a la calidad de la información obtenida. Esto también ha sido destacado por Lbarrera (2022), quien resalta los riesgos asociados con la mala gestión de datos. La autora menciona que los datos poco fiables no solo pueden perjudicar la atención al paciente, sino que también influyen en la economía de los sistemas de salud, ya que los fondos gubernamentales deben ser administrados correctamente y no desperdiciarse en situaciones no prioritarias. A

largo plazo, un manejo inadecuado de esos fondos puede generar sobrecostos y tensiones presupuestarias, comprometiendo la calidad y el acceso a los servicios médicos para la comunidad.

Figura 2

Gasto en salud alrededor del mundo



Nota. El gráfico representa el gasto en salud per cápita en diversos países seleccionados durante el año 2018, destacando las variaciones entre regiones y su impacto en los sistemas de atención médica. Fuente: Statista (2019).

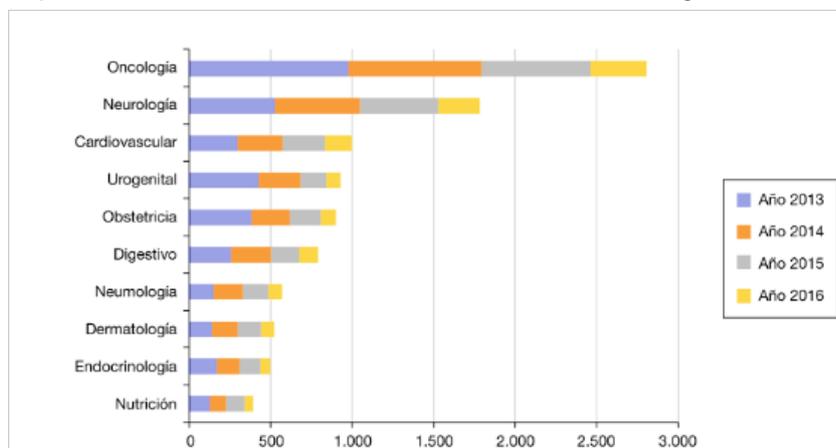
En el ámbito nacional, el Machine Learning ha promovido importantes avances en distintas áreas, destacándose especialmente en el sector médico. Sin embargo, la implementación eficaz de estas herramientas ha enfrentado diversos retos. La capacidad de las instituciones de salud para manejar y proteger grandes cantidades de datos es fundamental para obtener predicciones precisas. Asimismo, garantizar la integridad de la información y su accesibilidad resulta esencial para llevar a cabo análisis de manera eficiente.

Dentro de este contexto, se ofrece un enfoque esclarecedor en una investigación de 2020 que evaluó el impacto del aprendizaje automático en las predicciones diagnósticas de pacientes con COVID-19 en el Hospital María

Auxiliadora, destacando que la exactitud de los datos desempeña un papel crucial en proyectos de esta índole (Makishi & Rodríguez, 2020).

Figura 3

Especialidades médicas clave en el desarrollo de algoritmos ML



Nota. La figura muestra las especialidades médico-sanitarias más implicadas en el uso del Machine Learning entre 2013 y 2016. Adaptado de *Los profesionales que quieran humanizar la atención sanitaria tendrán un gran aliado en la inteligencia artificial*, por J. Beunza, 2020, Elsevier.

El empleo del Machine Learning en la predicción de enfermedades no es un fenómeno reciente, como señala Bolaños (2020). Estas técnicas ya se utilizaban en la detección de patologías mamarias de tipo oncológico mucho antes de la pandemia de COVID-19. Sin embargo, el autor destaca que, a pesar de la alta calidad de los datos actuales, persisten desafíos para predecir con precisión nuevos casos de COVID-19 debido a la escasez de datos disponibles o su carácter confidencial.

En el contexto empresarial, Redlab Perú S.A.C., uno de los laboratorios clínicos más destacados del país, enfrenta una encrucijada tecnológica centrada en la evolución y optimización de la interpretación de información clínica. A pesar del vasto volumen de datos que gestiona, el desafío principal radica en su capacidad para procesar y proporcionar diagnósticos precisos en tiempo real. Este reto surge, en gran medida, de la necesidad de analizar eficazmente los datos provenientes de pruebas diagnósticas, análisis clínicos y registros electrónicos de salud.

Ante los retos que enfrenta actualmente, Redlab ha reconocido la necesidad de integrar tecnologías basadas en Machine Learning, con el objetivo de transformar y agilizar los procesos de diagnóstico, garantizando interpretaciones más precisas y confiables. Si bien ya se han puesto en marcha varias iniciativas, persisten desafíos relacionados con los tiempos de entrega y la carga laboral del equipo. A pesar de ello, Redlab mantiene un firme compromiso con la mejora continua y la excelencia en sus servicios. Aunque la formación y el desarrollo del talento humano son componentes fundamentales, la sostenibilidad y escalabilidad de sus operaciones podrían depender, en gran medida, de la innovación tecnológica. En este contexto, un modelo avanzado de Machine Learning se perfila como una solución clave, no solo para optimizar la eficiencia en los diagnósticos, sino también para asegurar una integración efectiva y fluida de los resultados en los informes clínicos entregados a los pacientes.

La gestión eficiente de diagnósticos clínicos y su integración en informes para pacientes es un desafío actual que RedLab Perú no puede pasar por alto. Para abordar este reto, la solución necesita ser más amplia que simplemente expandir su equipo humano. Es vital que consideren adaptaciones tecnológicas, tales como la optimización de su infraestructura de TI, el despliegue de sistemas tecnológicos sofisticados para el manejo de datos clínicos, y, en particular, la integración de aplicaciones de aprendizaje automático enfocadas en optimizar y agilizar el análisis diagnóstico.

En el escenario actual, RedLab Perú S.A.C. enfrenta el desafío de adaptarse a las exigencias tecnológicas para mejorar sus procesos de diagnóstico clínico. La implementación de algoritmos de Machine Learning se presenta como una

oportunidad clave para optimizar la precisión y eficiencia de sus análisis, garantizando un servicio médico de calidad para sus pacientes.

Dado el contexto actual, surge una pregunta central en nuestra indagación: ¿En qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora el análisis predictivo de diagnósticos clínicos en RedLab Perú S.A.C.?

Para abordar este problema de forma integral, es esencial considerar algunas cuestiones específicas: ¿En qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de exactitud de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.? Además, ¿en qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de precisión de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.? Asimismo, ¿en qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de sensibilidad de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.? Por otra parte, ¿en qué medida la implementación de un modelo basado en algoritmos de Machine Learning reduce el tiempo requerido para procesar datos con fines diagnósticos en RedLab Perú S.A.C.? ¿En qué medida la implementación de un modelo basado en algoritmos de Machine Learning disminuye el tiempo requerido para predecir diagnósticos clínicos en RedLab Perú S.A.C.? Y, por último, ¿en qué medida la implementación de un modelo basado en algoritmos de Machine Learning incrementa el nivel de confianza en la toma de decisiones relacionadas con diagnósticos clínicos en RedLab Perú S.A.C.?.

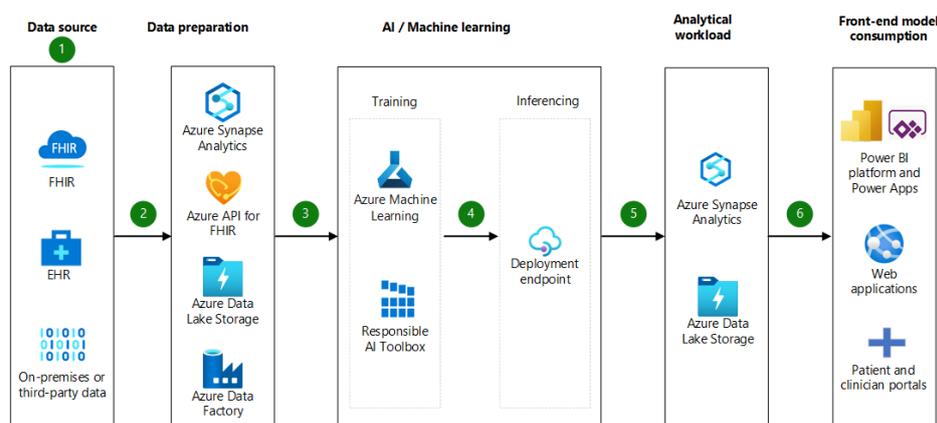
Por otra parte, la adaptabilidad de los modelos de Machine Learning les otorga un valor significativo en el ámbito médico. A medida que se agregan nuevos datos,

estos modelos pueden ajustarse y adaptarse para refinar sus predicciones, lo que significa que están en constante evolución y mejoramiento, respondiendo a las dinámicas y cambios en los datos médicos. Esta característica es esencial, especialmente en un campo donde las condiciones, tratamientos y tecnologías están en continuo desarrollo y revisión.

Un factor adicional que refuerza la utilidad de estos algoritmos es su capacidad de personalización. Los modelos pueden ser ajustados y afinados para adaptarse a distintos contextos y conjuntos de datos, permitiendo una implementación a medida según las necesidades de cada institución o aplicación médica en particular. Esto se traduce en diagnósticos y tratamientos más precisos, adaptados a las particularidades de cada paciente y contexto.

Figura 4

Arquitectura de modelo de predicción de riesgo para cirugías



Nota. La figura muestra la arquitectura de un modelo de predicción de riesgo en cirugías basado en Machine Learning. Adaptado de *Modelos de predicción de riesgos para cirugías - Azure Architecture Center*, por Microsoft (s. f.)

Es fundamental destacar el ahorro significativo de tiempo y recursos que los modelos basados en Machine Learning pueden aportar al sector médico. La automatización en el análisis de datos y la capacidad de procesar grandes volúmenes de información con rapidez permiten al personal de salud delegar tareas

repetitivas y enfocarse en los aspectos humanos del cuidado, tomando decisiones informadas y respaldadas por análisis más precisos y relevantes. Sin embargo, también es imprescindible abordar los desafíos que estas tecnologías plantean, como las implicaciones éticas en el manejo de datos personales, la interpretación adecuada de los patrones generados por los modelos y la necesidad de proporcionar formación especializada al personal sanitario para garantizar una implementación efectiva en esta área de rápido desarrollo.

Esta investigación tuvo como objetivo general determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora el análisis predictivo de diagnósticos clínicos en RedLab Perú S.A.C., y como objetivos específicos se plantearon los siguientes: determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de exactitud de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C., determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de precisión de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C., determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de sensibilidad de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C., determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning reduce el tiempo requerido para procesar datos con fines diagnósticos en RedLab Perú S.A.C., determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning disminuye el tiempo requerido para predecir diagnósticos clínicos en RedLab Perú S.A.C., y por último, determinar en qué medida la implementación de un modelo basado en

algoritmos de Machine Learning incrementa el nivel de confianza en la toma de decisiones relacionadas con diagnósticos clínicos en RedLab Perú S.A.C.

Dentro del marco investigativo, es esencial reconocer posibles limitaciones al desplegar el modelo de Machine Learning. Una de estas es la limitación técnica, donde es plausible enfrentar restricciones intrínsecas vinculadas a la tecnología empleada. Factores como la precisión de los algoritmos seleccionados, la cantidad y la excelencia de la información a disposición pueden ser limitantes, influyendo en la precisión de las predicciones y comprometiendo la eficacia general del modelo. Además, el despliegue opera bajo una cronología preestablecida, introduciendo una limitación temporal que delimita el margen para las fases esenciales del proceso: recolección y procesamiento de datos, implementación y calibración del modelo, y las fases de prueba y análisis de resultados. Finalmente, enfrentamos una limitación espacial, donde la dimensión geográfica y el acceso a los datos clínicos son cruciales. Ciertas restricciones en la disponibilidad o dispersión geográfica de estos datos pueden presentar desafíos al alcance y aplicabilidad del modelo, lo que resalta la importancia de tener en cuenta estas consideraciones en el diseño y evaluación del proyecto.

A nivel internacional, se han llevado a cabo diversos estudios que destacan el uso del aprendizaje automático en la predicción de la mortalidad en el ámbito médico. Khalaji et al. (2022) analizaron la precisión de modelos predictivos tras un injerto de bypass de arteria coronaria (CABG) y determinaron que el modelo de Regresión Logística (LR) presentó un desempeño sobresaliente, con un Área Bajo la Curva (AUC) de 0.81 (0.77-0.85). De manera similar, Zhang et al. (2022) investigaron la mortalidad intrahospitalaria en pacientes con enfermedad arterial periférica (PAD) y concluyeron que los modelos Light Gradient Boosting y XGBoost

(XGB) destacaron al alcanzar un AUC cercano a 0.85. Estos estudios, en conjunto, enfatizan la relevancia del aprendizaje automático en el campo médico, proporcionando herramientas valiosas para la toma de decisiones clínicas.

Tan et al. (2021) analizaron la predicción de cinco resultados clínicos en pacientes mayores con influenza en el departamento de emergencias, utilizando modelos como Random Forest (RF), XGBoost y Regresión Logística (LR). Random Forest destacó con un Área Bajo la Curva (AUC) de 0.840 para hospitalización, 0.765 para neumonía y 0.857 para sepsis. Por su parte, XGBoost y LR tuvieron un desempeño notable en predicciones específicas, con XGBoost alcanzando un AUC de 0.902 para ingreso a UCI y LR logrando un AUC de 0.889 para mortalidad intrahospitalaria. En otro estudio, Zhang et al. (2020) se enfocaron en la predicción de eventos cardíacos adversos mayores en pacientes con dolor en el pecho atendidos en el departamento de emergencias. Este trabajo empleó modelos como Random Forest, Regresión Logística, Support Vector Classifier (SVC) y K-Nearest Neighbors (KNN). Random Forest sobresalió con métricas como una exactitud de 0.915 y un AUC de 0.915, mientras que otros modelos, como la Regresión Logística, también demostraron buenos resultados, con una exactitud de 0.868 y un AUC de 0.868. Ambos estudios resaltan el impacto potencial de los algoritmos de inteligencia artificial en la optimización de decisiones médicas en tiempo real, aprovechando grandes volúmenes de datos para mejorar la precisión y la calidad de los resultados clínicos.

En el ámbito nacional, el aprendizaje automático ha sido empleado en el sector médico para optimizar los desenlaces clínicos desde diversas perspectivas. En el tratamiento de anomalías arteriovenosas cerebrales, se utilizaron técnicas como Naive Bayes y Regresión Logística, destacando la Máquina de Soporte

Vectorial (SVM) por su eficiencia superior en la predicción de resultados post-radiocirugía (Salas, 2021). Asimismo, en la identificación temprana de embarazos en riesgo en la región Junín, se implementaron algoritmos basados en árboles de decisión y redes neuronales, superando la precisión de los métodos tradicionales y subrayando la importancia de anticipar complicaciones en contextos prenatales (Gamarra, 2020).

Ambos estudios resaltan el potencial del aprendizaje automático como herramienta clave para mejorar la toma de decisiones clínicas. Estas tecnologías han demostrado su capacidad para abordar problemas médicos complejos, desde el tratamiento de malformaciones cerebrales hasta la atención prenatal, ofreciendo soluciones más precisas y eficaces en la práctica médica.

En el marco de los avances tecnológicos aplicados al sector salud, Cabrejos (2022) y Reyes (2023) abordan problemáticas distintas, pero comparten un objetivo común: mejorar la eficiencia y precisión de los métodos clínicos. Cabrejos (2022) se enfoca en el ámbito de la radiología, específicamente en el diagnóstico de neumonía, una enfermedad que representa una de las principales causas de mortalidad infantil a nivel global. Utilizando redes neuronales convolucionales y algoritmos como ResNet-50, AlexNet y VGG, su investigación logró agilizar el proceso de identificación de esta enfermedad mediante placas radiográficas, aumentando la precisión diagnóstica y demostrando el impacto de las técnicas computacionales en la mejora de la salud pública.

Por su parte, Reyes (2023) explora una problemática relacionada con la exposición de los trabajadores de salud a biocontaminantes en los laboratorios clínicos. Su propuesta tecnológica se centra en un sistema lector de capilares sanguíneos basado en visión artificial para determinar hematocrito, empleando

técnicas avanzadas como las Redes Neuronales Artificiales (ANN) y Máquinas de Soporte Vectorial (SVM). Este sistema no solo busca reducir el riesgo para los trabajadores de salud, sino también incrementar la precisión y rapidez en los análisis clínicos, destacando cómo la tecnología puede transformar la práctica médica al ofrecer soluciones más seguras y eficientes.

Otro autor destacado en este campo es Díaz (2021), quien en su tesis titulada *Predicción del rendimiento académico utilizando la técnica de árboles de decisión en los programas de maestría de educación en la escuela de posgrado de la Universidad Nacional José Faustino Sánchez Carrión* implementó el modelo de Árboles de Decisión (Decision Tree) como herramienta analítica principal. Este modelo demostró ser particularmente eficaz, alcanzando una exactitud del 87.76%. Los hallazgos destacan la capacidad de los Árboles de Decisión para predecir patrones complejos como la deserción estudiantil en contextos académicos, subrayando su utilidad en la mejora de los procesos educativos.

Para comprender de manera integral el enfoque de este estudio, resulta fundamental definir y describir algunos conceptos clave que sustentan la investigación. Esto permite establecer un marco teórico sólido y garantizar que las variables analizadas sean claramente entendidas en el contexto del problema planteado.

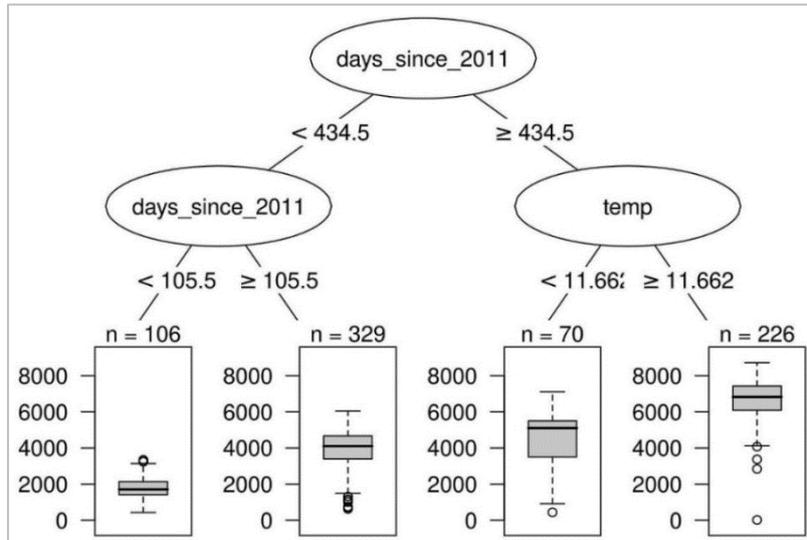
Este trabajo se centra en dos variables principales: un modelo basado en algoritmos de Machine Learning y el análisis predictivo de diagnósticos clínicos. Ambas variables están respaldadas por una revisión exhaustiva de la literatura experta, lo que refuerza la credibilidad y confiabilidad de esta investigación, asegurando un enfoque riguroso y fundamentado en prácticas científicas.

Los métodos de Machine Learning han revolucionado diversos campos al proporcionar herramientas analíticas avanzadas que superan las limitaciones de los enfoques tradicionales, permitiendo resolver problemas que antes parecían insuperables. Estas técnicas ofrecen la capacidad de desarrollar modelos personalizados y precisos a partir de extensos conjuntos de datos recolectados de diversas fuentes. Según Wang et al. (2020), el procedimiento para construir estos modelos incluye varios pasos fundamentales: preprocesamiento de los datos, selección de características relevantes, equilibrado de clases mediante técnicas de sobremuestreo, construcción del modelo, validación y evaluación final. Este enfoque permite generar predicciones que facilitan la toma de decisiones informadas en contextos complejos, optimizando los resultados y aportando valor en áreas como el diagnóstico médico.

Al seleccionar modelos de aprendizaje automático, es esencial optar por aquellos que sean robustos, eficientes y altamente adaptables, asegurándose de que se ajusten adecuadamente a las características del conjunto de datos y al problema específico que se busca resolver. Entre las diversas opciones disponibles, los árboles de decisión (Decision Trees) destacan notablemente por su estructura jerárquica, donde cada nodo terminal representa una categoría o resultado específico, y los nodos intermedios corresponden a atributos con umbrales bien definidos que guían el proceso de toma de decisiones. La clasificación de un caso particular se realiza siguiendo una ruta clara y estructurada desde la raíz del árbol hasta la hoja que coincide con las características del caso, lo que evidencia la capacidad de los árboles de decisión para facilitar tanto la interpretación como la clasificación precisa dentro de un marco de datos predefinido, mejorando la eficiencia y la comprensión del modelo (Molnar, 2022).

Figura 5

Representación del modelo Decision Tree



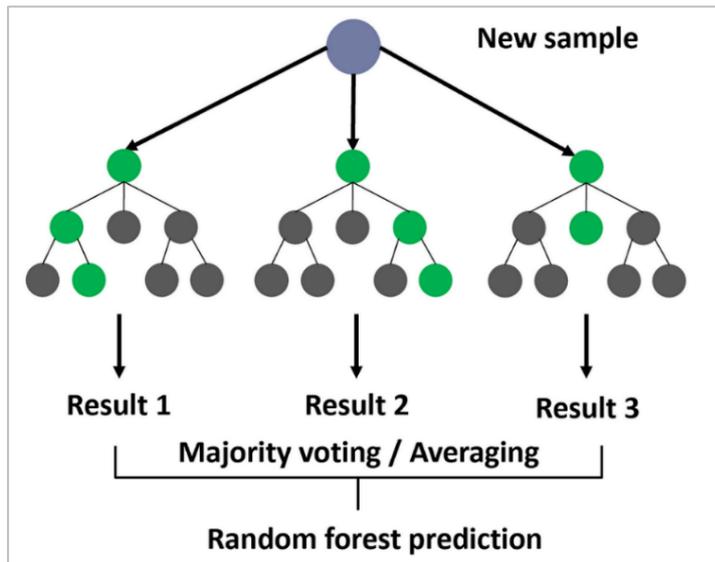
Nota. En la figura, se observa un diagrama de representación del modelo Decision Tree, que ilustra la estructura jerárquica de este tipo de algoritmo, con nodos que representan atributos y umbrales para guiar la clasificación. Fuente: Molnar, C. (2022)

Posteriormente, se tomó en cuenta el modelo de Bosque Aleatorio (Random Forest, RF) por su reconocida eficiencia en el aprendizaje conjunto y su capacidad para manejar grandes volúmenes de datos con múltiples variables de manera simultánea. Este modelo se distingue por su capacidad para ofrecer una alta precisión predictiva, reducir el riesgo de sobreajuste mediante la combinación de resultados de múltiples árboles de decisión y adaptarse a diferentes tipos de problemas analíticos, lo que lo convierte en una herramienta clave dentro del análisis realizado (Aria et al., 2021).

$$\bar{F}_l(z_l) = \frac{1}{N} \sum_{i=1}^N \hat{F}(z_l, z_i, \neg)$$

Donde:

- N: Número de elementos.
- z_l : Argumento.
- $\bar{F}_l(z_l)$: Función.

Figura 6*Arquitectura de Bosque aleatorio*

Nota. En la figura, se observa el diagrama de arquitectura del modelo de Bosque Aleatorio (Random Forest). Fuente: Yehoshua, R. (2023)

En el contexto de los modelos de Machine Learning, la Regresión Logística (Logistic Regression) ha sido ampliamente reconocida por su capacidad para manejar predicciones en eventos binarios. Según el estudio de Song et al. (2023), la formación y validación de este modelo se lleva a cabo mediante la selección aleatoria de pacientes, manteniendo una proporción de 3 a 1. Asimismo, el análisis multivariable incluye únicamente las variables con relevancia estadística identificadas en una evaluación preliminar. Finalmente, el modelo se perfecciona a través de técnicas progresivas de selección de variables, tanto incrementales como decrementales, permitiendo identificar las características más relevantes para la predicción.

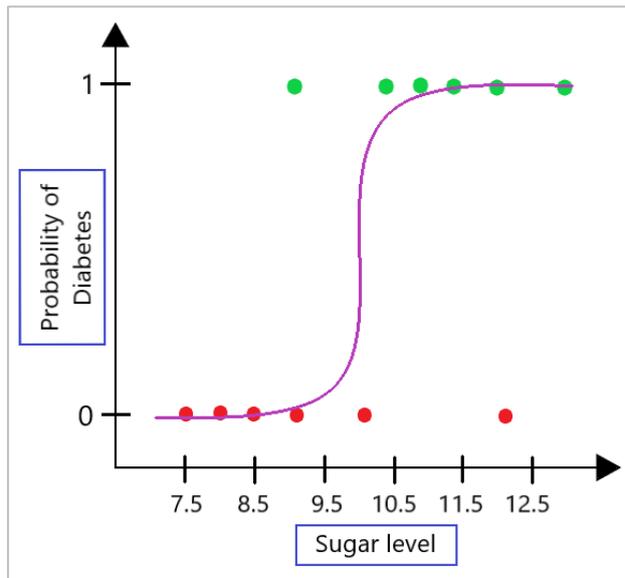
$$P = \frac{1}{1 + e^{-g(x)}}, \quad g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n.$$

Donde:

- e: Base del logaritmo natural
- β_0 : Es el término de intercepción o sesgo del modelo.

Figura 7

Estimación de Riesgo de Diabetes usando regresión logística



Nota. En la figura, se observa el uso de Logistic Regression aplicado al análisis de riesgo de diabetes.

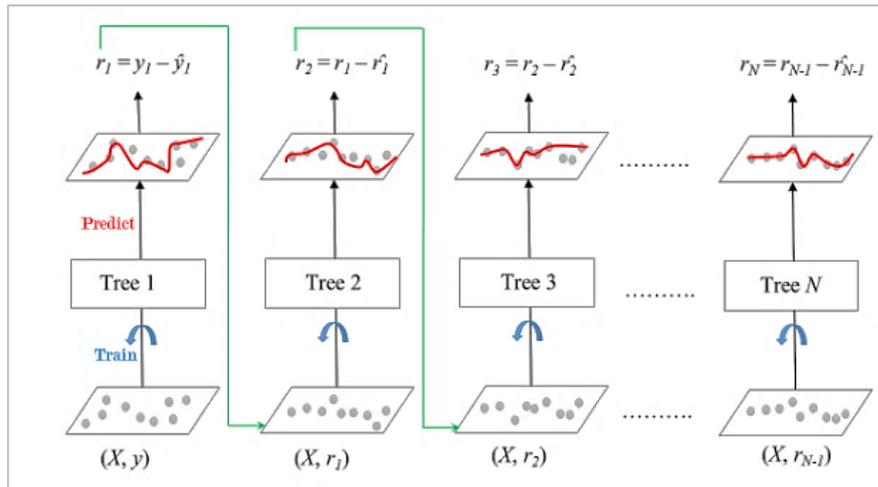
Fuente: Cost function in Logistic Regression in Machine Learning (GeeksforGeeks, 2019)

En el contexto del aprendizaje automático, se consideró el modelo de Gradient Boosting (GB) debido a su capacidad para construir un modelo predictivo de manera iterativa, mejorando constantemente mediante la rectificación de errores cometidos en etapas previas. Este enfoque se caracteriza por su habilidad para combinar modelos base simples, como en el caso del EGBM (una Máquina de Soporte Vectorial con núcleo de Función de Base Radial, SVM-RBF), que se perfeccionan iterativamente mediante la incorporación de nuevos modelos que se enfocan en corregir los errores existentes. De este modo, se logra aumentar significativamente la precisión del modelo final, optimizando el rendimiento y adaptándose a la naturaleza compleja de los datos (AlShourbaji et al., 2023).

$$r_{i,n} = -\frac{\partial L(y_n, F(x_n))}{\partial F(x_n)}$$

Donde:

- ∂ : Derivada parcial.
- L: Función de pérdida.

Figura 8*Ejemplo del uso Gradient Boosting*

Nota. En la figura, se presenta un ejemplo del proceso iterativo de Gradient Boosting, donde se utilizan múltiples árboles de decisión para corregir los errores de predicción en cada iteración. Este enfoque permite mejorar progresivamente la precisión del modelo predictivo al ajustar los resultados en función de los residuos generados en las etapas anteriores. Fuente: GeeksforGeeks (2020)

El análisis predictivo emplea algoritmos de alto rendimiento diseñados para pronosticar y anticipar resultados relacionados con la salud de los pacientes. Este enfoque, aplicado en un contexto clínico, tiene como principal objetivo optimizar la toma de decisiones de los profesionales sanitarios. Para ello, se realiza un análisis exhaustivo que incluye la exploración, preparación y preprocesamiento de conjuntos de datos clínicos. Posteriormente, estos datos se validan mediante diversas métricas de rendimiento que evalúan tanto la calidad como la efectividad de las predicciones, considerando también el tiempo como un factor crítico en la implementación de técnicas de análisis predictivo. Además, la integración del análisis predictivo con sistemas de apoyo a la toma de decisiones permite avanzar hacia la implementación de análisis prescriptivos, ampliando su utilidad en el ámbito clínico (Divyashree & Prasad, 2022).

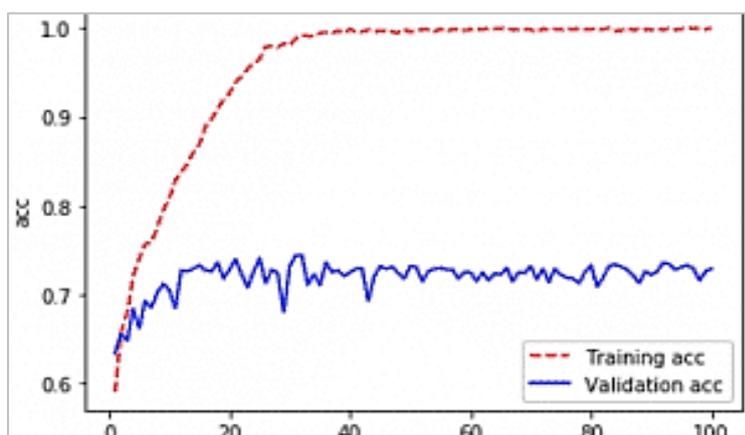
Las métricas de rendimiento desempeñan un papel crucial en la evaluación de la eficiencia y precisión de los algoritmos utilizados en análisis predictivo. Entre

las principales métricas se encuentran la Exactitud (Accuracy), la Curva ROC y el F1-Score, que ofrecen una visión integral de la operatividad y confiabilidad del algoritmo. Estas herramientas permiten comprender cómo se comporta el modelo en diferentes escenarios y garantizan que las decisiones basadas en estos algoritmos sean fundamentadas y fiables (Divyashree & Prasad, 2022).

La exactitud (Accuracy) es una métrica esencial en la evaluación de clasificadores, calculada como la suma de los Verdaderos Positivos (TP) y los Verdaderos Negativos (TN), dividida entre el total de casos evaluados, que incluye TP, TN, Falsos Positivos (FP) y Falsos Negativos (FN). Mohan et al. (2019) refuerzan esta perspectiva al destacar que la exactitud se fundamenta en la matriz de confusión, un marco analítico clave para evaluar el desempeño de los modelos. Por su parte, Noor, Ibrahim, Lah y Abdullah (2021) subrayan que esta métrica adquiere especial relevancia en contextos con conjuntos de datos desbalanceados, y su combinación con otras métricas, como la sensibilidad y la especificidad, puede optimizar la interpretación de los resultados y mejorar las predicciones en escenarios clínicos complejos.

Figura 9

Accuracy

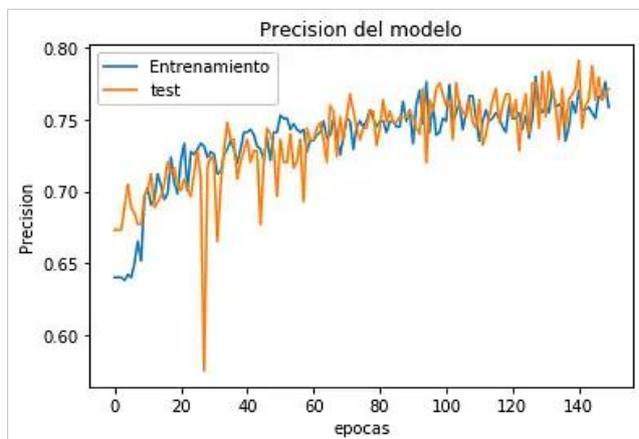


Nota. En la figura se observa la evolución de la exactitud (Accuracy) durante el entrenamiento y la validación de un modelo. Fuente: Torres.ai (s.f.)

En el ámbito del análisis de modelos predictivos, es esencial garantizar que las predicciones sean tanto precisas como relevantes. La métrica de precisión se define como el porcentaje de instancias relevantes identificadas entre todas las instancias recuperadas o clasificadas como positivas por el modelo. Esta métrica resulta crucial, ya que destaca la capacidad del modelo para identificar correctamente instancias relevantes mientras minimiza la inclusión de resultados irrelevantes o falsos positivos. Por ende, una alta precisión refleja que las predicciones realizadas por el modelo, al clasificar una instancia como positiva, son mayoritariamente correctas y relevantes en el contexto de aplicación (Srimekha et al., 2022).

Figura 10

Precisión



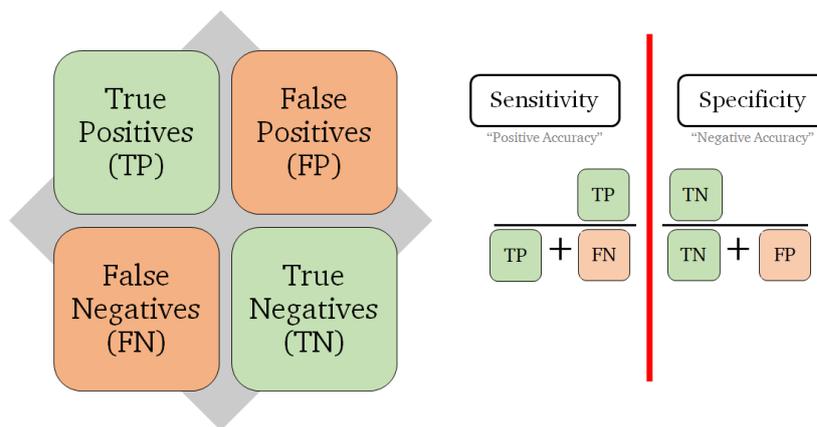
Nota. En la figura, se observa el gráfico de precisión del modelo durante las épocas de entrenamiento y prueba. Fuente: Fernández, R. (s.f.)

Dentro del ámbito de la evaluación de modelos clasificadores, es esencial garantizar que las decisiones tomadas por estos sean precisas y altamente relevantes. La métrica de sensibilidad se destaca como un elemento clave, definida como la capacidad del modelo para identificar correctamente la clase positiva o minoritaria dentro del conjunto de datos. Su relevancia es especialmente notable en contextos donde la detección de clases menos representadas resulta crítica. En este

sentido, la sensibilidad se posiciona como un indicador fiable de la habilidad del modelo para identificar y clasificar con precisión eventos o instancias poco frecuentes, pero significativamente relevantes para el ámbito de estudio. Una sensibilidad elevada refleja la robustez del modelo para no pasar por alto casos minoritarios que son esenciales en el contexto en que se aplica (Alahmar & Benlamri, 2020).

Figura 11

Sensibilidad y especificidad



Nota. En la figura se presentan los conceptos de sensibilidad y especificidad, ilustrando la relación entre verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN). Adaptado de *What are classification metrics?*, por Y. Serpa, 2020, *Towards Data Science*.

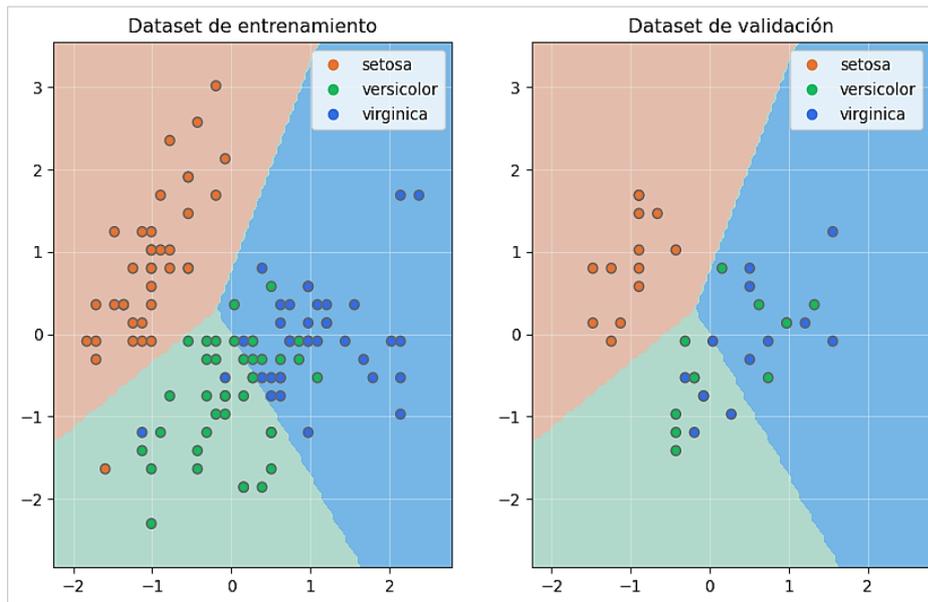
Los modelos de predicción actuales sobresalen por su celeridad en la presentación de resultados, lo cual es fundamental en el ámbito médico. Esta capacidad de proporcionar predicciones rápidas se traduce en una herramienta invaluable para la toma de decisiones clínicas, donde la agilidad en el análisis puede marcar una diferencia significativa en los tratamientos y en el bienestar de los pacientes (Tan et al., 2021).

La importancia de considerar el tiempo en el desempeño de los modelos de aprendizaje automático (ML) es crucial para garantizar resultados óptimos y eficientes. Según Wang et al. (2020), existe una relación directa entre el tiempo que se invierte en entrenar un modelo y su rendimiento final. Esta dinámica sugiere que,

a medida que se dedica más tiempo al entrenamiento, pueden producirse variaciones significativas en las métricas de rendimiento, ya sea mejorando la precisión o revelando áreas de ajuste. Esto resalta la necesidad de evaluar cuidadosamente el equilibrio entre la precisión del modelo y el tiempo invertido en su entrenamiento y prueba, asegurando un uso eficiente de los recursos disponibles.

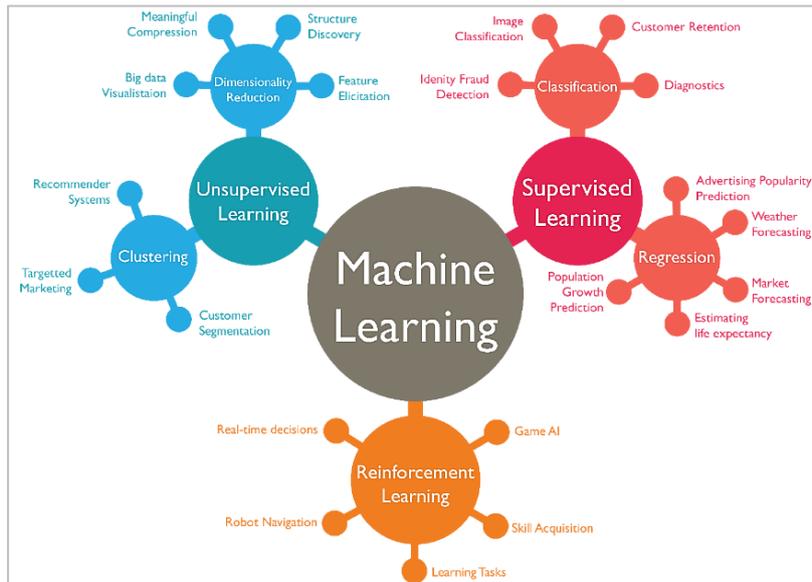
Figura 12

Entrenamiento de un modelo ML



Nota. En la figura, se observa el entrenamiento y la validación de un modelo de aprendizaje automático con el conjunto de datos Iris. Fuente: InteractiveChaos (s.f.).

La implementación del aprendizaje automático (ML) ha transformado radicalmente el manejo de grandes volúmenes de información, permitiendo un procesamiento y análisis de datos mucho más rápido, preciso y eficiente. Esta tecnología no solo reduce significativamente los tiempos de respuesta, sino que también facilita la identificación de patrones, tendencias y correlaciones complejas que, con métodos tradicionales, habrían permanecido ocultos o serían difíciles de detectar. De esta manera, el ML se consolida como una herramienta esencial para extraer conocimientos valiosos de datos masivos, optimizar procesos y apoyar la toma de decisiones basada en evidencia (Srimedha et al., 2022).

Figura 13**Clasificación de Machine Learning**

Nota. Clasificación de Machine Learning según las principales categorías: aprendizaje supervisado, no supervisado y por refuerzo, destacando sus aplicaciones más comunes. Fuente: ICCSI (2023).

El tiempo para predecir se refiere al intervalo requerido por el modelo tanto para su entrenamiento como para la realización de posibles predicciones. Esta métrica es esencial para evaluar la eficiencia de algoritmos en entornos en tiempo real y aplicaciones críticas donde la rapidez es fundamental (Ghorbani et al., 2020).

Se indica que la predicción en tiempo real es una herramienta prometedora para asistir a los médicos en la toma de decisiones y en la comunicación de explicaciones claras a los pacientes y sus familias. Zhang et al. (2020) destacan la importancia de evaluar el impacto de estas herramientas en las decisiones clínicas de los médicos, incluyendo factores como la aceptación, el rechazo, las preocupaciones, posibles cambios en sus prácticas clínicas y los resultados observados en los pacientes. Por su parte, Tan et al. (2021) subrayan que el uso de enfoques basados en datos masivos y aprendizaje automático en tiempo real puede optimizar la atención médica, especialmente en pacientes mayores, al mejorar la precisión de los resultados y la eficacia de los sistemas hospitalarios.

La toma de decisiones en el ámbito médico reviste una importancia crítica, ya que en ello puede residir la clave para la salud y supervivencia de los pacientes. Un caso ilustrativo es el de la predicción de la hipotensión intradiálisis (IDH) durante la hemodiálisis. Este proceso presenta desafíos significativos, ya que los modelos tradicionales no logran identificar las intrincadas relaciones inherentes a la IDH. Sin embargo, con los avances recientes, se han desarrollado modelos más sofisticados que ofrecen un apoyo valioso en la toma de decisiones clínicas (Mendoza, 2022).

En relación a la terminología clave utilizada a lo largo de esta investigación, el aprendizaje automático se destaca como un método avanzado enfocado en la identificación y análisis de patrones dentro de grandes volúmenes de datos. Esta metodología resulta extremadamente útil para la mejora en procesos decisionales, especialmente en el sector sanitario, donde la vasta generación de información juega un papel crucial. Según Mohan et al. (2019), el proceso de Machine Learning comienza con el preprocesamiento de datos, sigue con la selección de características relevantes y culmina con la clasificación y evaluación del rendimiento del modelo. Este enfoque meticuloso permite alcanzar predicciones con un alto grado de precisión, lo que demuestra el potencial transformador del aprendizaje automático para convertir grandes cantidades de datos en insights significativos y aplicables en la práctica médica.

La estructura de aprendizaje automático se refiere a sistemas capaces de desentrañar interpretaciones novedosas de interacciones complejas y no lineales entre variables. Estos modelos tienen la habilidad de discernir la relevancia, ya sea cierta o incierta, de variables destacadas, proporcionando una comprensión más profunda de los datos. Además, la identificación de estas variables, tanto las

previamente reconocidas como las recién descubiertas, sirve como punto de partida para investigaciones posteriores en el campo (Beecy et al., 2020).

En la práctica del aprendizaje automático, evaluar el desempeño de los modelos predictivos frente a datos futuros es un paso fundamental para garantizar su efectividad. Tanto los modelos de regresión como los de clasificación requieren métricas de rendimiento adecuadas que permitan medir su precisión y utilidad en distintos contextos. Para conjuntos de datos balanceados, indicadores como la precisión y la exactitud son ampliamente utilizados. Sin embargo, en escenarios con datos desequilibrados, métricas más avanzadas como ROC/AUC son preferidas, ya que ofrecen una evaluación más precisa del rendimiento del modelo y su capacidad de generalización (DataSource.ai, 2023).

La predicción del diagnóstico clínico implica el análisis de datos e información para anticipar resultados médicos o condiciones de salud futuras. En el ámbito académico, este enfoque ha adquirido gran relevancia debido a su capacidad para establecer relaciones significativas entre grandes volúmenes de datos clínicos y los posibles desenlaces asociados. La precisión y rapidez en estas predicciones resultan fundamentales, especialmente en entornos críticos como las unidades de cuidados intensivos, donde cada segundo puede marcar una diferencia. En este contexto, las notas médicas se destacan como una fuente esencial de información, ya que contienen detalles clave sobre el estado y las condiciones del paciente, incluso en ausencia de otras herramientas diagnósticas. La interpretación precisa y la integración adecuada de estas notas en modelos predictivos permiten una comprensión más profunda de los factores que afectan la salud, facilitando diagnósticos más certeros y tratamientos oportunos (Mugisha & Paik, 2022).

Con respecto a la metodología utilizada para el desarrollo, CRISP-ML(Q) se presenta como una metodología de aseguramiento de calidad específicamente diseñada para evaluar si las tareas clave en el ámbito del aprendizaje automático (Machine Learning) se llevan a cabo siguiendo los estándares establecidos por las mejores prácticas de la industria y las recomendaciones de la literatura académica. Según Studer et al. (2020), esta metodología no solo busca garantizar la calidad técnica, sino también promover un enfoque estructurado y replicable en proyectos de Machine Learning, maximizando la confiabilidad y efectividad de los resultados.

A continuación, se mostrará una tabla con las fases de esta metodología:

Tabla 1

Fases de CRISP-ML (Q)

ID	Fases	Descripción
01	Comprensión del Negocio y los Datos	En la fase inicial, se definen metas empresariales, se convierten en objetivos de ML y se verifica la información para evaluar la viabilidad del proyecto.
02	Preparación de la data	Luego de comprender los datos, se acondiciona la información para el modelado, realizando ajustes si aparecen errores en etapas posteriores como modelado o despliegue.
03	Modelado	La etapa de modelización tiene como objetivo desarrollar uno o varios modelos que se ajusten a las especificaciones y restricciones establecidas.
04	Evaluación	El análisis crítico es clave en el desarrollo de soluciones de ML, validando que el modelo sea efectivo, robusto y claro, y asegurando que los datos no se mezclen entre los conjuntos de entrenamiento, validación y evaluación.
05	Despliegue	La implementación de un modelo ML se centra en su aplicación efectiva, eligiendo el hardware de inferencia que mejor se ajuste a los requisitos, con recursos escalables en la nube o restricciones en sistemas embebidos.
06	Monitoreo y mantenimiento	La gestión del ciclo de vida de los modelos ML es crucial para prevenir la disminución de precisión y errores en predicciones, causada por diferencias entre los datos actuales y los usados en entrenamiento.

CAPÍTULO II

METODOLOGÍA

2.1. Tipo y diseño de investigación

Tipo de investigación

Aplicada

En el marco de este estudio actual se aplica la solución a la problemática, buscando el análisis predictivo de diagnósticos clínicos en RedLab Perú S.A.C.

Nivel de investigación

Explicativa

Este estudio actual tiene como objetivo descubrir las relaciones de causa y efecto que surgen de la implementación de un modelo basado en algoritmos de machine learning en el análisis predictivo de diagnósticos clínicos.

Diseño de la investigación

Pre-experimental: El diseño pre-experimental implica la exposición de un solo grupo a la variable de intervención, seguida de una evaluación u observación posterior al tratamiento (Claudio, 2024).

La configuración de este estudio se basa en un diseño pre-experimental. Para todos los KPIs, incluidos los relacionados con la precisión, exactitud, sensibilidad, y tiempos de procesamiento y predicción, así como el "nivel de confianza en la toma de decisiones", se utiliza un enfoque exclusivamente de post-test, ya que no se dispone de una medición previa.

Tabla 2

Conceptualización del diseño experimental

G	O1	X	O2
Grupo experimental: A este grupo es al que se le someterá el estímulo	Datos de Pre prueba: Prueba de los KPIs de la variable dependiente que se mide de la pre prueba del grupo experimental.	Condición experimental: Estímulo o condición aplicada.	Datos de Post prueba: Prueba de los KPIs de la variable dependiente que se mide de la post prueba del grupo experimental.

De la tabla anterior, se indica que:

Ge (Grupo experimental): Representa las pruebas médicas de laboratorio realizadas en RedLab Perú S.A.C.

O1: Valores obtenidos por cada KPI en la pre-prueba, que se realiza antes de implementar el modelo de Machine Learning.

X: Intervención realizada, que es la implementación y utilización del modelo de Machine Learning para predecir diagnósticos clínicos.

O2: Valores obtenidos por cada KPI en la post-prueba, que se realiza después de implementar el modelo de Machine Learning.

Descripción

De donde se puede explicar que se está desarrollando una investigación pre-experimental en RedLab Perú S.A.C. para evaluar la influencia del Machine Learning en el ámbito sanitario. El estudio se enfoca en un grupo experimental, evaluado mediante KPIs antes (O1) y después (O2) de la intervención del modelo (X). Los resultados post-intervención se analizan para evidenciar la contribución del Machine Learning en el laboratorio clínico.

2.2. Población, muestra y muestreo

Población

Según las ideas presentadas por Mendoza (2022), la población se refiere a la totalidad de individuos que conforman una comunidad o sociedad y que tienen una relación directa o indirecta con el objeto de estudio. Esta definición enfatiza la importancia de considerar a todos los miembros relevantes para garantizar que estén informados y que los resultados obtenidos en la investigación puedan ser aplicados o entendidos en el contexto de esa población específica.

Para la presente investigación, la población considerada incluye las pruebas médicas realizadas en el área de análisis clínico de Redlab Perú, abarcando una amplia variedad de biomarcadores. Estas pruebas fueron realizadas por profesionales tecnólogos y han sido registradas en un determinado periodo de tiempo.

N = 120 pruebas médicas de laboratorio

Muestra

Según explican Vizcaíno et al. (2023), una muestra es un subconjunto seleccionado de una población o grupo, diseñado para que sus características y variaciones representen de manera precisa y equitativa las propiedades de la población o grupo estudiado. Este enfoque permite extrapolar los hallazgos obtenidos en la muestra al conjunto total, facilitando conclusiones más generalizables y representativas.

La muestra de la investigación estará compuesta por 26 pruebas médicas de laboratorio realizadas en Redlab Perú, seleccionadas mediante un muestreo por conveniencia y de acuerdo con criterios de inclusión y exclusión.

N = 26 pruebas médicas de laboratorio

Tipo de muestra

Según menciona Mendoza (2022), el muestreo constituye una selección representativa de variables aleatorias extraídas de un grupo poblacional más grande para luego ser estudiadas.

Muestreo no probabilístico

El muestreo no probabilístico se distingue porque no proporciona a cada miembro de la población una posibilidad definida y distinta de cero de ser elegido para la muestra. Existen distintos tipos de muestreo no probabilístico, entre los que

destacan: por conveniencia, intencional, por cuotas y de bolas de nieve. Cada uno tiene sus propias características, ventajas y limitaciones.

De acuerdo con Hernández (2021), el muestreo por conveniencia se caracteriza por la selección de unidades de análisis basándose principalmente en su facilidad de acceso. Esto implica que las unidades son elegidas debido a su proximidad y la accesibilidad que ofrecen al investigador, facilitando el proceso de recolección de datos sin necesariamente garantizar la representatividad de la población estudiada.

Para este estudio, se ha elegido la técnica de muestreo por conveniencia debido a la accesibilidad y facilidad de acceso a las pruebas médicas de RedLab Perú S.A.C., sustentada por criterios de inclusión y exclusión claramente definidos. Este enfoque se considera apropiado para los objetivos de la investigación.

Tabla 3

Criterios de elección de la muestra

ID	Criterios de inclusión	Criterios de exclusión
01	Exámenes con mayor frecuencia en el año 2022 y 2023.	Exámenes no solicitados comúnmente
02	Exámenes más resaltantes para la salud pública	Exámenes con datos históricos insuficientes

La tabla presentada muestra los criterios para la selección rigurosa de exámenes clínicos, basados en los más comunes en RedLab Perú durante 2022 y 2023, así como aquellos cruciales para el diagnóstico de enfermedades graves, como las cardíacas, obstétricas y otras de alta complejidad. Además, se han priorizado los estudios con alta relevancia médica y utilidad clínica. Se han excluido los exámenes con baja demanda y representación en el conjunto de datos de la empresa.

2.3. Hipótesis

Hipótesis general

La implementación de un modelo basado en algoritmos de Machine Learning mejora el análisis predictivo de diagnósticos clínicos en RedLab Perú S.A.C.

Hipótesis específicas

La implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de exactitud de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.

La implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de precisión de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.

La implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de sensibilidad de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.

La implementación de un modelo basado en algoritmos de Machine Learning reduce el tiempo requerido para procesar datos con fines diagnósticos en RedLab Perú S.A.C.

La implementación de un modelo basado en algoritmos de Machine Learning disminuye el tiempo requerido para predecir diagnósticos clínicos en RedLab Perú S.A.C.

La implementación de un modelo basado en algoritmos de Machine Learning incrementa el nivel de confianza en la toma de decisiones relacionadas con diagnósticos clínicos en RedLab Perú S.A.C.

2.4. Variables y operacionalización

Variable independiente

Tabla 4*Definiciones para la variable independiente*

Variable independiente	Definición conceptual	Definición operacional
Modelo basado en algoritmos de Machine Learning (X)	Un modelo basado en algoritmos de Machine Learning se refiere a un sistema diseñado para analizar y aprender de grandes volúmenes de datos con el objetivo de generar predicciones precisas y confiables. El desarrollo de estos modelos implica etapas fundamentales como el preprocesamiento de datos, la identificación de características relevantes y la validación del modelo. La precisión y utilidad del sistema dependen directamente de la ejecución meticulosa de cada fase, garantizando resultados que puedan contribuir de manera eficaz a la toma de decisiones en contextos complejos (Wang et al., 2020)	<p>Cuando es NO, es porque no existe el modelo basado en algoritmos de Machine Learning en la empresa RedLab Perú S.A.C, y la situación actual del problema persistiría.</p> <p>Cuando es SÍ, se aplica la solución (Modelo ML) con la expectativa de obtener mejores resultados.</p>

Tabla 5*Clasificación de la variable independiente*

Dimensión	Indicador	Definición operacional
	Presencia-Ausencia	En un grupo de estudio se implementa el modelo basado en algoritmos de Machine Learning, mientras que en otro grupo no se utiliza en la actualidad.

Variable dependiente

Tabla 6

Definición para la variable dependiente

Variable dependiente	Definición conceptual	Definición operacional
Análisis predictivo de diagnósticos clínicos (Y)	<p>El análisis predictivo ha emergido como una herramienta esencial en el ámbito clínico, ofreciendo soluciones innovadoras para interpretar datos complejos y mejorar la toma de decisiones médicas.</p> <p>En este contexto, se basa en el uso de algoritmos avanzados para prever resultados relacionados con la salud de los pacientes, con el objetivo de fortalecer y respaldar las decisiones médicas. Este enfoque requiere la exploración, procesamiento y validación de conjuntos de datos clínicos mediante diversas métricas de rendimiento. Estas métricas no solo miden la precisión de las predicciones, sino que también consideran la eficiencia en la implementación de estas técnicas. Al combinar el análisis predictivo con sistemas de apoyo a la toma de decisiones, su utilidad se amplifica, ofreciendo un marco más robusto para orientar acciones médicas específicas (Divyashree & Prasad, 2022)</p>	<p>La eficacia del modelo clínico se evalúa en base a tres métricas: exactitud, que debe superar el 85% para ser óptima; precisión; y sensibilidad.</p> <p>Además, se monitorizan dos tiempos críticos: el necesario para procesar los datos y el requerido para realizar la predicción diagnóstica. Ambos deben ser menores a 5 segundos.</p> <p>Finalmente, la confianza en el modelo se mide utilizando una escala de Likert, con respuestas que varían desde TCO (Totalmente conforme) hasta TDCO (Totalmente disconforme).</p>

Tabla 7*Operacionalización para variable dependiente*

Dimensión	Indicador	Tipo	Unidad de medida	Índice	Técnica	Instrumento	Fuente verificación
Métricas de rendimiento	Tasa de exactitud de la predicción	Cuantitativo	Proporción / Porcentaje (%)	>85%	Observación	Ficha de registro	Reportes de Software
	Tasa de sensibilidad de la predicción						
	Tiempo requerido para procesar data						
Tiempo	Tiempo requerido para predecir el diagnóstico clínico		Segundos	< 5s			Cronómetro
Toma de decisiones	Nivel de confianza en la toma de decisiones	Ordinal	Escala Likert	[Totalmente conforme, Bastante conforme, Moderadamente conforme, Poco conforme, Totalmente disconforme]	Encuesta	Cuestionario	Análisis de encuesta

2.5. Técnicas e instrumentos de recolección de datos, validez y confiabilidad

Técnicas e instrumentos de la investigación

En el contexto de este estudio actual se consideró utilizar las técnicas de observación directa y encuesta. Por lo que los instrumentos usados son ficha de registro y cuestionario respectivamente.

Tabla 8

Técnicas e instrumentos de recolección de información

Técnicas	Uso	Instrumentos
Observación directa <ul style="list-style-type: none"> No participante 	<ul style="list-style-type: none"> Se observarán los valores en porcentajes (%) que brindan las métricas ML en los modelos. Se observará el tiempo en segundos que se toma el procesamiento de data. Se observará el tiempo en segundos que toma la predicción de resultados clínicos 	Ficha de registro Anexo 08, 09, 10 Ficha de registro Anexo 11 Ficha de registro Anexo 12
Encuesta <ul style="list-style-type: none"> Cerrado 	Se realizará un cuestionario para ver el nivel de confianza en la toma de decisiones del personal del laboratorio clínico	Cuestionario Anexo 13

Técnicas e instrumentos de la investigación documental

Tabla 9

Técnicas e instrumentos de la investigación documental

Técnicas	Instrumentos
Revisión de <ul style="list-style-type: none"> Tesis Internet Revistas 	<ul style="list-style-type: none"> Fichas Computador Libreta de apuntes

Validez y confiabilidad del instrumento

La fiabilidad de la herramienta utilizada fue corroborada por un panel de tres especialistas, que analizaron su pertinencia, relevancia y claridad, concluyendo su adecuación para la aplicación.

Tabla 10

Validez de los instrumentos por expertos

	Experto 1	Experto 2	Experto 3
Nombre	Jose Carlos Anicama Silva	Ivan Carlo Azabache Petrlik	Carlos Alberto Lon Kan
Grado	Maestría	Doctorado	Maestría
Especialidad	Temático	Temático	Temático
Institución	Universidad Autónoma del Perú	Universidad Nacional Federico Villarreal	Universidad Autónoma del Perú

La fiabilidad del instrumento se estableció utilizando el método Test-Retest junto al índice de fiabilidad Alfa de Cronbach, aplicados sobre la observación de la pre-prueba.

Tabla 11

Correlaciones para el Indicador KPI 4: Tiempo de procesamiento de data

		TP Fecha 1	TP Fecha 2
	Correlación de Pearson:	1,000	,755**
TP Fecha 1	Significación (bilateral):	-	,000
	N	45	45
	Correlación de Pearson:	,755**	1
TP Fecha 2	Significación (bilateral):	,000	-
	N	45	45

Nota. Elaborado con el Software SPSS. TP = Tiempo de Procesamiento.

** La correlación es significativa en el nivel 0,01 (bilateral).

En este estudio se empleó la metodología Test-Retest, utilizando la correlación de Pearson para evaluar la consistencia. Los resultados mostraron una correlación superior al 70%, lo que indica una alta interrelación entre los datos.

Para el indicador KPI 6: Toma de decisiones

A fin de determinar la consistencia de los datos, se aplicó el coeficiente Alfa de Cronbach. Los hallazgos mostraron un valor que excede el 90%, implicando así una calidad de datos excepcional.

Tabla 12

Estadísticas de fiabilidad

Alfa de Cronbach	N de elementos
,901	10

Nota. Estadísticas de fiabilidad, SPSS

Como el alfa de Cronbach supero el 90% de confiabilidad, esto refleja una alta consistencia interna en el cuestionario utilizado, es decir sus ítems están correlacionados entre si y miden de manera coherente el mismo constructo.

Tabla 13

Estadísticas de total de elemento

Elemento	M (sup.)	V (sup.)	CC	AC (sup.)
VAR00001	43,50	7,500	0,984	0,870
VAR00002	43,50	7,500	0,984	0,870
VAR00003	43,33	9,867	0,000	0,912
VAR00004	43,50	7,500	0,984	0,870
VAR00005	43,83	5,767	0,846	0,897
VAR00006	43,50	10,300	-0,229	0,939
VAR00007	43,50	7,500	0,984	0,870
VAR00008	43,33	9,867	0,000	0,912
VAR00009	43,50	7,500	0,984	0,870
VAR00010	43,50	7,500	0,984	0,870

Nota. La tabla presenta estadísticas relacionadas con la consistencia interna de los elementos. M (sup.) corresponde a la media de la escala si el elemento se ha suprimido, lo que indica cómo afecta la eliminación del elemento al promedio general de la escala. V (sup.) representa la varianza de la escala si el elemento se ha suprimido, mostrando la dispersión de los valores restantes. CC se refiere a la correlación total corregida, que mide la relación entre el elemento y el resto de la escala, excluyendo al elemento evaluado. AC (sup.) es el Alfa de Cronbach si el elemento se ha suprimido, mostrando el impacto en la confiabilidad de la escala al eliminar el elemento.

2.6. Procedimientos

Este estudio tiene como propósito aplicar el Aprendizaje Automático en la predicción de diagnósticos clínicos, tomando como referencia los pacientes que proporcionan sus muestras para análisis en laboratorios clínicos, culminando con la deliberación de resultados. En este marco, el "análisis predictivo de resultados clínicos" se considera como la variable dependiente. Para alcanzar el objetivo, se realizó un meticuloso repaso de investigaciones análogas a nivel local y global, con el propósito de examinar las soluciones previamente formuladas. Durante este procedimiento, se realizó una evaluación detallada y completa de la variable dependiente y la independiente. Se recopiló información de tesis y publicaciones científicas para establecer una base de antecedentes y referencias teóricas. A partir de este conjunto de datos, se identificaron dimensiones e indicadores relacionados con la sostenibilidad. En la ejecución de este análisis investigador, se consideraron datos obtenidos de pacientes, donde se incluye información relacionada con los exámenes de laboratorio llevados a cabo y los logros obtenidos en los mismos. El área administrativa del laboratorio es la encargada de proporcionar la data de pacientes. Posteriormente, se adoptó la metodología CRISP-ML(Q), la cual es un enfoque estructurado para estandarizar el ciclo de vida del aprendizaje automático en diferentes industrias, con un énfasis en la garantía de calidad. Esta metodología se divide en seis fases: entendimiento del negocio y de los datos, preparación de los datos, construcción y ajuste del modelo, evaluación, despliegue del modelo, y monitoreo y mantenimiento. En este proyecto específico, el enfoque CRISP-ML(Q) facilitó el reconocimiento y la comprensión de las tendencias y conexiones presentes en la información de los pacientes, un aspecto crucial para el desarrollo y mejora de los modelos de Machine Learning. Estos modelos fueron entrenados y validados con

los datos proporcionados, y se seleccionó el que proporcionaba los resultados más precisos y coherentes para el objetivo de la investigación.

En este contexto, se implementaron variadas estructuras de Machine Learning, tales como Gradient Boosting, Decision Tree, Logistic Regression, Random Forest, con fines de realizar el análisis comparativo e implementar el modelo que tiene mejor rendimiento.

Para el registro de valores en métricas de exactitud, precisión y sensibilidad se consideran fichas de registro (Anexo 8, 9 y 10), mientras que para medir el tiempo de predicción se consideran también fichas (Anexo 11 y 12).

A fin de verificar las hipótesis planteadas en este estudio, se empleó la matriz de confusión suministrada por la biblioteca scikit-learn de Python, lo que facilitó la valoración de la efectividad del algoritmo. Además, se empleó el módulo metrics de scikit-learn para calcular el coeficiente de Kappa de Cohen, permitiendo verificar el grado de acuerdo y la significancia estadística.

Tabla 14

Pruebas médicas de laboratorio del dataset estudiado

ID	Campo del dataset	Descripción
001	VB12 II	Relacionado con la vitamina B12
002	Anti-TPO	Anticuerpos contra la peroxidasa tiroidea, asociados con enfermedades tiroideas autoinmunes
003	Anti-Tg	Anticuerpos contra la tiroglobulina
004	CEA	Antígeno carcinoembrionario, utilizado como marcador tumoral.
005	CA19-9	Un marcador tumoral asociado principalmente con el cáncer de páncreas
006	CA125	Marcador tumoral asociado con el cáncer de ovario
007	AFP	Alfa-fetoproteína, un marcador tumoral

008	25-OH VitD	25-hidroxivitamina D, un marcador de la vitamina D en el cuerpo
009	CA15-3	Marcador tumoral utilizado en el seguimiento del cáncer de mama
010	BNP	Péptido natriurético cerebral, relacionado con insuficiencia cardíaca
011	Anti-TP	Anticuerpos contra el treponema pallidum, agente causante de la sífilis
012	Troponin-I	Una proteína del músculo cardíaco, indicador de daño cardíaco
013	HIV Combo	Prueba combinada para la detección del VIH
014	HBsAg	Antígeno de superficie del virus de la hepatitis B
015	ANA	Anticuerpos antinucleares, asociados con enfermedades autoinmunitarias
016	Anti-HBc	Anticuerpos contra el núcleo del virus de la hepatitis B
017	Anti-HCV	Anticuerpos contra el virus de la hepatitis C
018	Anti-CCP	Anticuerpos contra el péptido cíclico citrulinado, relacionados con la artritis reumatoide
019	E2	Estradiol, una hormona esteroide sexual femenina
020	ACTH	Hormona adrenocorticotrópica, que estimula la producción de cortisol
021	Free PSA	Una forma de antígeno prostático específico, útil en el diagnóstico de problemas de próstata
022	TPSA	Área de superficie polar total, común en química
023	Ferritin	Una proteína que almacena hierro en el cuerpo
024	Testosterone	Testosterona, una hormona sexual masculina
025	HCG	Gonadotropina coriónica humana, asociada con el embarazo
026	Prolactin	Prolactina, una hormona producida por la hipófisis

La información incluida en el dataset del laboratorio clínico será validada, luego codificada, seguidamente se registrará y por último se tabulará para el análisis de los datos, haciendo uso de Python, Jupyter Notebooky Google Colab. Los resultados de este análisis serán representados a través de gráficos y tablas estadísticas, así como la interpretación de los datos tomando como base los indicadores y variables que se medirán, por lo cual ello se apoya en el software estadístico SPSS.

Cada modelo de Machine Learning tiene su desempeño característico, el cual se determina en función de diversas métricas de rendimiento. Como tal, es esencial documentar y rastrear estos valores de manera sistemática. Una vez evaluado el rendimiento de cada modelo se procederá a seleccionar al que ofrece mayor desempeño. Por lo tanto, las fichas de registro (Anexo 9, 10 y 11) cobran gran relevancia en este contexto. Estas fichas no solo permiten realizar un seguimiento del rendimiento del modelo de Machine Learning, sino que también proporciona una visión clara y comparativa del desempeño relativo de varios modelos en diferentes conjuntos de datos o tareas. Al proporcionar esta visión integral, las fichas de registro facilitan la identificación del modelo que ofrece el mejor desempeño en un escenario específico.

Además, esta documentación metódica del desempeño del modelo puede ser crucial para comprender cómo las modificaciones en los datos de entrada, los parámetros del modelo o las características seleccionadas pueden influir en los resultados. Por último, puede proporcionar valiosos insights para mejorar el rendimiento de los modelos en futuras iteraciones del proceso de Machine Learning, haciendo posible ajustes y optimizaciones basados en evidencia práctica.

La métrica Accuracy (exactitud) se determina dividiendo el número de predicciones acertadas entre el total de predicciones efectuadas, proporcionando así

el porcentaje de predicciones correctas. Nasim et al. (2022), en *A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics*, sostienen que este índice de exactitud ilustra la capacidad predictiva del modelo. De manera similar, Jia et al. (2022), en *PE_DIM: An Efficient Probabilistic Ensemble Classification Algorithm for Diabetes Handling Class Imbalance Missing Values*, destacan que la métrica de accuracy indica la proporción de casos correctamente categorizados en comparación con la cantidad total de muestras examinadas.

$$\text{Accuracy} = \left(\frac{TP+TN}{TP+TN+FP+FN} \right) \times 100 \quad (1)$$

La métrica de precisión, según Nasim et al. (2022), se define como la proporción de predicciones positivas correctas, representada por los verdaderos positivos, en relación con la suma de verdaderos positivos y falsos positivos. Este indicador evalúa la eficacia del modelo al generar predicciones positivas que son realmente exactas, destacando su capacidad para minimizar los falsos positivos.

$$\text{Precision (\%)} = \left(\frac{TP}{TP+FP} \right) \times 100 \quad (2)$$

La sensibilidad, también conocida como tasa de verdaderos positivos o recall, se calcula como la fracción de verdaderos positivos dividida entre la suma de verdaderos positivos y falsos negativos. Esta métrica tiene como objetivo garantizar el reconocimiento integral de los casos relevantes dentro de un conjunto de datos. Nasim et al. (2022), en su investigación titulada *A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics*, destacan que el recall mide la capacidad del modelo para detectar y clasificar correctamente los casos positivos, lo que resulta crucial en contextos donde los falsos negativos tienen un impacto significativo.

$$\text{Sensitivity (\%)} = \left(\frac{TP}{TP+FN} \right) \times 100 \quad (3)$$

Para evaluar la hipótesis general de la investigación, se establecen 6 Key Performance Indicators (KPIs) que evalúan diversas facetas relacionadas con la aplicación del modelo de aprendizaje automático: la exactitud de la predicción, la precisión de la predicción, la sensibilidad de la predicción, el tiempo requerido para el procesamiento de data, el tiempo requerido para la predicción clínica y el nivel de confianza en la toma de decisiones.

Procedimientos de análisis

Los procedimientos de análisis serán realizados en:

- El software Python, por ser open source y notablemente versátil para la manipulación y transformación de estructuras de datos, así como para el uso de varias medidas asociadas al aprendizaje automático. Su amplia gama de bibliotecas especializadas facilita la implementación de algoritmos avanzados, y su flexibilidad permite adaptarse a una amplia variedad de tareas de análisis.
- SPSS, por ser un software de análisis estadístico ampliamente utilizado y efectivo. Gracias a su habilidad para procesar grandes cantidades de datos, se emplea para la creación de análisis descriptivos e inferenciales, siendo especialmente beneficiosa en la ejecución de pruebas de hipótesis, análisis de varianza y regresión. Su interfaz gráfica es intuitiva y sus herramientas de análisis son precisas, lo que convierte a SPSS en un recurso valioso para nuestra investigación, la cual necesita un análisis minucioso y riguroso de los datos. Adicionalmente, SPSS facilita la visualización de datos y resultados, lo cual mejora la interpretación y comprensión de estos.

2.7. Análisis de datos

Software estadístico

Para la gestión y evaluación de datos en nuestro estudio, se utilizará SPSS versión 28, una aplicación informática destinada a la evaluación estadística altamente eficaz y ampliamente utilizado. Este programa será empleado para la gestión y manipulación de grandes volúmenes de datos, con el objetivo de identificar patrones, tendencias y correlaciones dentro de nuestro conjunto de información.

SPSS es especialmente beneficioso en la ejecución de análisis de hipótesis, análisis de varianza y regresión. Su interfaz gráfica es intuitiva y sus herramientas de análisis son precisas, convirtiendo a SPSS en un recurso valioso para nuestra investigación. Tras el procesamiento de los datos y la obtención de los resultados preliminares, el uso de SPSS permitirá realizar un análisis detallado y la interpretación de dichos resultados, permitiéndonos hacer inferencias válidas y significativas. Al final, estos análisis nos permitirán llegar a conclusiones útiles y orientadas a la toma de decisiones basadas en evidencia.

Etapas del análisis

Tabla 15

Etapas del análisis de datos en SPSS

Fase	Descripción
1.- Selección del software	- SPSS v. 26®
2.- Ejecución de programas	
3.- Exploración de los datos	- Distribución de frecuencias gráficas - Las medidas de tendencia central - Medidas de la variabilidad - Otras estadísticas descriptivas
4.- Análisis mediante pruebas estadísticas	- Análisis estadístico inferencial
5.- Análisis adicionales	

6.- Preparación de los resultados

- Tablas, gráficos, figuras, cuadros, etc.

2.8. Aspectos éticos

Criterios éticos

La ejecución de este estudio se llevó a cabo con un enfoque en criterios éticos estrictamente alineados con las directrices establecidas en el Informe Belmont. Estos principios son de suma importancia ya que proporcionan un marco para el comportamiento ético en la realización de investigaciones.

Según lo que comenta la autora Fernández (2020) el informe Belmont es un conjunto de "Principios éticos y pautas para la protección de los seres humanos en la investigación" publicado por la Comisión Nacional para la Protección de los Sujetos Humanos ante la Investigación Biomédica y de Comportamiento de Estados Unidos.

En consecuencia, el estudio actual tiene su fundamento en:

- **El respeto a las personas:** Aseguramos este cumplimiento proporcionando las citas adecuadas a las investigaciones que han servido como fundamentos para el desarrollo tanto de las contribuciones teóricas como prácticas.
- **Beneficencia:** Los avances logrados en este estudio proveen ventajas para toda la población que utiliza tecnologías de la información y se encuentra vinculada a desafíos de salud que necesitan del análisis de muestras.

Criterios de rigor científico

En el marco de este estudio actual se implementó una serie de procesos para estructurar las contribuciones teóricas y prácticas, basándose en los siguientes criterios científicos:

- **Credibilidad:** Se realizó una evaluación de técnicas de machine learning, y las conclusiones obtenidas se integraron en el modelo de machine learning diseñado especialmente para predecir resultados en el entorno clínico.
- **Objetividad:** El análisis predictivo de resultados clínicos que se pretende dar a conocer se fundamenta en investigaciones previas, parámetros técnicos y neutrales.
- **Originalidad:** Se incluyen las citas bibliográficas con el propósito de respaldar la originalidad del contenido.
- **Veracidad:** La información empleada en este estudio es auténtica y se preserva la confidencialidad de los datos.
- **Relevancia:** Los datos utilizados en esta investigación son genuinos y se mantiene la privacidad de la información.

CAPÍTULO III
RESULTADOS

Pruebas de normalidad

Los resultados obtenidos se presentan a continuación para cada uno de los Indicadores Clave de Desempeño (KPIs). Estos fueron analizados durante las etapas de pre-evaluación y post-evaluación, enfocándose especialmente en la normalidad estadística de la información.

Prueba de normalidad del KPI 1: Tasa de exactitud

Plantear la hipótesis

Ho: Los datos siguen una distribución normal.

Ha: Los datos no siguen una distribución normal

Criterio de decisión

Si $p < 0,05$ rechazamos la Ho y aceptamos la Ha.

Si $p \geq 0,05$ aceptamos la Ho y rechazamos la Ha

Tabla 16

Pruebas de normalidad KPI 1

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP1_Pre	,139	26	,200*	,925	26	,058
KP1_Post	,477	26	,000	,504	26	,000

Nota. Elaborado con el Software SPSS Versión 28.

*. Esto es un límite inferior de la significación verdadera

a. Corrección de significación de Lilliefors

Decisión y conclusión

En el análisis de pre-prueba, dado que $p=0,058$ es mayor que 0,05, Se concluye que los datos siguen una distribución normal, lo cual acepta la hipótesis nula (Ho). En la post-prueba, con un valor de $p=0,000$, que es menor que 0,05, se rechaza Ho, indicando que los datos no se distribuyen normalmente. Por lo tanto, para el KPI 1 relacionado con la tasa de exactitud, se optó por emplear métodos estadísticos no paramétricos.

Prueba de normalidad del KPI 2: Tasa de precisión

Plantear la hipótesis

Ho: Los datos tienen una distribución normal

Ha: Los datos no tienen una distribución normal

Criterio de decisión

Si $p < 0,05$ rechazamos la Ho y aceptamos la Ha.

Si $p \geq 0,05$ aceptamos la Ho y rechazamos la Ha

Tabla 17

Pruebas de normalidad KPI 2

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP2_Pre	,118	26	,200*	,942	26	,148
KP2_Post	,483	26	,000	,520	26	,000

Nota. Elaborado con el Software SPSS Versión 28.

*. Esto es un límite inferior de la significación verdadera

a. Corrección de significación de Lilliefors

Decisión y conclusión

Dado que en la etapa de pre-prueba el valor de p es 0,148, superando el umbral de 0,05, se acepta la hipótesis nula (Ho), lo cual indica que los datos están en concordancia con una distribución normal. En contraste, en la etapa de post-prueba, con un valor de p de 0,000, menor a 0,05, se rechaza Ho, indicando una distribución no normal de los datos. Por lo tanto, se decidió utilizar métodos estadísticos no paramétricos en el análisis de estos datos.

Prueba de normalidad del KPI 3: Tasa de sensibilidad

Plantear la hipótesis

Ho: Los datos tienen una distribución normal

Ha: Los datos no tienen una distribución normal

Criterio de decisión

Si $p < 0,05$ rechazamos la H_0 y aceptamos la H_a .

Si $p \geq 0,05$ aceptamos la H_0 y rechazamos la H_a

Tabla 18

Pruebas de normalidad KPI 3

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP3_Pre	,149	26	,145	,898	26	,014
KP3_Post	,483	26	,000	,520	26	,000

Nota. Elaborado con el Software SPSS Versión 28.

a. Corrección de significación de Lilliefors

Decisión y conclusión

Durante la fase de pre-prueba, el valor de p fue de 0,014, que está por debajo del umbral de significancia establecido en 0,05, llevando a un rechazo de la hipótesis nula (H_0). Este resultado indica que la distribución de los datos no sigue un patrón normal, lo que sugiere posibles asimetrías o variabilidad en los datos. De manera similar, en la fase de post-prueba, se observó un valor de p de 0,000, también inferior a 0,05, lo cual reafirma el rechazo de la H_0 y confirma que la distribución sigue siendo no normal, tal como ocurrió en la fase de pre-prueba. Ante esta situación y para garantizar la validez de los análisis, se implementaron métodos de estadística no paramétrica, los cuales son más adecuados para trabajar con datos que no cumplen con los supuestos de normalidad.

Prueba de normalidad del KPI 4: Tiempo de procesamiento de datos

Plantear la hipótesis

H_0 : Los datos tienen una distribución normal

H_a : Los datos no tienen una distribución normal

Criterio de decisión

Si $p < 0,05$ rechazamos la H_0 y aceptamos la H_a .

Si $p \geq 0,05$ aceptamos la H_0 y rechazamos la H_a

Tabla 19*Pruebas de normalidad KPI 4*

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP4_Pre	,128	26	,200*	,931	26	,080
KP4_Post	,365	26	,000	,655	26	,000

Nota. Elaborado con el Software SPSS Versión 28.

*. Esto es un límite inferior de la significación verdadera

a. Corrección de significación de Lilliefors

Decisión y conclusión

En la etapa de pre-prueba, el valor de p es 0,080, mayor que 0,05, lo que lleva a aceptar la hipótesis nula (H_0), indicando una distribución normal de los datos. En contraste, en la etapa de post-prueba, el valor de p es 0,000, menor que 0,05, lo que lleva al rechazo de H_0 y sugiere una distribución no normal de los datos. Por lo tanto, se optó por utilizar técnicas de estadística no paramétrica.

Prueba de normalidad del KPI 5: Tiempo de predicción**Plantear la hipótesis**

H_0 : Los datos tienen una distribución normal

H_a : Los datos no tienen una distribución normal

Criterio de decisión

Si $p < 0,05$ rechazamos la H_0 y aceptamos la H_a .

Si $p \geq 0,05$ aceptamos la H_0 y rechazamos la H_a

Tabla 20*Pruebas de normalidad KPI 5*

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP5_Pre	,095	26	,200*	,958	26	,347
KP5_Post	,354	26	,000	,801	26	,000

Nota. Elaborado con el Software SPSS Versión 28.

*. Esto es un límite inferior de la significación verdadera

a. Corrección de significación de Lilliefors

Decisión y conclusión

Durante la fase de pre-prueba, se observó un valor de p de 0,347, que excede el umbral de 0,05, lo que llevó a la aceptación de la hipótesis nula (H_0) y sugirió una distribución normal de los datos. Sin embargo, en la fase de post-prueba, el valor de p fue de 0,000, inferior a 0,05, lo que indica el rechazo de H_0 y, por lo tanto, una distribución no normal de los datos. Por consiguiente, se decidió utilizar métodos estadísticos no paramétricos.

Prueba de normalidad del KPI 6: Nivel de confianza en la toma de decisiones

Plantear la hipótesis

H_0 : Los datos tienen una distribución normal

H_a : Los datos no tienen una distribución normal

Criterio de decisión

Si $p < 0,05$ rechazamos la H_0 y aceptamos la H_a .

Si $p \geq 0,05$ aceptamos la H_0 y rechazamos la H_a

Tabla 21

Pruebas de normalidad KPI 6

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP6_Pre	,199	16	,092	,934	16	,282
KP6_Post	,346	16	,000	,657	16	,000

Nota. Elaborado con el Software SPSS Versión 28.

a. Corrección de significación de Lilliefors

Decisión y conclusión

En la fase preliminar, con un valor de p de 0,282, que es superior al umbral de significancia de 0,05, se procede a aceptar la hipótesis nula (H_0), lo cual indica que los datos siguen una distribución normal, cumpliendo con los supuestos estadísticos necesarios para este tipo de análisis. Por otro lado, en la fase posterior, el valor de p es de 0,000, claramente menor que 0,05, lo que resulta en un rechazo

de la H_0 y confirma la no normalidad de la distribución de los datos en esta etapa. Este cambio en la distribución de los datos llevó a la decisión de emplear técnicas de estadística no paramétrica, consideradas más adecuadas para analizar información que no cumple con los criterios de normalidad.

Estadística descriptiva

Estadística descriptiva del KPI 1

Los resultados alcanzados del KPI 1 se presentan a continuación, mostrando la tasa de exactitud de predicción obtenida tanto en la fase de pre-prueba como en la fase de post-prueba.

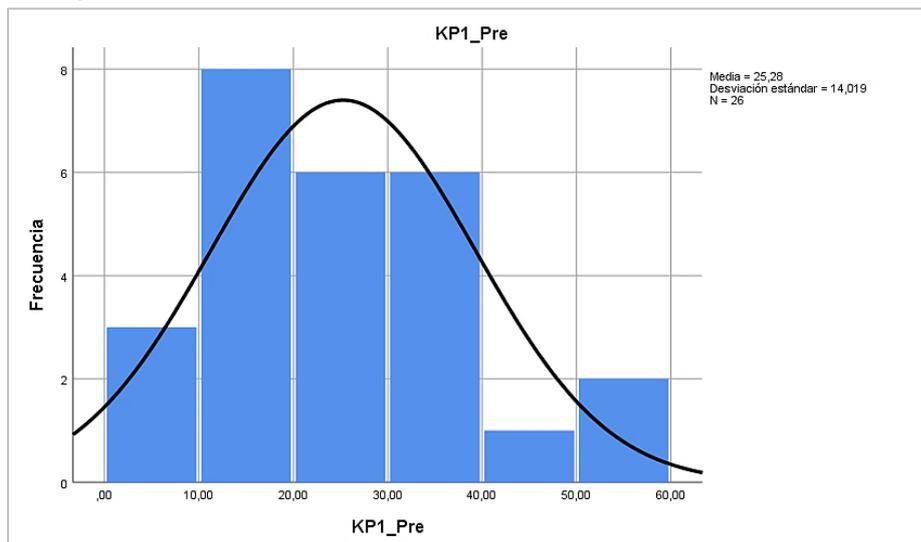
Tabla 22

Análisis de la tasa de exactitud

		KP1_Pre	KP1_Post
N	Válido	26	26
	Perdidos	0	0
Media		25,2758	99,5000
Mediana		21,6450	100,0000
Moda		7,26 ^a	100,00
Desviación estándar		14,01917	1,14018
Varianza		196,537	1,300
Asimetría		,925	-2,193
Error estándar de asimetría		,456	,456
Curtosis		,380	3,666
Error estándar de curtosis		,887	,887
Rango		52,69	4,00
Mínimo		7,26	96,00
Máximo		59,95	100,00
Suma		657,17	2587,00
Percentiles	25	15,0425	100,0000
	50	21,6450	100,0000
	75	34,6550	100,0000

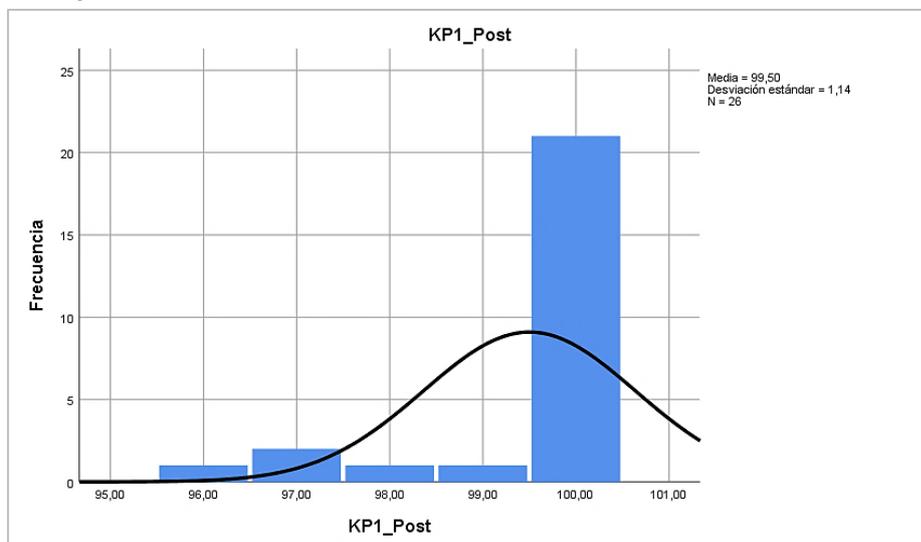
Nota. Elaborado con el Software SPSS Versión 28.

a. Existen múltiples modos. Se muestra el valor más pequeño.

Figura 14*Histograma pre KP1: Tasa de exactitud*

Nota. Elaborado con el Software SPSS Versión 28.

La Figura 14 ilustra la distribución de la tasa de exactitud inicial (KP1_Pre) antes de la intervención, evidenciando que la mayoría de las mediciones se concentran en valores bajos, con una mayor frecuencia entre 10 y 30. La curva de distribución superpuesta refleja una asimetría positiva, consistente con la variabilidad observada en las métricas previas. Esto respalda la interpretación de un nivel inicial de precisión limitado, caracterizado por una dispersión considerable de los datos.

Figura 15*Histograma post KPI 1: Tasa de exactitud*

Nota. Elaborado con el Software SPSS Versión 28.

La Figura 15 presenta un histograma que muestra la distribución de la tasa de exactitud posterior a la intervención (KP1_Post), evidenciando un cambio significativo en los resultados. La mayoría de las mediciones se concentran en el valor máximo de exactitud (100), con una curva de distribución superpuesta que confirma una agrupación casi perfecta de los datos. Con una media de 99,50 y una desviación estándar de 1,14, se observa una notable reducción en la variabilidad en comparación con la medición previa. Esto indica que la intervención logró estandarizar las tasas de exactitud, alcanzando niveles consistentemente altos en prácticamente todas las observaciones.

El análisis del KPI1, que evalúa la exactitud antes y después de una intervención específica, revela una mejora notable en todas las métricas estadísticas. Originalmente, la media de exactitud antes de la intervención era de 25,2758, lo que indica un nivel de precisión inicial relativamente bajo. Tras la intervención, esta media aumentó significativamente a 99,5000, demostrando una mejora considerable en la exactitud.

Además, la mediana subió de 21,6450 a 100,0000. Esto confirma que, después de la intervención, al menos la mitad de las mediciones alcanzaron la tasa de exactitud máxima. Antes de la intervención, la moda era de 7,26, lo que reflejaba una prevalencia de tasas de exactitud más bajas. Sin embargo, en la post-prueba, la moda se estableció firmemente en 100,00, lo que indica que la mayoría de los resultados después de la intervención fueron óptimos.

Respecto a la dispersión de los datos, se consideran tanto la desviación estándar como la varianza, reduciéndose dramáticamente, pasando de 14,01917 a 1,14018 y de 196,537 a 1,300, respectivamente. Esto indica una mayor uniformidad en los resultados después de la intervención. La asimetría de los datos cambió de una

ligera inclinación positiva (0,925) a una negatividad marcada (-2,193). Además, la curtosis aumentó de 0,380 a 3,666 en la post-prueba, sugiriendo una agrupación más fuerte de los resultados en torno a la alta tasa de exactitud.

El rango de los datos también muestra esta mejora, disminuyendo de 52,69 en la pre-prueba a solo 4,00 en la post-prueba, lo que destaca una notable disminución en la variabilidad de las tasas de exactitud. El valor mínimo mejoró de 7,26 a 96,00, y el valor máximo se mantuvo constante en 100,00, asegurando que incluso el desempeño más bajo estuvo muy cerca del nivel óptimo.

Por último, los percentiles 25, 50 (que corresponde a la mediana) y 75, que antes mostraban una mayor dispersión en las tasas de exactitud, se alinearon todos en 100,0000 en la post-prueba. Esto demuestra que, como mínimo, el 75% de las mediciones de exactitud alcanzaron el nivel más alto posible después de la intervención.

Estos resultados subrayan el impacto significativo de la intervención en la mejora de la tasa de exactitud, evidenciado no solo por el aumento de la media y la mediana hacia resultados cercanos al máximo, sino también por la notable reducción de la variabilidad y el considerable aumento en el rendimiento mínimo.

Estadística descriptiva del KPI 2

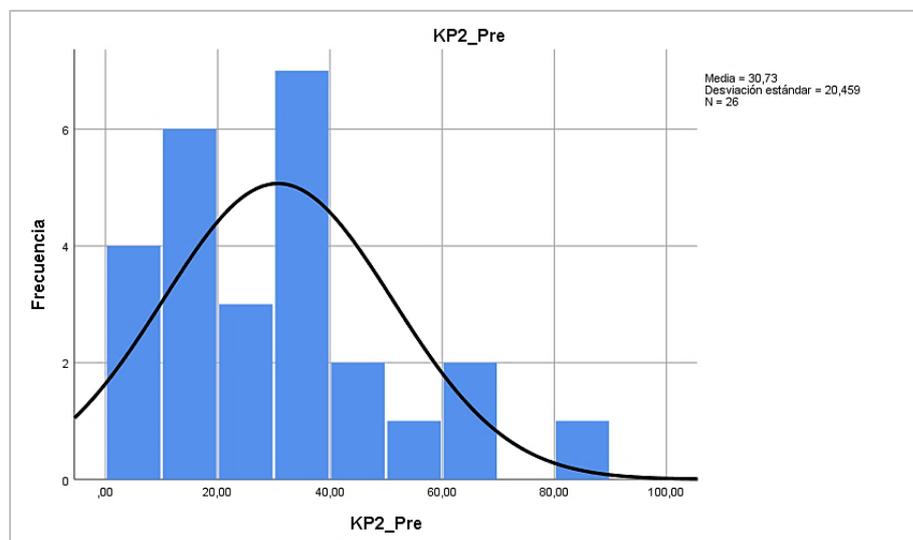
Los resultados del Indicador Clave de Rendimiento (KPI) 2, que mide la tasa de precisión en las predicciones realizadas por el modelo, se exponen a continuación tanto para la pre-prueba como para la post-prueba, permitiendo una comparación clara y cuantitativa del rendimiento antes y después de la implementación de la solución propuesta. Este análisis proporciona una visión más detallada y objetiva sobre la capacidad del modelo para generar predicciones precisas en ambos escenarios, destacando las mejoras alcanzadas tras la intervención.

Tabla 23*Análisis de la tasa de precisión*

		KP2_Pre	KP2_Post
N	Válido	26	26
	Perdidos	0	0
Media		30,7300	99,3462
Mediana		29,8450	100,0000
Moda		3,55 ^a	100,00
Desviación estándar		20,45878	1,44062
Varianza		418,562	2,075
Asimetría		,772	-2,057
Error estándar de asimetría		,456	,456
Curtosis		,258	3,098
Error estándar de curtosis		,887	,887
Rango		78,57	5,00
Mínimo		3,55	95,00
Máximo		82,12	100,00
Suma		798,98	2583,00
Percentiles	25	13,7600	100,0000
	50	29,8450	100,0000
	75	40,8425	100,0000

Nota. Elaborado con el Software SPSS Versión 28.

a. Existen múltiples modos. Se muestra el valor más pequeño.

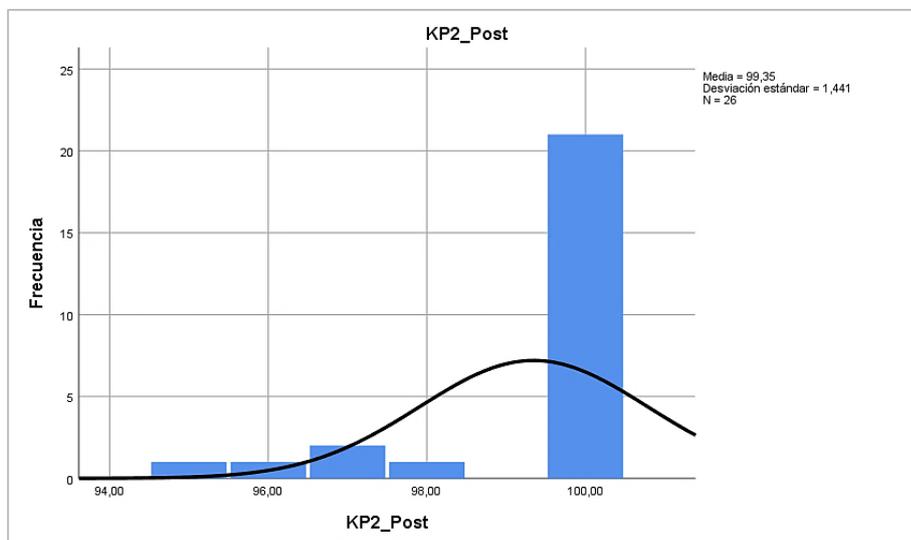
Figura 16*Histograma pre KP2: Tasa de precisión*

Nota. Elaborado con el Software SPSS Versión 28.

La Figura 16 muestra un histograma que representa la distribución de la tasa de precisión del modelo antes de la intervención (KP2_Pre). Los valores se distribuyen en un rango amplio, con una concentración mayoritaria entre 20 y 40, lo que refleja un rendimiento inicial moderado. La curva de distribución superpuesta sugiere una ligera asimetría positiva, indicando que algunos valores superaron los niveles promedio de precisión, aunque de forma menos frecuente. Con una media de 30,73 y una desviación estándar de 20,46, los datos evidencian una alta dispersión y variabilidad en las mediciones iniciales, lo que resalta la necesidad de optimización del modelo en este escenario previo a la implementación de la solución.

Figura 17

Histograma post KP2: Tasa de precisión



Nota. Elaborado con el Software SPSS Versión 28.

La Figura 17 presenta un histograma que refleja la distribución de la tasa de precisión del modelo después de la intervención (KP2_Post), evidenciando una mejora significativa en los resultados. La mayoría de las mediciones se concentran en el valor máximo de 100, lo que indica que el modelo alcanzó un rendimiento óptimo en la mayoría de los casos. La curva de distribución superpuesta confirma esta tendencia, con una media de 99,35 y una desviación estándar de 1,44, lo que muestra

una notable reducción en la variabilidad en comparación con la pre-prueba. Estos resultados destacan la efectividad de la intervención para optimizar la capacidad del modelo en la generación de predicciones precisas y consistentes.

Al examinar el KPI2, que mide la tasa de precisión antes y después de una intervención esencial, se observan cambios notables en todas las métricas estadísticas. Inicialmente, en la fase pre-intervención, la media de precisión registrada fue de 30,7300, proporcionando una base para entender el nivel de precisión inicial. Sin embargo, tras la intervención, esta media experimentó un salto considerable, alcanzando 99,3462 en la fase post-intervención. Este notable aumento refleja una mejora significativa hacia una tasa de precisión casi óptima.

Además, la mediana, que es el valor medio de la tasa de precisión, también mostró un incremento sustancial. Pasó de 29,8450 en la Pre-prueba a un valor ideal de 100,0000 en la Post-prueba. Este cambio indica que, después de la intervención, más de la mitad de las mediciones alcanzaron la perfección en términos de precisión.

La moda en la fase pre-intervención fue de 3,55, lo que podría indicar una mayor frecuencia de tasas de precisión inferiores. En cambio, en la fase post-intervención, la moda se situó en 100,00, demostrando que la mayoría de los resultados después de la intervención alcanzaron el nivel máximo.

En cuanto a la dispersión de los datos, observamos una disminución drástica tanto en la desviación estándar como en la varianza, pasando de 20,45878 a 1,44062 y de 418,562 a 2,075, respectivamente. Esto señala una gran reducción en la variabilidad de los datos y una mayor consistencia en la tasa de precisión post-intervención.

La asimetría de los datos disminuyó ligeramente de 0,772 a -2,057, sugiriendo un cambio en la distribución hacia una agrupación en el extremo superior de la escala.

Además, la curtosis aumentó de 0,258 a 3,098, lo que implica una distribución más puntiaguda en la Post-prueba con una concentración de valores alrededor del máximo.

El rango de la tasa de precisión también refleja esta mejora, disminuyendo significativamente de 78,57 en la pre-prueba a solo 5,00 en la post-prueba, lo que enfatiza la reducción de la variabilidad. Los valores mínimo y máximo mejoraron notablemente, con el mínimo subiendo de 3,55 a 95,00 y el máximo manteniéndose en 100,00, asegurando que incluso el desempeño más bajo estuvo muy cerca del nivel óptimo en la post-prueba.

Finalmente, los percentiles 25, 50 (que es la mediana) y 75, que anteriormente reflejaban una mayor dispersión en las tasas de precisión, se alinearon todos en 100,0000 en la Post-prueba. Esto indica que al menos el 75% de las mediciones de precisión alcanzaron el nivel más alto posible después de la intervención.

Estos resultados subrayan el impacto significativo de la intervención en el KPI2, evidenciando no solo un aumento en la media y la mediana hacia niveles casi óptimos de precisión, sino también una marcada reducción en la variabilidad y una mejora sustancial en el rendimiento mínimo.

Los percentiles reflejan esta mejora consistente y significativa; mientras que en la pre-prueba había una mayor dispersión con el 25% de las mediciones igual o inferior a 13,7600 y el 75% igual o inferior a 40,8425, en la post-prueba todos los cuartiles alcanzaron el valor máximo de 100,0000, lo que indica que al menos el 75% de las observaciones lograron una precisión óptima y uniforme, eliminando prácticamente cualquier variabilidad en los resultados superiores.

Estos datos estadísticos, en su totalidad, evidencian un aumento considerable y significativo en la tasa de precisión debido a la intervención realizada, mostrando un

incremento notable y sostenido en la media y la mediana hacia resultados casi óptimos, una disminución marcada en la variabilidad que refleja mayor consistencia en los datos, y una mejora sustancial en el rendimiento mínimo, acercándolo de manera destacada al nivel máximo posible.

Estadística descriptiva del KPI 3

Los resultados del Indicador Clave de Rendimiento (KPI) 3, que evalúa la tasa de sensibilidad en las predicciones, se presentan a continuación, tanto para la fase pre-prueba como para la post-prueba, permitiendo analizar en detalle cómo la intervención impactó en la capacidad del modelo para identificar correctamente los casos positivos.

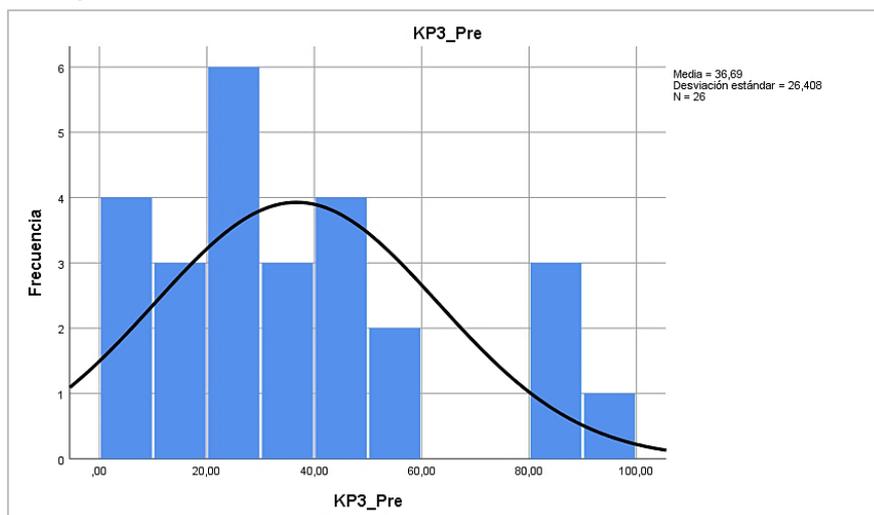
Tabla 24

Análisis de la tasa de sensibilidad

		KP3_Pre	KP3_Post
N	Válido	26	26
	Perdidos	0	0
Media		36,6888	99,3462
Mediana		29,6750	100,0000
Moda		3,51 ^a	100,00
Desviación estándar		26,40771	1,44062
Varianza		697,367	2,075
Asimetría		,874	-2,057
Error estándar de asimetría		,456	,456
Curtosis		-,056	3,098
Error estándar de curtosis		,887	,887
Rango		88,96	5,00
Mínimo		3,51	95,00
Máximo		92,47	100,00
Suma		953,91	2583,00
Percentiles	25	19,0500	100,0000
	50	29,6750	100,0000
	75	47,6200	100,0000

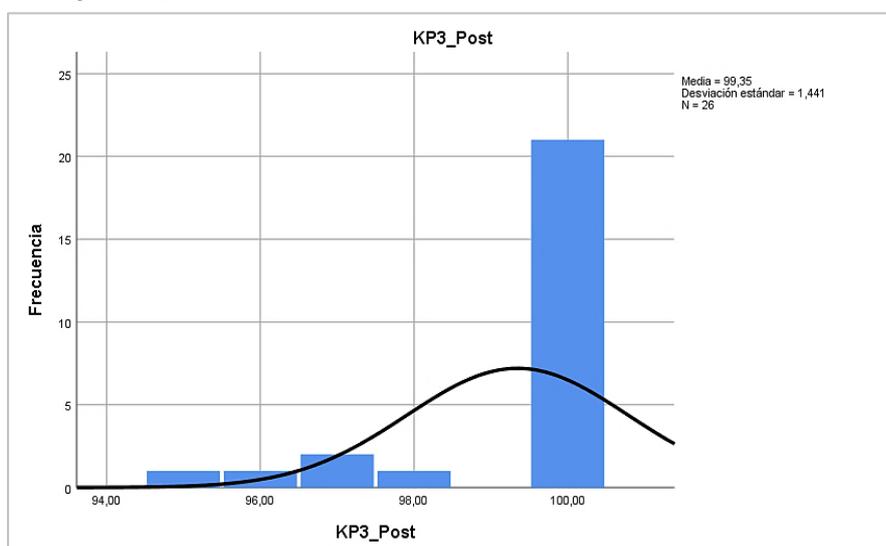
Nota. Elaborado con el Software SPSS Versión 28.

a. Existen múltiples modos. Se muestra el valor más pequeño.

Figura 18*Histograma pre KP3: Tasa de sensibilidad*

Nota. Elaborado con el Software SPSS Versión 28.

La Figura 18 muestra un histograma de la tasa de sensibilidad antes de la intervención (KP3_Pre), con la mayoría de los valores concentrados entre 20 y 40. La media de 36,69 y la desviación estándar de 26,41 evidencian una alta variabilidad y un desempeño inicial limitado, reflejando una dispersión significativa en las mediciones. Estos resultados destacan la necesidad de mejoras en la sensibilidad para lograr resultados más consistentes y acercar el rendimiento a niveles óptimos.

Figura 19*Histograma post KP3: Tasa de sensibilidad*

Nota. Elaborado con el Software SPSS Versión 28.

La Figura 19 presenta un histograma que refleja la distribución de la tasa de sensibilidad después de la intervención (KP3_Post). La mayoría de las mediciones se concentran en el valor máximo de 100, evidenciando un desempeño casi óptimo del modelo en la detección de casos positivos. La media de 99,35 y la desviación estándar de 1,44 muestran una reducción significativa en la variabilidad, lo que confirma la consistencia y uniformidad de los resultados tras la implementación de la solución propuesta.

El análisis del KPI3, enfocado en medir la tasa de sensibilidad, ha revelado un progreso significativo entre las mediciones realizadas antes y después de una intervención clave. En la fase previa a la intervención, la media se registró en 36,6888, reflejando un nivel inicial de sensibilidad frente a ciertos estímulos o condiciones. Posteriormente, esta media experimentó un aumento dramático en la fase posterior a la intervención, alcanzando un valor de 99,3462. Este notable incremento indica que la intervención tuvo un impacto sustancial, llevando la sensibilidad a un nivel casi óptimo.

La mediana también mostró un aumento considerable, pasando de 29,6750 en la fase pre-intervención a un óptimo de 100,0000 en la fase post-intervención. Este cambio subraya que más de la mitad de las mediciones después de la intervención reflejaron la máxima sensibilidad posible. Además, la moda, que indica el valor que aparece con mayor frecuencia en el conjunto de datos, cambió de un bajo 3,51 en la pre-prueba a un máximo de 100,00 en la post-prueba. Esto indica que la mayoría de las observaciones alcanzaron el valor más alto posible después de la intervención.

En cuanto a la dispersión de los datos, se notó una marcada reducción tanto en la desviación estándar como en la varianza, pasando de 26,40771 a 1,44062 y de 697,367 a 2,075, respectivamente. Este cambio sugiere que las mediciones de la

sensibilidad se volvieron mucho más consistentes y concentradas alrededor del valor medio después de la intervención.

La asimetría, que mide la simetría en la distribución de los datos, mostró un cambio de una ligera positividad (0,874) en la pre-prueba a una negatividad marcada (-2,057) en la post-prueba, reflejando un agrupamiento de los datos hacia el extremo superior de la escala. La curtosis, que indica la 'agudeza' de la distribución, pasó de ser ligeramente negativa (-0,056) a positiva y alta (3,098) en la post-prueba. Esto sugiere una concentración de los valores cerca del valor medio, indicativa de una distribución con un pico más alto y colas más delgadas.

El rango de la tasa de sensibilidad también se redujo de manera significativa, de 88,96 en la pre-prueba a solo 5,00 en la post-prueba, lo que destaca una disminución notable en la variabilidad de los resultados. Los valores mínimos y máximos reafirman este patrón de mejora, pasando de 3,51 y 92,47, respectivamente, en la pre-prueba, a 95,00 y 100,00 en la post-prueba. Esto confirma que incluso el desempeño más bajo mejoró significativamente y se acercó al nivel ideal tras la intervención.

En resumen, estos resultados destacan el impacto significativo y positivo de la intervención en el KPI3, mostrando un notable aumento en la media y la mediana hacia niveles casi óptimos de sensibilidad, junto con una marcada reducción en la variabilidad, lo que refleja una mayor consistencia en los resultados, y una considerable mejora en el rendimiento mínimo, acercándolo al nivel máximo posible.

Por último, los percentiles 25, 50 y 75 se elevaron uniformemente a 100,0000 en la post-prueba, lo que evidencia que al menos el 75% de las mediciones alcanzaron la tasa de sensibilidad máxima después de la intervención, demostrando un alto grado de uniformidad y efectividad en los resultados logrados.

Estadística descriptiva del KPI 4

Los resultados del Indicador Clave de Rendimiento (KPI) 4, que evalúa el tiempo necesario para el procesamiento de datos, se presentan a continuación tanto para la pre-prueba como para la post-prueba, permitiendo analizar en detalle cómo la intervención impactó en la eficiencia operativa del sistema. Este análisis busca evidenciar no solo una posible reducción en los tiempos de procesamiento, sino también una mayor consistencia y optimización en el rendimiento general del modelo en ambos escenarios.

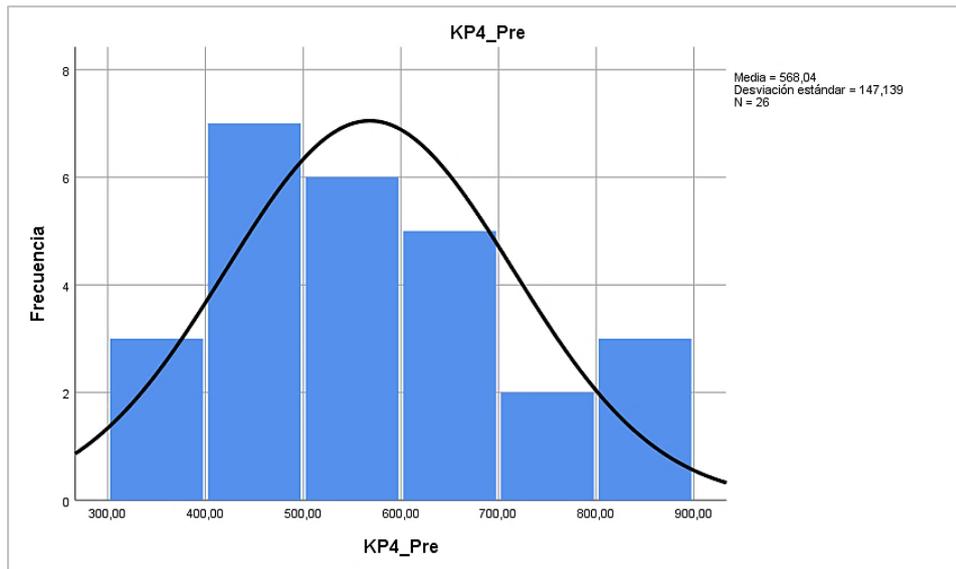
Tabla 25

Análisis del tiempo de procesamiento de datos

		KP4_Pre	KP4_Post
N	Válido	26	26
	Perdidos	0	0
Media		568,0385	1,0009588
Mediana		572,5000	1,0011100
Moda		369,00 ^a	1,00111
Desviación estándar		147,13898	,00029479
Varianza		21649,878	,000
Asimetría		,682	-1,752
Error estándar de asimetría		,456	,456
Curtosis		-,245	2,073
Error estándar de curtosis		,887	,887
Rango		509,00	,00097
Mínimo		369,00	1,00018
Máximo		878,00	1,00115
Suma		14769,00	26,02493
Percentiles	25	448,0000	1,0008050
	50	572,5000	1,0011100
	75	637,0000	1,0011300

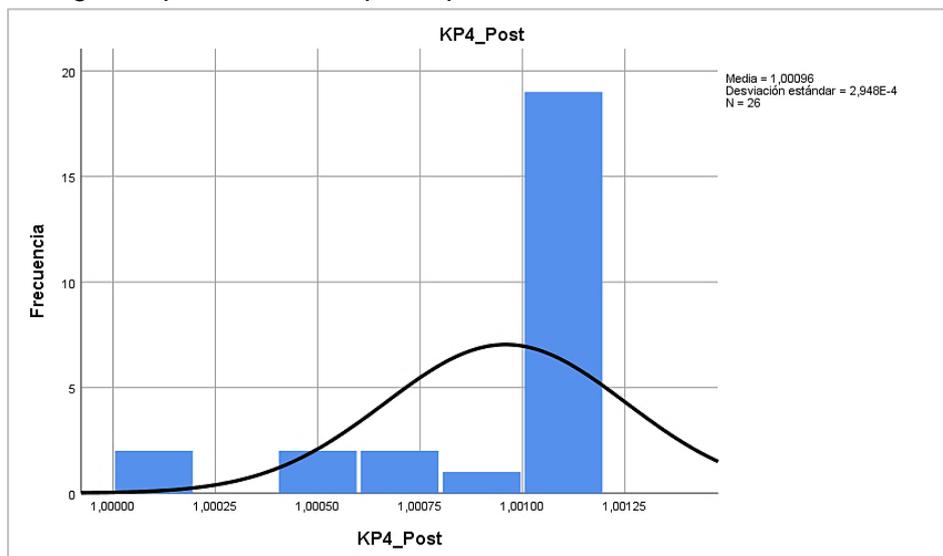
Nota. Elaborado con el Software SPSS Versión 28.

a. Existen múltiples modos. Se muestra el valor más pequeño.

Figura 20*Histograma pre KP4: Tiempo de procesamiento de datos*

Nota. Elaborado con el Software SPSS Versión 28.

La Figura 20 muestra un histograma del tiempo de procesamiento de datos antes de la intervención (KP4_Pre), con valores concentrados entre 400 y 700 milisegundos. La media de 568,04 y la desviación estándar de 147,14 reflejan una notable variabilidad, evidenciando una eficiencia inicial limitada y la necesidad de optimizar los tiempos de procesamiento.

Figura 21*Histograma post KP4: Tiempo de procesamiento de datos*

Nota. Elaborado con el Software SPSS Versión 28.

La Figura 21 muestra un histograma del tiempo de procesamiento de datos después de la intervención (KP4_Post), con la mayoría de los valores concentrados en torno a 1 milisegundo. La media de 1,00096 y la desviación estándar de 0,00029 reflejan una drástica reducción en la variabilidad y una alta consistencia en los tiempos de procesamiento, destacando la efectividad de la intervención para optimizar significativamente la eficiencia del sistema.

El análisis del KPI4, que se centra en evaluar el tiempo de procesamiento de datos, ha revelado una transformación significativa y positiva como resultado de una intervención específica. Antes de esta intervención, la media de tiempo de procesamiento se situaba en 568,0385 segundos, indicando un período considerable para procesar datos. Sin embargo, tras la implementación de mejoras, esta media se redujo drásticamente a tan solo 1,0009588 segundos en la post-prueba. Esta disminución es un claro indicativo de una eficiencia notablemente mejorada en el procesamiento de datos.

La mediana, que proporciona una medida de tendencia central menos influenciada por valores atípicos en comparación con la media, también experimentó una disminución importante, pasando de 572,5000 segundos en la pre-prueba a tan solo 1,0011100 segundos en la post-prueba. Este cambio sugiere que más de la mitad de los tiempos de procesamiento después de la intervención fueron excepcionalmente breves.

En relación con la moda, que representa el valor más frecuente, se observa que cambió de 369,00 segundos en la fase Pre-intervención a 1,0011100 segundos en la fase Post-intervención. Esto sugiere una consistencia significativa en los tiempos de procesamiento después de la intervención.

La desviación estándar y la varianza, que miden la dispersión de los datos, mostraron una disminución significativa, pasando de 147,13898 a 0,00029479 y de 21649,878 a prácticamente cero, respectivamente. Estos cambios indican que los tiempos de procesamiento después de la intervención se volvieron extremadamente consistentes y se agruparon estrechamente alrededor de la media.

La asimetría y la curtosis también experimentaron cambios significativos. La asimetría pasó de ser moderadamente positiva (0,682), lo que sugería una cola más larga hacia los tiempos más altos, a negativa (-1,752) en la post-prueba, reflejando una concentración de tiempos de procesamiento en el extremo más bajo. La curtosis cambió de un valor ligeramente negativo (-0,245) a un valor positivo (2,073), lo que indica una distribución más puntiaguda con una concentración más alta de valores alrededor de la media en la post-prueba.

Además, el rango del tiempo de procesamiento se redujo drásticamente de 509,00 segundos a solo 0,00097 segundos, destacando una notable disminución en la variabilidad. Los valores mínimo y máximo pasaron de 369,00 y 878,00 segundos en la pre-prueba a 1,00018 y 1,00115 segundos en la post-prueba, confirmando una mejora significativa y una estandarización en el tiempo de procesamiento.

Por último, los percentiles muestran que, mientras en la pre-prueba había diversidad en los tiempos de procesamiento, en la post-prueba los percentiles 25, 50 y 75 se agruparon cerca de 1,001 segundos, evidenciando una notable homogeneización y una reducción significativa en la dispersión de los tiempos.

Estadística descriptiva del KPI 5

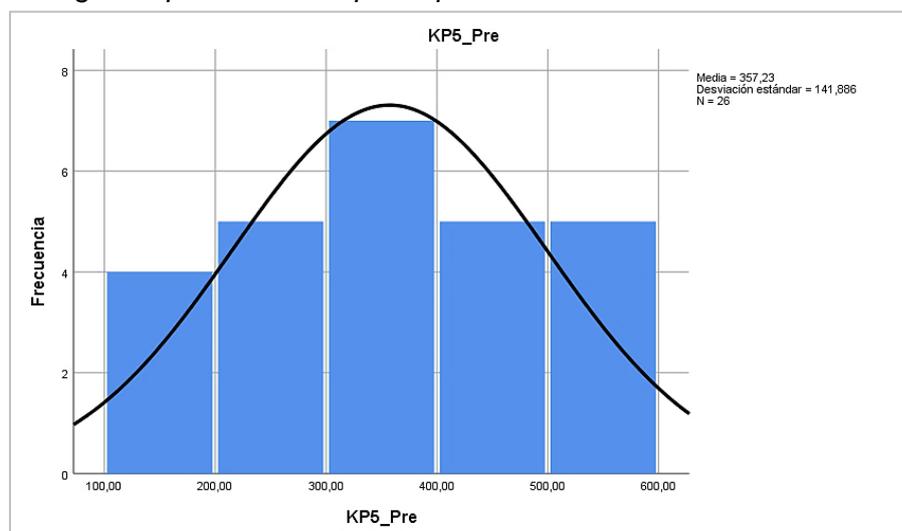
A continuación, se presentan los resultados obtenidos del KPI 5, que corresponde al tiempo requerido para predecir, tanto en la pre-prueba como en la post-prueba.

Tabla 26*Análisis del tiempo de predicción*

		KP5_Pre	KP5_Post
N	Válido	26	26
	Perdidos	0	0
Media		357,2308	,00358
Mediana		352,5000	,00300
Moda		120,00 ^a	,003
Desv. Desviación		141,88624	,001102
Varianza		20131,705	,000
Asimetría		,065	,955
Error estándar de asimetría		,456	,456
Curtosis		-,846	,038
Error estándar de curtosis		,887	,887
Rango		467,00	,004
Mínimo		120,00	,002
Máximo		587,00	,006
Suma		9288,00	,093
Percentiles	25	245,0000	,00300
	50	352,5000	,00300
	75	458,2500	,00425

Nota. Elaborado con el Software SPSS Versión 28.

a. Existen múltiples modos. Se muestra el valor más pequeño.

Figura 22*Histograma pre KP5: Tiempo de predicción*

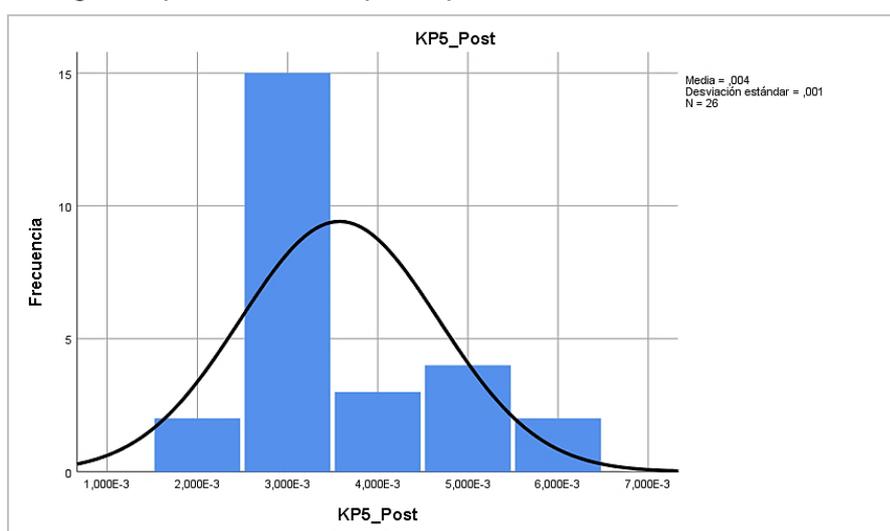
Nota. Elaborado con el Software SPSS Versión 28.

La Figura 22 muestra un histograma que representa la distribución del tiempo de predicción antes de la intervención (KP5_Pre). Los valores están concentrados

principalmente entre 300 y 400 milisegundos, con una media de 357,23 y una desviación estándar de 141,89, lo que refleja una notable dispersión en los tiempos registrados. La curva de distribución superpuesta indica una ligera simetría, evidenciando una variabilidad considerable en los datos iniciales. Estos resultados sugieren la necesidad de optimizar los tiempos de predicción para alcanzar un desempeño más consistente y eficiente.

Figura 23

Histograma post KP5: Tiempo de predicción



Nota. Elaborado con el Software SPSS Versión 28.

La Figura 23 muestra un histograma del tiempo de predicción después de la intervención (KP5_Post), evidenciando una mejora significativa en la eficiencia. La mayoría de los valores están concentrados alrededor de 0,004 segundos, con una media de 0,004 y una desviación estándar de 0,001, lo que refleja una notable reducción en la variabilidad. La curva de distribución superpuesta confirma la uniformidad en los tiempos de predicción, destacando la efectividad de la intervención para optimizar el desempeño y alcanzar un nivel de consistencia elevado.

El análisis del KPI5, que evalúa un parámetro específico, muestra un cambio significativo y positivo como resultado de una intervención importante. Antes de la intervención, la media en la fase pre-intervención era de 357,2308, lo que sugiere un

valor considerable para este parámetro. Sin embargo, después de la intervención, la media disminuyó notablemente a solo 0,00358 en la fase post-intervención. Esta reducción indica una mejora considerable en la eficiencia o eficacia del parámetro medido.

La mediana, que es resistente a los valores atípicos y representa el valor central del conjunto de datos, también experimentó una disminución notable, pasando de 352,5000 a solo 0,00300. Esto implica que más de la mitad de los valores después de la intervención se situaron en torno a este valor extremadamente bajo.

En cuanto a la moda, que representa el valor más frecuente en el conjunto de datos, se observa un cambio de 120,00 en la fase Pre-intervención a 0,003 en la fase Post-intervención. Este cambio indica una notable uniformidad en los valores medidos después de la intervención.

La desviación estándar y la varianza, que son indicadores de la dispersión de los datos, mostraron una disminución significativa, pasando de 141,88624 a solo 0,001102 y de 20131,705 a prácticamente cero, respectivamente. Estos cambios sugieren que los valores medidos después de la intervención se volvieron extremadamente consistentes y estrechamente agrupados alrededor de la media.

La asimetría y la curtosis también experimentaron cambios notables. La asimetría pasó de ser casi nula (0,065) a moderadamente positiva (0,955), lo que puede indicar un ligero cambio en la distribución de los datos hacia valores más altos. La curtosis cambió de un valor ligeramente negativo (-0,846) a un valor casi neutro (0,038), sugiriendo una distribución más normal en la post-prueba.

El rango del parámetro medido, que indica la diferencia entre los valores mínimo y máximo, se contrajo drásticamente de 467,00 a solo 0,004, destacando una reducción significativa en la variabilidad de los valores observados. Esta contracción

evidencia el impacto positivo de la intervención en la estabilización de los resultados. Los valores mínimo y máximo pasaron de 120,00 y 587,00 en la pre-prueba a 0,002 y 0,006 en la post-prueba, lo que confirma una estandarización notable y una mayor consistencia en los valores después de la intervención.

Finalmente, los percentiles revelan que, aunque en la pre-prueba había una notable diversidad, en la post-prueba los percentiles 25, 50 y 75 se concentraron cerca entre sí, alrededor de valores muy bajos.

Estadística descriptiva del KPI 6

Se presentan los resultados del KPI 6, correspondiente al nivel de confianza en la toma de decisiones, tanto en la pre-prueba como en la post-prueba.

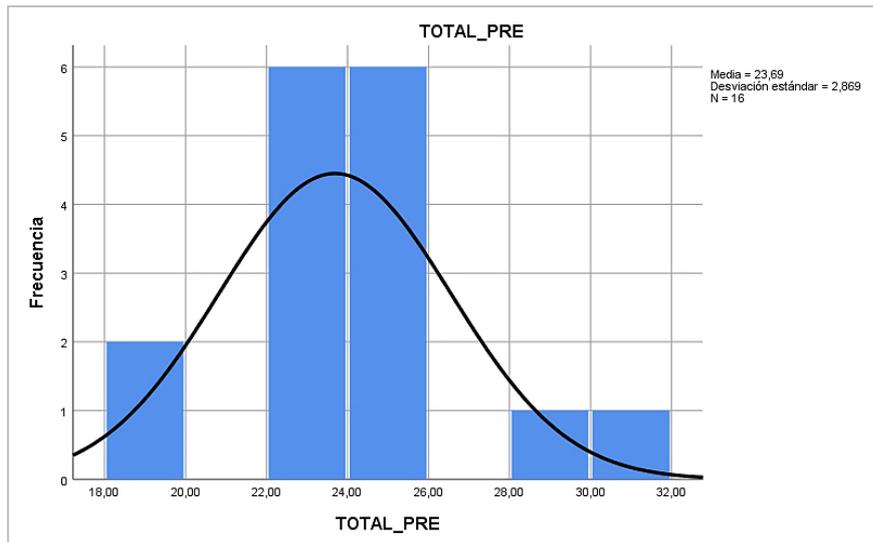
Tabla 27

Análisis del nivel de confianza en la toma de decisiones

		TOTAL_PRE	TOTAL_POST
N	Válido	16	16
	Perdidos	10	10
Media		23,6875	40,8125
Mediana		23,5000	46,0000
Moda		25,00	50,00
Desviación estándar		2,86865	14,14788
Varianza		8,229	200,163
Asimetría		,378	-1,675
Error estándar de asimetría		,564	,564
Curtosis		,627	1,286
Error estándar de curtosis		1,091	1,091
Rango		11,00	40,00
Mínimo		19,00	10,00
Máximo		30,00	50,00
Suma		379,00	653,00
Percentiles	25	22,0000	42,2500
	50	23,5000	46,0000
	75	25,0000	50,0000

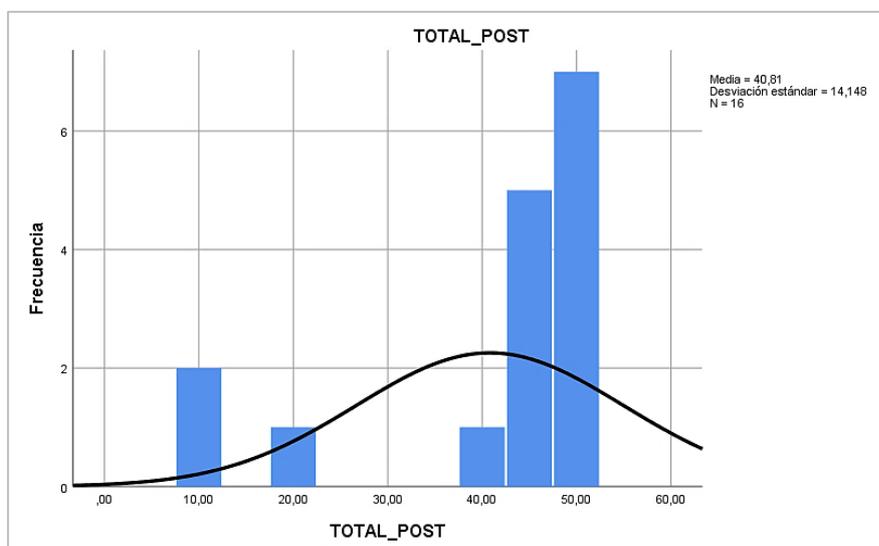
Nota. Elaborado con el Software SPSS Versión 28.

a. Existen múltiples modos. Se muestra el valor más pequeño.

Figura 24*Histograma pre KP6: Nivel de confianza en la toma de decisiones*

Nota. Elaborado con el Software SPSS Versión 28.

La Figura 24 muestra un histograma del nivel de confianza en la toma de decisiones antes de la intervención (TOTAL_PRE), con valores concentrados entre 22 y 26. La media de 23,69 y la desviación estándar de 2,87 reflejan una moderada dispersión, indicando un nivel de confianza inicial limitado y margen para mejoras en la seguridad en la toma de decisiones.

Figura 25*Histograma post KP6: Nivel de confianza en la toma de decisiones*

Nota. Elaborado con el Software SPSS Versión 28.

La Figura 25 muestra un histograma del nivel de confianza en la toma de decisiones después de la intervención (TOTAL_POST). Los valores están concentrados principalmente entre 40 y 50, con una media de 40,81 y una desviación estándar de 14,15, lo que indica un aumento significativo en el nivel de confianza. La curva de distribución refleja una mayor dispersión respecto a la pre-prueba, pero evidencia un cambio positivo hacia niveles superiores de confianza en las decisiones tras la intervención.

La evaluación de dos conjuntos de datos, etiquetados como TOTAL_PRE y TOTAL_POST, nos ofrece una perspectiva interesante sobre los cambios ocurridos tras una intervención o evento. Ambos conjuntos contienen 16 observaciones válidas, aunque se señala la pérdida de 10 datos en cada caso, lo que sugiere que algunas mediciones no estaban disponibles o no se registraron.

En la fase previa a la intervención (TOTAL_PRE), la media se situaba en 23,6875. Este valor representa el promedio de todas las mediciones recogidas antes de la intervención. Posteriormente, en la fase posterior a la intervención (TOTAL_POST), la media aumentó a 40,8125, lo que indica un incremento significativo en el valor medio del parámetro medido tras la intervención.

La mediana, que divide el conjunto de datos en dos mitades iguales, también mostró un cambio notable. En TOTAL_PRE era de 23,5000, mientras que en TOTAL_POST ascendió a 46,0000. Este cambio en la mediana sugiere que más de la mitad de las mediciones en TOTAL_POST fueron iguales o superiores a 46, indicando un desplazamiento hacia valores más altos.

En cuanto a la moda, que es el valor que aparece con mayor frecuencia en el conjunto de datos, observamos un aumento de 25,00 en TOTAL_PRE a 50,00 en

TOTAL_POST. Este cambio refleja una concentración de frecuencia en valores más altos después de la intervención.

La desviación estándar y la varianza, que indican la dispersión de los datos alrededor de la media, mostraron cambios significativos. La desviación estándar pasó de 2,86865 en TOTAL_PRE a 14,14788 en TOTAL_POST, y la varianza de 8,229 a 200,163. Estos aumentos sugieren que hubo una mayor variabilidad en las mediciones después de la intervención.

En términos de asimetría y curtosis, que describen la forma de la distribución de los datos, hubo cambios notables. La asimetría cambió de una ligera positividad (0,378) en TOTAL_PRE a una negatividad más pronunciada (-1,675) en TOTAL_POST, lo que podría indicar una distribución más sesgada hacia los valores más bajos después de la intervención. La curtosis aumentó de 0,627 a 1,286, sugiriendo una distribución con un pico más alto y colas más pronunciadas en TOTAL_POST.

El rango de los datos, que es la diferencia entre los valores mínimo y máximo, se expandió de 11,00 en TOTAL_PRE a 40,00 en TOTAL_POST, reflejando un aumento en la variabilidad de los valores medidos. Los valores mínimo y máximo se incrementaron de 19,00 y 30,00 en TOTAL_PRE a 10,00 y 50,00 en TOTAL_POST, respectivamente. Este cambio muestra que, aunque el valor mínimo disminuyó, el valor máximo aumentó, ampliando la gama de valores observados.

Por último, los percentiles muestran que en TOTAL_PRE, el 25% de los datos eran menores o iguales a 22,0000, el 50% menores o iguales a 23,5000 y el 75% menores o iguales a 25,0000. En TOTAL_POST, estos valores aumentaron a 42,2500, 46,0000 y 50,0000, respectivamente, lo que indica un desplazamiento general hacia valores más altos en la distribución.

Estos resultados indican que la intervención ha tenido como consecuencia un aumento en el nivel de confianza en la toma de decisiones, tal como lo demuestra el incremento en la media, mediana y moda. Sin embargo, este aumento también ha venido acompañado de una mayor variabilidad entre los individuos o las medidas, como lo reflejan los aumentos en la desviación estándar y la varianza.

Contrastación de hipótesis

A continuación, se presentan las medias obtenidas de los KPI 1, KPI 2, KPI 3, KPI 4, KPI 5 y KPI 6.

Tabla 28

Medias de los KPIs

Indicador	Pre-prueba	Post-prueba	Comentario
Tasa de exactitud de la predicción	25,2758	99,5000	
Tasa de precisión de la predicción	30,7300	99,3462	
Tasa de sensibilidad de la predicción	36,6888	99,3462	
Tiempo de procesamiento de datos	568,0385	1,0009588	
Tiempo de predicción	357,2308	,00358	
Nivel de confianza en la toma de decisiones			Indicador cualitativo (no contrastado)

Contrastación del KPI 1: Tasa de exactitud de la predicción

Este proceso implica la evaluación del efecto que la implementación de la solución propuesta tiene sobre la precisión con la que se pueden predecir los resultados. Se comparan dos conjuntos de datos: uno recopilado antes de la adopción de la solución (pre-prueba) y otro después de su implementación (post-prueba). El objetivo es verificar cómo la solución mejora la habilidad de predecir

resultados de manera correcta. A continuación, se presentan los valores correspondientes a ambas pruebas, destacando las mejoras en la precisión y la reducción de errores en la predicción.

Tabla 29

Valores pre-prueba del KPI 1

Valores pre-prueba KPI 1						
7.49	15.27	34.39	12.17	23.09	9.33	28.26
22.99	17.30	31.89	45.32	7.26	16.55	55.22
12.49	31.37	19.63	39.18	59.95	15.30	24.66
20.30	37.91	20.04	14.36	35.45		

La tabla 29 presenta los valores de pre-prueba del KPI 1, que representan la tasa de exactitud de diversas pruebas médicas realizadas antes de la implementación de la solución propuesta. Cada valor numérico (7.49, 15.27, 34.39, etc.) refleja la exactitud de los resultados clínicos previos, comparándolos con un resultado de referencia considerado correcto y permitiendo identificar inconsistencias o errores en las predicciones iniciales. Estos datos se obtuvieron revisando registros históricos detallados para determinar cuántos de los resultados iniciales coincidían con los resultados de seguimiento o confirmaciones posteriores, proporcionando una línea base clave para medir el impacto de la solución.

Tabla 30

Valores post-prueba del KPI 1

Valores post-prueba KPI 1						
100.0	100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	97.0	100.0	97.0
96.0	100.0	98.0	100.0	99.0		

La tabla 30 muestra los valores de post-prueba del KPI 1, reflejando la tasa de exactitud de las mismas pruebas médicas después de la implementación de la solución propuesta. Los valores alcanzan consistentemente 100.0, con pocas excepciones de 97.0, 96.0 y 98.0, indicando una mejora significativa y sostenida en la precisión de los resultados. Esta notable uniformidad en los valores destaca la efectividad y robustez de la solución implementada en aumentar la exactitud de las predicciones, reduciendo casi completamente los errores, optimizando la confiabilidad de las pruebas y demostrando un desempeño superior en comparación con los datos pre-prueba.

Hipótesis alternativa (Ha):

La implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de exactitud de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.

Hipótesis nula (Ho):

La implementación de un modelo basado en algoritmos de Machine Learning no mejora la tasa de exactitud de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.

a) Planteamiento de la hipótesis:

μ_1 = Corresponde a la media de la tasa de exactitud de la predicción en la pre-prueba.

μ_2 = Corresponde a la media de la tasa de exactitud de la predicción en la post-prueba.

$$Ho: \mu_1 = \mu_2$$

$$Ha: \mu_1 < \mu_2$$

b) Criterios de decisión:

Si $p \geq 0.05$, aceptamos la Ho y rechazamos la Ha.

Si $p \leq 0.05$, rechazamos la H_0 y rechazamos la H_a .

Tabla 31

Resultados de las pruebas de normalidad: KPI 1

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP1_Pre	,139	26	,200*	,925	26	,058
KP1_Post	,477	26	,000	,504	26	,000

Nota. Elaborado con el Software SPSS Versión 28.

*. Esto es un límite inferior de la significación verdadera

a. Corrección de significación de Lilliefors

c) Prueba estadística:

Como en cada KPI evaluado se visualiza que en el pre hay distribución normal, pero en el post es distribución no normal, entonces se aplicaría estadística no paramétrica, con Wilcoxon.

d) Decisión estadística:

Tabla 32

Prueba de Wilcoxon KPI 1

	Hipótesis nula	Prueba	Sig.	Decisión
1	La mediana de diferencias entre KP1_Pre y KP1_Post es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,000	Rechace la hipótesis nula.

Nota. Elaborado con el Software SPSS Versión 28. El nivel de significación es de ,000.

Para el indicador KPI 1, se utilizó la prueba de Wilcoxon de rangos con signo en muestras relacionadas con el fin de evaluar si existía una diferencia estadísticamente relevante en las mediciones de una variable, tanto previo como posterior a una intervención o evento. Esta prueba es particularmente útil cuando las mediciones se realizan en los mismos sujetos en dos momentos distintos, permitiendo una comparación directa de los cambios producidos por la intervención.

La hipótesis nula del análisis estableció que no hay diferencia en la mediana de las mediciones de KP1 antes (KP1_Pre) y después (KP1_Post) de la intervención.

La mediana representa un valor central en un conjunto de datos, indicando su posición media. Tomando eso en cuenta, la hipótesis nula sugiere que, en términos de mediana, las mediciones de KP1 no se ven afectadas por la intervención; es decir, la intervención no tiene un efecto significativo sobre la mediana de KP1.

El resultado de la prueba mostró un valor de significación (Sig.) de 0,000. Este valor es una medida de probabilidad que indica cuán probable es obtener los resultados observados si la hipótesis nula fuera cierta. En la práctica estadística, un valor de p (el valor de significación) menor que 0.05 generalmente se considera como indicativo de una diferencia significativa. Un valor de p tan extremadamente bajo como 0,000 sugiere que la probabilidad de observar una diferencia como la encontrada, o mayor, si la hipótesis nula fuera cierta, es prácticamente nula. Esto constituye una sólida evidencia estadística en contradicción con la hipótesis nula.

Debido a este valor de p muy pequeño, se opta por rechazar la hipótesis nula. Esto implica que las mediciones de KP1 muestran una diferencia significativa en su mediana antes y después de la intervención. Eso quiere decir que la intervención o evento analizado ha tenido un efecto considerable y estadísticamente significativo en la variable KP1, alterando su mediana de manera que no se puede atribuir a la variabilidad aleatoria o a las fluctuaciones ordinarias en las mediciones.

Contrastación del KPI 2: Tasa de precisión de la predicción

El análisis de este indicador se centra en determinar la mejora en la precisión de las predicciones realizadas a raíz de la implementación de la solución propuesta. Para ello, se examina la relación entre las predicciones acertadas y el total de predicciones realizadas, comparando los resultados obtenidos antes de la introducción de la solución (pre-prueba) y después de su puesta en marcha (post-prueba). El propósito es cuantificar el incremento en la tasa de predicciones precisas

y cómo esto contribuye a una operación más eficiente. Los valores de ambos periodos se presentan a continuación, resaltando la disminución en la tasa de fallos y el aumento en la confiabilidad de las predicciones.

Tabla 33

Valores pre-prueba del KPI 2

Valores pre-prueba KPI 2						
8.70	36.57	45.54	16.10	32.59	3.55	26.80
19.97	20.30	43.58	4.00	12.26	14.06	82.12
4.65	60.29	12.86	34.54	58.56	19.71	34.14
27.10	38.68	33.93	39.93	68.45		

La tabla 33 presenta los valores de pre-prueba del KPI 2, que representan la tasa de precisión de las predicciones realizadas antes de la implementación de la solución propuesta. Cada valor numérico (8.70, 36.57, 45.54, etc.) refleja la proporción de resultados clínicos previos que, ya fueran positivos o negativos, efectivamente correspondieron a casos confirmados en comparación con el total de resultados entregados. Estos datos permiten establecer una línea base para medir el impacto de la intervención. La variabilidad en los valores indica diferentes niveles de precisión en las pruebas antes de la intervención, proporcionando una referencia para comparar con los resultados post-prueba.

Tabla 34

Valores post-prueba del KPI 2

Valores post-prueba KPI 2						
100.0	100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	95.0	100.0	97.0
96.0	100.0	97.00	100.0	98.0		

La tabla 34 muestra los valores de post-prueba del KPI 2, reflejando la tasa de precisión de las predicciones después de la implementación de la solución propuesta. Los valores consistentemente altos, mayoritariamente 100.0, con pocas excepciones de 95.0, 97.0 y 98.0, indican una mejora sustancial en la precisión de las predicciones. Esta consistencia resalta la efectividad de la solución en aumentar la confiabilidad, reducir los errores en las predicciones y garantizar que el modelo se adapte de manera eficiente a los datos procesados, fortaleciendo su aplicabilidad en escenarios reales.

Hipótesis alternativa (Ha):

La implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de precisión de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.

Hipótesis nula (Ho):

La implementación de un modelo basado en algoritmos de Machine Learning no mejora la tasa de precisión de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.

a) Planteamiento de la hipótesis:

μ_1 = Corresponde a la media de la tasa de precisión de la predicción en la pre-prueba.

μ_2 = Corresponde a la media de la tasa de precisión de la predicción en la post-prueba.

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 < \mu_2$$

b) Criterios de decisión:

Si $p \geq 0.05$, aceptamos la H_0 y rechazamos la H_a .

Si $p < 0.05$, rechazamos la H_0 y rechazamos la H_a .

Tabla 35*Resultados de las pruebas de normalidad: KPI 2*

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP2_Pre	,118	26	,200*	,942	26	,148
KP2_Post	,483	26	,000	,520	26	,000

Nota. Elaborado con el Software SPSS Versión 28.

*. Esto es un límite inferior de la significación verdadera

a. Corrección de significación de Lilliefors

c) Prueba estadística:

Como en cada KPI evaluado se visualiza que en el pre hay distribución normal, pero en el post es distribución no normal, entonces se aplicaría estadística no paramétrica, con Wilcoxon.

d) Decisión estadística:

Tabla 36*Prueba de Wilcoxon KPI 2*

	Hipótesis nula	Prueba	Sig.	Decisión
1	La mediana de diferencias entre KP2_Pre y KP2_Post es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,000	Rechace la hipótesis nula.

Nota. Elaborado con el Software SPSS Versión 28. El nivel de significación es de ,000.

Para el KPI 2 se empleó prueba de rangos con signo de Wilcoxon para muestras relacionadas, con el objetivo de examinar si existe una diferencia estadísticamente significativa en las mediciones de una variable, antes y después de una intervención o evento. Esta prueba es particularmente útil cuando las mediciones se realizan en los mismos sujetos en dos momentos distintos, permitiendo una comparación directa de los cambios producidos por la intervención.

La hipótesis nula del análisis estableció que no hay diferencia en la mediana de las mediciones de KP2 antes (KP2_Pre) y después (KP2_Post) de la intervención. La mediana es una medida de tendencia central que representa el valor medio en un

conjunto de datos. Según la hipótesis nula, en términos de mediana, las mediciones de KP2 no se ven afectadas por la intervención; es decir, la intervención no tiene un efecto significativo sobre la mediana de KP2.

El resultado de la prueba mostró un valor de significación (Sig.) de 0,000. Este valor es una medida de probabilidad que indica cuán probable es obtener los resultados observados si la hipótesis nula fuera cierta. En la práctica estadística, un valor de p (el valor de significación) menor que 0.05 generalmente se considera como indicativo de una diferencia significativa. Un valor de p tan extremadamente bajo como 0,000 sugiere que la probabilidad de observar una diferencia como la encontrada, o mayor, si la hipótesis nula fuera cierta, es prácticamente nula. Esto proporciona una fuerte evidencia estadística en contra de la hipótesis nula.

Por lo tanto, en base a este valor de p extremadamente bajo, se toma la decisión de rechazar la hipótesis nula. Esto implica que las mediciones de KP2 muestran una diferencia significativa en su mediana antes y después de la intervención. Eso quiere decir que la intervención o evento analizado ha tenido un efecto considerable y estadísticamente significativo en la variable KP2, alterando su mediana de manera que no se puede atribuir a la variabilidad aleatoria o a las fluctuaciones ordinarias en las mediciones.

Contrastación del KPI 3: Tasa de sensibilidad de la predicción

Este indicador clave de rendimiento evalúa cómo la introducción de la solución impacta en la capacidad de identificar correctamente los casos positivos. Se compara la eficacia antes (pre-prueba) y después (post-prueba) de aplicar la solución, con el fin de determinar mejoras en la detección acertada de situaciones que requieren atención. Los datos recabados se exponen a continuación, evidenciando el aumento en la tasa de sensibilidad post-implementación.

Tabla 37*Valores pre-prueba del KPI 3*

Valores pre-prueba KPI 3						
4.82	18.18	40.67	21.59	26.59	4.90	46.67
58.06	32.73	25.95	92.47	6.37	45.42	50.47
3.51	35.29	83.33	88.02	85.08	19.34	22.41
23.84	46.47	26.62	12.21	32.90		

La tabla 37 presenta los valores de pre-prueba del KPI 3, que representan la tasa de sensibilidad en la clasificación correcta de resultados clínicos por los tecnólogos médicos antes de la implementación de la solución. Cada valor (4.82, 18.18, 40.67, etc.) refleja la eficacia inicial en identificar casos positivos utilizando las pruebas médicas y rangos referenciales disponibles en ese momento, proporcionando una base para comparar con los resultados post-intervención.

Tabla 38*Valores post-prueba del KPI 3*

Valores post-prueba KPI 3						
100.0	100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	95.0	100.0	97.0
96.0	100.0	97.0	100.0	98.0		

La tabla 38 muestra los valores de post-prueba del KPI 3, reflejando la tasa de sensibilidad en la clasificación correcta de resultados clínicos por el modelo de Machine Learning después de la implementación de la solución. Los valores consistentemente altos, mayoritariamente 100.0, con algunas excepciones de 95.0, 97.0 y 98.0, indican una mejora significativa en la capacidad del modelo para identificar correctamente los casos positivos, demostrando la alta efectividad de la solución implementada. Este sobresaliente desempeño destaca no solo la precisión

del modelo, sino también su confiabilidad para aplicarse en entornos clínicos reales, donde la detección precisa de casos positivos es crucial para la toma de decisiones médicas.

Hipótesis alternativa (Ha):

La implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de sensibilidad de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.

Hipótesis nula (Ho):

La implementación de un modelo basado en algoritmos de Machine Learning no mejora la tasa de sensibilidad de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.

a) Planteamiento de la hipótesis:

μ_1 = Corresponde a la media de la tasa de sensibilidad de la predicción en la pre-prueba.

μ_2 = Corresponde a la media de la tasa de sensibilidad de la predicción en la post-prueba.

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 < \mu_2$$

b) Criterios de decisión:

Si $p \geq 0.05$, aceptamos la H_0 y rechazamos la H_a .

Si $p < 0.05$, rechazamos la H_0 y rechazamos la H_a .

Tabla 39

Resultados de las pruebas de normalidad: KPI 3

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP3_Pre	,149	26	,145	,898	26	,014
KP3_Post	,483	26	,000	,520	26	,000

Nota. Elaborado con el Software SPSS Versión 28.

a. Corrección de significación de Lilliefors

c) Prueba estadística:

Como en cada KPI evaluado se visualiza que en el pre hay distribución normal, pero en el post es distribución no normal, entonces se aplicaría estadística no paramétrica, con Wilcoxon.

d) Decisión estadística

Tabla 40

Prueba de Wilcoxon KPI 3

	Hipótesis nula	Prueba	Sig.	Decisión
1	La mediana de diferencias entre KP3_Pre y KP3_Post es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,000	Rechace la hipótesis nula.

Nota. Elaborado con el Software SPSS Versión 28. El nivel de significación es de ,000.

En el caso del indicador KPI 3, se utilizó la prueba de Wilcoxon de rangos con signo en muestras relacionadas para analizar si se observa una diferencia estadísticamente relevante en las mediciones de una variable antes y después de una intervención o evento. Esta prueba es particularmente útil cuando las mediciones se realizan en los mismos sujetos en dos momentos distintos, permitiendo una comparación directa de los cambios producidos por la intervención.

La hipótesis nula del análisis estableció que no hay diferencia en la mediana de las mediciones de KP3 antes (KP3_Pre) y después (KP3_Post) de la intervención. La mediana representa una medida de centralización que señala el valor medio en un conjunto de datos. Tomando eso en cuenta, la hipótesis nula sugiere que, en términos de mediana, las mediciones de KP3 no se ven afectadas por la intervención; es decir, la intervención no tiene un efecto significativo sobre la mediana de KP3.

El resultado de la prueba mostró un valor de significación (Sig.) de 0,000. Este valor es una medida de probabilidad que indica cuán probable es obtener los resultados observados si la hipótesis nula fuera cierta. En la práctica estadística, un

valor de p (el valor de significación) menor que 0.05 generalmente se considera como indicativo de una diferencia significativa. Un valor de p tan extremadamente bajo como 0,000 sugiere que la probabilidad de observar una diferencia como la encontrada, o mayor, si la hipótesis nula fuera cierta, es prácticamente nula. Esto proporciona una fuerte evidencia estadística en contra de la hipótesis nula.

Un valor de p tan extraordinariamente reducido, indicado por 0,000, implica que la posibilidad de presenciar una discrepancia del tamaño observado o superior, bajo la suposición de que la hipótesis nula es correcta, es casi inexistente. Esto implica que las mediciones de KP3 muestran una diferencia significativa en su mediana antes y después de la intervención. Eso quiere decir que la intervención o evento analizado ha tenido un efecto considerable y estadísticamente significativo en la variable KP3, alterando su mediana de manera que no se puede atribuir a la variabilidad aleatoria o a las fluctuaciones ordinarias en las mediciones.

Contrastación del KPI 4: Tiempo de procesamiento de datos

Este KPI mide el impacto de la solución propuesta en la agilización del procesamiento de datos. Se analiza la duración del procesamiento antes de la implementación (pre-prueba) y tras ella (post-prueba), con el fin de evidenciar cualquier reducción significativa en el tiempo requerido. Los resultados se muestran a continuación, destacando la eficiencia ganada en el tratamiento de los datos.

Tabla 41

Valores pre-prueba del KPI 4

Valores pre-prueba KPI 4						
578	579	606	872	369	458	587
472	392	878	542	816	603	439
701	623	451	508	765	471	567
633	387	414	409	649		

La tabla 41 presenta los valores de pre-prueba del KPI 4, que representan el tiempo promedio de procesamiento de datos antes de la implementación de la solución propuesta. Cada valor numérico (578, 579, 606, etc.) indica el tiempo en segundos que le tomaba al personal correspondiente, como los tecnólogos médicos, recopilar, organizar y procesar los datos de las pruebas médicas antes de llegar a una conclusión de resultado clínico. Estos tiempos reflejan la eficiencia inicial del proceso de manejo de datos, proporcionando una línea base para comparar con los tiempos post-intervención.

Tabla 42

Valores post-prueba del KPI 4

Valores post-prueba KPI 4						
1.00111	1.00111	1.00113	1.00108	1.00110	1.00081	1.00018
1.00111	1.00018	1.00114	1.00114	1.00113	1.00110	1.00062
1.00113	1.00111	1.00106	1.00114	1.00115	1.00079	1.00113
1.00058	1.00112	1.00112	1.00111	1.00055		

La tabla 42 muestra los valores de post-prueba del KPI 4, que representan el tiempo promedio de procesamiento de datos después de la implementación de la solución propuesta. Cada valor numérico (1.00111, 1.00113, 1.00108, etc.) indica el tiempo en segundos que le tomaba al modelo de Machine Learning recopilar, organizar y procesar los datos de las pruebas médicas antes de llegar a una conclusión de resultado clínico. Estos tiempos, significativamente reducidos en comparación con los valores pre-prueba, demuestran una mejora sustancial en la eficiencia del procesamiento de datos tras la intervención.

Hipótesis alternativa (Ha):

La implementación de un modelo basado en algoritmos de Machine Learning mejora el tiempo de procesamiento de los datos.

Hipótesis nula (Ho):

La implementación de un modelo basado en algoritmos de Machine Learning no mejora el tiempo de procesamiento de los datos.

a) Planteamiento de la hipótesis:

μ_1 = Corresponde a la media del tiempo de procesamiento de los datos en la pre-prueba.

μ_2 = Corresponde a la media del tiempo de procesamiento de los datos en la post-prueba.

$$H_0: \mu_1 = \mu_2 \quad H_a: \mu_1 < \mu_2$$

b) Criterios de decisión:

Si $p \geq 0.05$, aceptamos la H_0 y rechazamos la H_a .

Si $p < 0.05$, rechazamos la H_0 y rechazamos la H_a .

Tabla 43

Resultados de las pruebas de normalidad: KPI 4

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP4_Pre	,128	26	,200*	,931	26	,080
KP4_Post	,365	26	,000	,655	26	,000

Nota. Elaborado con el Software SPSS Versión 28.

*. Esto es un límite inferior de la significación verdadera

a. Corrección de significación de Lilliefors

c) Prueba estadística:

Como en cada KPI evaluado se visualiza que en el pre hay distribución normal, pero en el post es distribución no normal, entonces se aplicaría estadística no paramétrica, con Wilcoxon.

d) Decisión estadística

Tabla 44

Prueba de Wilcoxon KPI 4

Hipótesis nula	Prueba	Sig.	Decisión
----------------	--------	------	----------

1	La mediana de diferencias entre KP4_Pre y KP4_Post es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,000	Rechace la hipótesis nula.
---	--	---	------	----------------------------

Nota. Elaborado con el Software SPSS Versión 28. El nivel de significación es de ,000.

En el caso del KPI 4, se utilizó la prueba de rangos con signo de Wilcoxon para muestras emparejadas. El propósito de esta prueba fue determinar si hay una variación estadísticamente relevante en las mediciones de una variable específica, comparando los datos recogidos antes y después de una intervención o suceso particular. Esta prueba es particularmente útil cuando las mediciones se realizan en los mismos sujetos en dos momentos distintos, permitiendo una comparación directa de los cambios producidos por la intervención.

La hipótesis nula del análisis estableció que no hay diferencia en la mediana de las mediciones de KP4 antes (KP4_Pre) y después (KP4_Post) de la intervención. La mediana representa una métrica de tendencia central que señala el valor que se sitúa en el punto medio de un conjunto de datos. Tomando eso en cuenta, la hipótesis nula sugiere que, en términos de mediana, las mediciones de KP4 no se ven afectadas por la intervención; es decir, la intervención no tiene un efecto significativo sobre la mediana de KP4.

El resultado de la prueba mostró un valor de significación (Sig.) de 0,000. Este valor es una medida de probabilidad que indica cuán probable es obtener los resultados observados si la hipótesis nula fuera cierta. En la práctica estadística, un valor de p (el valor de significación) menor que 0.05 generalmente se considera como indicativo de una diferencia significativa. Un valor de p tan extremadamente bajo como 0,000 sugiere que la probabilidad de observar una diferencia como la encontrada, o mayor, si la hipótesis nula fuera cierta, es prácticamente nula. Esto proporciona una fuerte evidencia estadística en contra de la hipótesis nula.

Por lo tanto, en base a este valor de p extremadamente bajo, se toma la decisión de rechazar la hipótesis nula. Esto implica que las mediciones de KP4 muestran una diferencia significativa en su mediana antes y después de la intervención. Eso quiere decir que la intervención o evento analizado ha tenido un efecto considerable y estadísticamente significativo en la variable KP4, alterando su mediana de manera que no se puede atribuir a la variabilidad aleatoria o a las fluctuaciones ordinarias en las mediciones.

Contrastación del KPI 5: Tiempo de predicción

Este KPI evalúa la rapidez con la que la solución propuesta permite realizar predicciones tras su implementación. Comparando los tiempos de respuesta antes (pre-prueba) y después (post-prueba) de su incorporación, se busca cuantificar la mejora en la celeridad con la que se disponen las predicciones. A continuación, se presentan los tiempos registrados, resaltando la optimización en la velocidad de predicción gracias a la solución.

Tabla 45

Valores pre-prueba del KPI 5

Valores pre-prueba KPI 5						
425	120	152	324	486	407	337
587	297	582	367	162	581	403
316	346	249	135	541	449	218
359	548	233	381	283		

La tabla 45 presenta los valores de pre-prueba del KPI 5, que representan el tiempo de predicción antes de la implementación de la solución propuesta. Cada valor numérico (425, 120, 152, etc.) indica el tiempo en segundos que le tomaba al tecnólogo médico interpretar cada prueba médica y llegar a un resultado clínico, desde la finalización de la prueba hasta la entrega del resultado final. Estos tiempos

reflejan la eficiencia inicial del proceso de interpretación de resultados clínicos, proporcionando una línea base para comparar con los tiempos post-intervención.

Tabla 46

Valores post-prueba del KPI 5

Valores post-prueba KPI 5						
0.003	0.005	0.006	0.005	0.004	0.006	0.004
0.003	0.003	0.003	0.002	0.002	0.003	0.003
0.003	0.005	0.004	0.005	0.003	0.003	0.003
0.003	0.003	0.003	0.003	0.003		

La tabla 46 muestra los valores de post-prueba del KPI 5, que representan el tiempo de predicción después de la implementación de la solución propuesta. Cada valor numérico (0.003, 0.005, 0.006, etc.) indica el tiempo en segundos que le tomaba al modelo de Machine Learning interpretar cada prueba médica y llegar a un resultado clínico, desde la finalización de la prueba hasta la entrega del resultado final. Estos tiempos, significativamente reducidos en comparación con los valores pre-prueba, demuestran una mejora considerable en la eficiencia del proceso de interpretación de resultados clínicos tras la intervención.

Hipótesis alternativa (Ha):

La implementación de un modelo basado en algoritmos de Machine Learning mejora el tiempo de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C

Hipótesis nula (Ho):

La implementación de un modelo basado en algoritmos de Machine Learning no mejora el tiempo de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.

a) Planteamiento de la hipótesis:

μ_1 = Corresponde a la media del tiempo de la predicción en la pre-prueba.

μ_2 = Corresponde a la media del tiempo de la predicción en la post-prueba.

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 < \mu_2$$

b) Criterios de decisión:

Si $p \geq 0.05$, aceptamos la H_0 y rechazamos la H_a .

Si $p < 0.05$, rechazamos la H_0 y rechazamos la H_a .

Tabla 47

Resultados de las pruebas de normalidad KPI 5

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP5_Pre	,095	26	,200*	,958	26	,347
KP5_Post	,354	26	,000	,801	26	,000

Nota. Elaborado con el Software SPSS Versión 28.

*. Esto es un límite inferior de la significación verdadera

a. Corrección de significación de Lilliefors

c) Prueba estadística:

Como en cada KPI evaluado se visualiza que en el pre hay distribución normal, pero en el post es distribución no normal, entonces se aplicaría estadística no paramétrica, con Wilcoxon.

d) Decisión estadística

Tabla 48

Prueba de Wilcoxon KPI 5

	Hipótesis nula	Prueba	Sig.	Decisión
1	La mediana de diferencias entre KP5_Pre y KP5_Post es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,000	Rechace la hipótesis nula.

Nota. Elaborado con el Software SPSS Versión 28. El nivel de significación es de ,000.

En lo que respecta al KPI 5, se aplicó la prueba de rangos con signo de Wilcoxon para muestras relacionadas. Esta metodología se utilizó con la finalidad de investigar la presencia de una variación estadísticamente significativa en los valores

registrados de una variable específica, antes y después de la implementación de una intervención o acontecimiento. Esta prueba es particularmente útil cuando las mediciones se realizan en los mismos sujetos en dos momentos distintos, permitiendo una comparación directa de los cambios producidos por la intervención.

La hipótesis nula del análisis estableció que no hay diferencia en la mediana de las mediciones de KP5 antes (KP5_Pre) y después (KP5_Post) de la intervención. La mediana, como indicador de tendencia central, refleja el valor que ocupa la posición central en una serie de datos. Tomando eso en cuenta, la hipótesis nula sugiere que, en términos de mediana, las mediciones de KP5 no se ven afectadas por la intervención; es decir, la intervención no tiene un efecto significativo sobre la mediana de KP5.

El resultado de la prueba mostró un valor de significación (Sig.) de 0,000. Este valor es una medida de probabilidad que indica cuán probable es obtener los resultados observados si la hipótesis nula fuera cierta. En la práctica estadística, un valor de p (el valor de significación) menor que 0.05 generalmente se considera como indicativo de una diferencia significativa. Un valor de p tan extremadamente bajo como 0,000 sugiere que la probabilidad de observar una diferencia como la encontrada, o mayor, si la hipótesis nula fuera cierta, es prácticamente nula. Esto proporciona una fuerte evidencia estadística en contra de la hipótesis nula.

Así, debido a este valor de p excepcionalmente bajo, se procede a tomar la decisión de descartar la hipótesis nula. Esto implica que las mediciones de KP5 muestran una diferencia significativa en su mediana antes y después de la intervención. Eso quiere decir que la intervención o evento analizado ha tenido un efecto considerable y estadísticamente significativo en la variable KP5, alterando su

La tabla 49 presenta los valores de pre-prueba del KPI 6, que se obtuvieron de un cuestionario administrado a 16 personas con 10 preguntas cada una. Cada valor numérico (3.00, 2.00, 3.00, etc.) refleja las respuestas de los participantes antes de la implementación de la solución propuesta. Estos datos proporcionan una línea base para medir las percepciones y conocimientos iniciales de los participantes sobre el proceso en cuestión, permitiendo comparar con los resultados post-intervención para evaluar cambios y mejoras en las opiniones y comprensión de los involucrados.

Tabla 50

Valores post-prueba del KPI 6 (Cuestionario a 16 personas, 10 preguntas)

Valores post-prueba del KPI 6									
5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
5,00	5,00	5,00	5,00	4,00	5,00	5,00	5,00	5,00	5,00
5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
5,00	5,00	5,00	5,00	5,00	4,00	5,00	5,00	5,00	5,00
5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
4,00	4,00	5,00	4,00	3,00	5,00	4,00	5,00	4,00	4,00
1,00	1,00	1,00	2,00	1,00	1,00	1,00	1,00	1,00	1,00
1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
1,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
5,00	5,00	5,00	4,00	4,00	4,00	4,00	4,00	4,00	4,00
1,00	1,00	1,00	1,00	1,00	1,00	3,00	3,00	3,00	3,00
5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
5,00	5,00	4,00	5,00	4,00	4,00	5,00	5,00	4,00	5,00
5,00	4,00	4,00	5,00	4,00	4,00	4,00	5,00	4,00	5,00
4,00	5,00	5,00	4,00	5,00	4,00	5,00	4,00	5,00	4,00

La tabla 50 muestra los valores de post-prueba del KPI 6, obtenidos del mismo cuestionario administrado a 16 personas con 10 preguntas cada una después de la

implementación de la solución propuesta. Cada valor numérico (5.00, 4.00, 1.00, etc.) refleja las respuestas de los participantes tras la intervención, mostrando variaciones individuales que permiten evaluar cambios específicos en cada caso. Estos datos permiten comparar las percepciones y conocimientos de los participantes con respecto a los valores pre-prueba, proporcionando una visión detallada de las mejoras alcanzadas y evidenciando cualquier cambio positivo en sus opiniones y comprensión del proceso tras la implementación de la solución.

Hipótesis alternativa (Ha):

La implementación de un modelo basado en algoritmos de Machine Learning mejora el nivel de confianza en la toma de decisiones en RedLab Perú S.A.C.

Hipótesis nula (Ho):

La implementación de un modelo basado en algoritmos de Machine Learning no mejora el nivel de confianza en la toma de decisiones en RedLab Perú S.A.C.

a) Planteamiento de la hipótesis:

μ_1 = Corresponde a la media del nivel de confianza en la pre-prueba.

μ_2 = Corresponde a la media del nivel de confianza en la post-prueba.

$$H_0: \mu_1 = \mu_2 \quad H_a: \mu_1 < \mu_2$$

b) Criterios de decisión:

Si $p \geq 0.05$, aceptamos la H_0 y rechazamos la H_a .

Si $p < 0.05$, rechazamos la H_0 y rechazamos la H_a .

Tabla 51

Resultados de las pruebas de normalidad KPI 6

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KP6_Pre	,199	16	,092	,934	16	,282
KP6_Post	,346	16	,000	,657	16	,000

Nota. Elaborado con el Software SPSS Versión 28.

a. Corrección de significación de Lilliefors

c) Prueba estadística:

Como en cada KPI evaluado se visualiza que en el pre hay distribución normal, pero en el post es distribución no normal, entonces se aplicaría estadística no paramétrica, con Wilcoxon.

d) Decisión estadística

Tabla 52

Prueba de Wilcoxon KPI 6

	Hipótesis nula	Prueba	Sig.	Decisión
1	La mediana de diferencias entre TOTAL_PRE y TOTAL_POST es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,001	Rechace la hipótesis nula.

Nota. Elaborado con el Software SPSS Versión 28. El nivel de significación es de ,001.

Respecto al KPI 6, se utilizó la prueba de rangos con signo de Wilcoxon para muestras emparejadas. Esta técnica se seleccionó con el propósito de evaluar si hay un cambio estadísticamente importante en los valores medidos de una variable específica, comparando los periodos antes y después de realizar una intervención o suceso. Esta prueba es particularmente útil cuando las mediciones se realizan en los mismos sujetos en dos momentos distintos, permitiendo una comparación directa de los cambios producidos por la intervención.

La hipótesis nula del análisis estableció que no hay diferencia en la mediana de las mediciones de KP6 antes (TOTAL_PRE) y después (TOTAL_POST) de la intervención. La mediana, una métrica central de tendencia, representa el valor que se sitúa en la mitad exacta de una distribución de datos. Tomando eso en cuenta, la hipótesis nula sugiere que, en términos de mediana, las mediciones de KP6 no se ven afectadas por la intervención; es decir, la intervención no tiene un efecto significativo sobre la mediana de KP6.

El resultado de la prueba mostró un valor de significación (Sig.) de 0,001. Este valor es una medida de probabilidad que indica cuán probable es obtener los resultados observados si la hipótesis nula fuera cierta. En la práctica estadística, un valor de p (el valor de significación) menor que 0.05 generalmente se considera como indicativo de una diferencia significativa. Un valor de p tan extremadamente bajo como 0,001 sugiere que la probabilidad de observar una diferencia como la encontrada, o mayor, si la hipótesis nula fuera cierta, es prácticamente nula. Esto proporciona una fuerte evidencia estadística en contra de la hipótesis nula.

Por lo tanto, en base a este valor de p extremadamente bajo, se toma la decisión de rechazar la hipótesis nula. Esto implica que las mediciones de KP6 muestran una diferencia significativa en su mediana antes y después de la intervención. Eso quiere decir que la intervención o evento analizado ha tenido un efecto considerable y estadísticamente significativo en la variable KP6, alterando su mediana de manera que no se puede atribuir a la variabilidad aleatoria o a las fluctuaciones ordinarias en las mediciones.

CAPÍTULO IV

DISCUSIÓN

Los hallazgos de esta investigación representan un avance significativo en el análisis predictivo en el contexto clínico, especialmente con la implementación de algoritmos de aprendizaje automático. La evidencia obtenida y analizada demuestra claramente la necesidad de rechazar la hipótesis nula en todos los KPIs evaluados. Este rechazo se basa en la observación de que la mayoría de los valores p son menores a 0.05, lo cual respalda el enfoque estadístico no paramétrico utilizado y subraya la rigurosidad y precisión de nuestros métodos analíticos.

La efectividad del modelo de Machine Learning empleado se evidencia claramente en las medias óptimas alcanzadas en cada Indicador Clave de Rendimiento (KPI), destacando su capacidad para cumplir con los objetivos establecidos. En este sentido, el modelo de Árbol de Decisión ha demostrado ser significativamente más eficaz en comparación con otros modelos como Random Forest, Gradient Boosting y Regresión Logística. Este hallazgo resulta particularmente relevante al compararlo con estudios previos en el campo. Por ejemplo, investigaciones como la de Khalaji et al. (2022) subrayaron la efectividad de la Regresión Logística, mientras que Tan et al. (2021) resaltaron el desempeño superior del modelo Random Forest en escenarios específicos. Asimismo, nuestros resultados están en línea con los hallazgos de Díaz (2021), quien también empleó el modelo de Árbol de Decisión, logrando una precisión del 62.45%. Este análisis refuerza la solidez del modelo utilizado, consolidándolo como una herramienta confiable y efectiva para la resolución de problemas complejos en el ámbito del Machine Learning aplicado.

Al poner estos resultados en el contexto más amplio del análisis predictivo en la salud, se observa un avance significativo en la comprensión y aplicación de los algoritmos de Machine Learning. No obstante, es importante señalar las limitaciones

de nuestro estudio. La capacidad de generalizar nuestros resultados podría estar limitada por las características específicas de la muestra y el entorno de laboratorio en el que se realizó la investigación. Esta consideración es crucial para evitar extrapolaciones incorrectas y asegurar la relevancia clínica de nuestros hallazgos.

Mirando hacia el futuro, los resultados obtenidos abren muchas posibilidades para investigaciones adicionales y nuevas líneas de desarrollo en el ámbito del Machine Learning aplicado a la salud. Es esencial realizar estudios que exploren la aplicabilidad de estos modelos en diversos contextos clínicos y con conjuntos de datos más variados. Esto permitiría no solo validar y ampliar nuestras conclusiones sobre la eficacia del modelo de Árbol de Decisión, sino también contribuir al desarrollo de herramientas analíticas más avanzadas y personalizadas para la toma de decisiones en salud. Investigaciones futuras que incluyan una gama más amplia de variables, contextos y poblaciones podrían proporcionar una comprensión más detallada de la flexibilidad, robustez y limitaciones de los modelos de Machine Learning, mejorando su utilidad en la mejora de los resultados clínicos, la optimización de recursos y la reducción de costos en el sector salud. Estos esfuerzos serán fundamentales no solo para avanzar en el conocimiento, sino también en la implementación práctica de la inteligencia artificial, consolidándola como un pilar clave en la mejora continua de la atención médica y la gestión de la salud pública.

CAPÍTULO V
CONCLUSIONES

- Como primera conclusión, la Tabla 22 evidencia una mejora considerable en el KPI 1, que mide la tasa de exactitud, antes y después de la implementación de mejoras específicas. Se observa un aumento significativo del 74% en este indicador clave de rendimiento. Específicamente, la tasa de exactitud se incrementó de un 25,28% inicial a un impresionante 99,5% posterior a la implementación, destacando la eficacia de las estrategias adoptadas.
- Como segunda conclusión, la Tabla 23 presenta resultados destacados relacionados con el KPI 2, enfocado en la tasa de precisión. Al comparar los períodos antes y después de la implementación, se observa un incremento aproximado del 68% en este indicador. Antes de la implementación, la tasa de precisión era del 30,73%, mientras que después de la implementación, aumentó a un notable 99,35%, demostrando una mejora significativa en la precisión tras las medidas adoptadas.
- Como tercera conclusión, la Tabla 24 ilustra un análisis comparativo del KPI 3, enfocado en la tasa de sensibilidad, tanto en las fases pre como post implementación. Este indicador experimentó un incremento del 62%: inicialmente, la tasa de sensibilidad se situaba en un 36,69%, pero tras la implementación de mejoras estratégicas, se elevó a un impresionante 99,35%. Este notable aumento subraya la eficacia de las intervenciones realizadas en la mejora de la sensibilidad de la predicción.
- Como cuarta conclusión, la Tabla 25 aporta una visión comparativa del KPI 4, que mide el tiempo de procesamiento de la data, tanto en las etapas pre como post implementación. Se observó una disminución significativa de 567.04 segundos (9.45 minutos) en este KPI. Antes de la implementación, el tiempo de procesamiento era de 568,039 segundos, pero tras aplicar las mejoras, este

tiempo se redujo drásticamente, tardando ahora 1,00096 segundos. Esta notable disminución resalta la eficiencia alcanzada en el procesamiento de datos como resultado de las estrategias implementadas.

- Como quinta conclusión, presentada en la Tabla 26, se analiza el KPI 5, que mide el tiempo de predicción, comparando los períodos pre y post implementación. Se registró una reducción impresionante de 357.227 segundos (5.95 minutos) en este indicador. Inicialmente, el tiempo de predicción era de 357,2306 segundos (5.95 minutos), pero tras la implementación de mejoras significativas, este tiempo de predicción se redujo drásticamente a ahora demorar 0,0036 segundos en predecir. Esta notable disminución subraya la eficiencia mejorada en la velocidad de predicción gracias a las intervenciones realizadas.
- Como sexta conclusión, la Tabla 27 ofrece una comparativa del KPI 6, que evalúa el nivel de confianza, antes y después de la implementación. Se observó un aumento significativo en este KPI, pasando de un nivel de confianza de 23,6875 en la etapa pre implementación a 40,8125 en la etapa post implementación. Este incremento notable refleja una mejora sustancial en la confiabilidad y la certeza del modelo ML.

CAPÍTULO VI
RECOMENDACIONES

- Como recomendación principal, se sugiere la evaluación y mejora continua de los modelos de Machine Learning. Es crucial implementar un proceso regular de revisión y ajuste para asegurar que los algoritmos se mantengan actualizados con los últimos avances y datos en el ámbito médico. Esta práctica garantizará que los modelos conserven su precisión y efectividad a lo largo del tiempo.
- Como segunda recomendación, es crucial priorizar la capacitación del personal. Se recomienda implementar programas de formación continua para el equipo médico y técnico, asegurando que puedan utilizar y comprender eficientemente la tecnología de Machine Learning. Esto no solo mejorará la eficiencia, sino que también facilitará su integración en la práctica clínica diaria.
- Como tercera recomendación, se propone fortalecer la colaboración interdisciplinaria entre los equipos de tecnología, médicos y administrativos. Promover un entorno de trabajo cooperativo con comunicación abierta facilitará la integración efectiva de la tecnología de ML en los procesos clínicos y administrativos, asegurando decisiones informadas y holísticas que consideren tanto las necesidades clínicas como operativas.
- Como cuarta recomendación, se sugiere una evaluación exhaustiva de las cuestiones éticas y de privacidad en el manejo de datos clínicos. Es vital implementar protocolos rigurosos para proteger la información del paciente. Esto incluye cumplir con las normativas de protección de datos y realizar auditorías regulares para asegurar que la gestión de los datos se haga de manera ética y segura, fortaleciendo la confianza del paciente en RedLab Perú S.A.C.

REFERENCIAS

- AlShourbaji, I., Helian, N., Sun, Y., Hussien, A. G., Abualigah, L., & Elnaim, B. (2023). An efficient churn prediction model using gradient boosting machine and metaheuristic optimization. *Scientific Reports*, 13, 14441. <https://doi.org/10.1038/s41598-023-41093-6>
- Alghunaim, S., & Al-Baity, H. H. (2019). On the scalability of machine-learning algorithms for breast cancer prediction in big data context, *IEEE Access*, 7, 91535–91546. <https://doi.org/10.1109/access.2019.2927080>
- Alahmar, A., & Benlamri, R. (2020). SNOMED CT-based standardized e-clinical pathways for enabling big data analytics in healthcare, *IEEE access: practical innovations, open solutions*, 8, 1–1. <https://doi.org/10.1109/access.2020.2994286>
- Aria, M., Cuccurullo, C., & Gnasso, A. (2021). A comparison among interpretative proposals for Random Forests. *Machine Learning with Applications*, 6, 100094. <https://doi.org/10.1016/j.mlwa.2021.100094>
- Beecy, A. N., Gummalla, M., Sholle, E., Xu, Z., Zhang, Y., Michalak, K., Dolan, K., Hussain, Y., Lee, B. C., Zhang, Y., Goyal, P., Champion, T. R., Jr, Shaw, L. J., Baskaran, L., & Al'Aref, S. J. (2020). Utilizing electronic health data and machine learning for the prediction of 30-day unplanned readmission or all-cause mortality in heart failure. *Cardiovascular Digital Health Journal*, 1(2), 71–79. <https://doi.org/10.1016/j.cvdhj.2020.07.004>
- Beunza, J. (2020). *Los profesionales que quieran humanizar la atención sanitaria tendrán un gran aliado en la inteligencia artificial*. Elsevier. <https://www.elsevier.com/es-es/connect/medicina/juanjo-beunza-Manual-practico-de-inteligencia-artificial-en-entornos-sanitarios>

- Bolaños, X. (2020, 30 de abril). *Machine Learning Aplicado al Diagnóstico Médico*. Encora. <https://www.encora.com/es/blog/machine-learning-aplicado-al-diagnostico-medico>
- Cabrejos Yalán, V. M. (2022). *Las Redes Neuronales Convolucionales y la mejora en el diagnóstico de Neumonía – área de Radiología* [Tesis de maestría, Universidad Ricardo Palma]. Repositorio de la Universidad Ricardo Palma: <https://repositorio.urp.edu.pe/handle/20.500.14138/5267>
- Claudio Bautista, M. (2024). *Desarrollo de los hábitos de vida saludable en tiempos de pandemia, una propuesta basada en actividades físicas en estudiantes de educación primaria del Distrito de Llata, Huamalíes 2021* [Tesis de segunda especialidad, Universidad Nacional Hermilio Valdizán]. Repositorio de la Universidad Nacional Hermilio Valdizán. <https://repositorio.unheval.edu.pe/handle/20.500.13080/9960>
- DataSource.ai. (2023, 25 de septiembre). *Métricas de evaluación de modelos en el aprendizaje automático*. <https://www.datasource.ai/es/data-science-articles/metricas-de-evaluacion-de-modelos-en-el-aprendizaje-automatico>
- Díaz Landa, B. N. (2021). *Predicción de la Deserción Estudiantil Utilizando la Técnica de árboles de Decisión en la Escuela de Posgrado de la Universidad Nacional José Faustino Sánchez Carrión* [Tesis de maestría, Universidad Nacional José Faustino Sánchez Carrión]. Repositorio de la Universidad Nacional José Faustino Sánchez Carrión. <https://repositorio.unjfsc.edu.pe/handle/20.500.14067/5576>
- Dávalos, E., Paca, F., Aguilar, E., & Sullcahuamán, J. (2022). Inteligencia artificial y gobierno digital durante la COVID 19 en una institución prestadora de salud,

2021-2022. *Ciencia Latina Revista Científica Multidisciplinar*, 6(4), 3897–3931.

https://doi.org/10.37811/cl_rcm.v6i4.2896

Divyashree, N., & Prasad, K. S. N. (2022). Improved clinical diagnosis using predictive analytics. *IEEE Access: Practical Innovations, Open Solutions*, 10, 75158–75175. <https://doi.org/10.1109/access.2022.3190416>

Fernández, N. (2020). *El informe belmont y los principios éticos fundamentales*. Canal Jurídico. <https://www.inesem.es/revistadigital/juridico/informe-belmont/>

Fernández, R. (s.f.). *Entender el comportamiento del modelo durante el entrenamiento graficando la pérdida y la precisión*. Unipython. <https://unipython.com/entender-el-comportamiento-del-modelo-durante-el-entrenamiento-graficando-la-perdida-y-la-precision/>

Gamarra Moreno, A. H. (2020). *Modelos de aprendizaje automático basado en técnicas supervisadas para la predicción de la mortalidad perinatal en la región Junín* [Tesis doctoral, Universidad Nacional Federico Villarreal]. Repositorio de la Universidad Nacional Federico Villarreal. <https://repositorio.unfv.edu.pe/handle/20.500.13084/4355>

GeeksforGeeks. (2019). *Cost function in Logistic Regression in Machine Learning*. <https://www.geeksforgeeks.org/ml-cost-function-in-logistic-regression/>

GeeksforGeeks. (2020, 25 de agosto). *Gradient boosting in ML*. <https://www.geeksforgeeks.org/ml-gradient-boosting/>

Ghorbani, R., Ghousi, R., Makui, A., & Atashi, A. (2020). A new hybrid predictive model to predict the early mortality risk in intensive care units on a highly imbalanced dataset. *IEEE access: practical innovations, open solutions*, 8, 141066–141079. <https://doi.org/10.1109/access.2020.3013320>

- Hernández González, O. (2021). Aproximación a los distintos tipos de muestreo no probabilístico que existen. *Revista cubana de medicina general integral*, 37(3). http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21252021000300002
- Ibrahim, Z., Bean, D., Searle, T., Qian, L., Wu, H., Shek, A., Kraljevic, Z., Galloway, J., Norton, S., Teo, J., & Dobson, R. (2022). A knowledge distillation ensemble framework for predicting short- and long-term hospitalization outcomes from electronic health records data. *IEEE Journal of Biomedical and Health Informatics*, 26(1), 423–435. <https://doi.org/10.1109/JBHI.2021.3089287>
- ICCSI. (2023, 4 de julio). *Big Data, Machine Learning y Realidad Aumentada: Una revolución tecnológica*. <https://iccsi.com.ar/big-data-machine-learning-inteligencia-artificial-realidad-aumentada/>
- InteractiveChaos. (s.f.). *Entrenamiento y validación del modelo*. <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/entrenamiento-y-validacion-del-modelo>
- Jia, L., Wang, Z., Lv, S., & Xu, Z. (2022). PE_DIM: An efficient probabilistic ensemble classification algorithm for diabetes handling class imbalance missing values. *IEEE access: practical innovations, open solutions*, 10, 107459–107476. <https://doi.org/10.1109/access.2022.3212067>
- Joo, G., Song, Y., Im, H., & Park, J. (2020). Clinical implication of machine learning in predicting the occurrence of cardiovascular disease using big data (nationwide cohort data in Korea). *IEEE access: practical innovations, open solutions*, 8, 157643–157653. <https://doi.org/10.1109/access.2020.3015757>
- Khalaji, A., Behnoush, A. H., Jameie, M., Sharifi, A., Sheikhy, A., Fallahzadeh, A., Sadeghian, S., Pashang, M., Bagheri, J., Ahmadi, S. H., & Hosseini, K. (2022).

Machine learning algorithms for predicting mortality after coronary artery bypass grafting. *Frontiers in cardiovascular medicine*, 9, Article 977747. <https://doi.org/10.3389/fcvm.2022.977747>

Lbarrera. (2022). *Calidad de los datos en la sanidad: retos, limitaciones y medidas para mejorar la calidad*. Data Ladder. <https://dataladder.com/es/calidad-de-los-datos-en-la-sanidad-retos-limitaciones-y-medidas-para-mejorar-la-calidad/>

Liu, P., Jin, K., Jiao, Y., He, M. & Fei, S. (2021). Prediction of second primary lung cancer patient's survivability based on improved eigenvector centrality-based feature selection. *IEEE Access*, 9, 55663–55672. <https://doi.org/10.1109/access.2021.3063944>

Makishi Maeireizo, I., & Rodríguez López, M. E. (2020). *Machine Learning para predecir la progresión de la condición médica de pacientes COVID-19 en el hospital María Auxiliadora* [Tesis de pregrado, Universidad Científica del Sur]. Repositorio de la Universidad Científica del Sur. <https://repositorio.cientifica.edu.pe/handle/20.500.12805/1754>

MedlinePlus. (s.f.). *Cómo entender el resultado de sus pruebas de laboratorio*. <https://medlineplus.gov/spanish/pruebas-de-laboratorio/como-entender-el-resultado-de-sus-pruebas-de-laboratorio/>

Mendoza Vásquez, D., & Salazar Chávez, S. S. (2022). *Modelo Tecnológico utilizando herramientas de Machine Learning para apoyar la toma de decisiones en el Diagnóstico y Tratamiento de la Leucemia Pediátrica* [Tesis de pregrado, Universidad Peruana de Ciencias Aplicadas]. Repositorio académico de la Universidad Peruana de Ciencias Aplicadas. <https://repositorioacademico.upc.edu.pe/handle/10757/663417>

- Microsoft. (s.f.). *Modelos de predicción de riesgos para cirugías - Azure Architecture Center*. <https://learn.microsoft.com/es-es/azure/architecture/example-scenario/ai/risk-stratification-surgery>
- Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable*. <https://christophm.github.io/interpretable-ml-book/tree.html>
- Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access: practical innovations, open solutions*, 7, 81542–81554. <https://doi.org/10.1109/access.2019.2923707>
- Montisci, A., Palmieri, V., Vietri, M., Sala, S., Maiello, C., Donatelli, F., & Napoli, C. (2022). Big Data in cardiac surgery: real world and perspectives. *Journal of Cardiothoracic Surgery*, 17(1), 277. <https://doi.org/10.1186/s13019-022-02025-z>
- Mugisha, C., & Paik, I. (2022). Comparison of neural language modeling pipelines for outcome prediction from unstructured medical text notes. *IEEE access: practical innovations, open solutions*, 10, 16489–16498. <https://doi.org/10.1109/access.2022.3148279>
- Muñoz Jaramillo, V. D. (2021). *Evaluación de Modelos de Machine Learning para la Predicción de Crímenes en la Ciudad de Medellín* [Tesis de maestría, Universidad Nacional de Colombia]. Repositorio de la Universidad Nacional de Colombia. <https://repositorio.unal.edu.co/handle/unal/80976>
- Nasim, S., Almutairi, M., Munir, K., Raza, A., & Younas, F. (2022). A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics. *IEEE access: practical innovations, open solutions*, 10, 97610-97624. <https://doi.org/10.1109/access.2022.3205587>

- Nghiem, N., Atkinson, J., Nguyen, B., Tran-Duy, An. & Wilson, N. (2023). Predicting high health-cost users among people with cardiovascular disease using machine learning and nationwide linked social administrative datasets. *Health Economics Review*, 13(1), 9. <https://doi.org/10.1186/s13561-023-00422-1>
- Noor, N. S. E. M., Ibrahim, H., Lah, M. H. C., & Abdullah, J. M. (2021). Improving outcome prediction for traumatic brain injury from imbalanced datasets using RUSBoosted trees on electroencephalography spectral power. *IEEE access: practical innovations, open solutions*, 9, 121608–121631. <https://doi.org/10.1109/access.2021.3109780>
- Reyes Carrascal, L. A. (2023). *Diseño de un sistema lector de capilares sanguíneos para determinar hematocrito usando visión artificial, 2023* [Tesis de maestría, Universidad Nacional de Cajamarca]. Repositorio de la Universidad Nacional de Cajamarca. <https://repositorio.unc.edu.pe/handle/20.500.14074/5739>
- Salas Ojeda, M. I. (2021). *Algoritmos de aprendizaje automático para predecir la curación de malformaciones arteriovenosas cerebrales tratadas con radiocirugía* [Tesis de maestría, Universidad Ricardo Palma]. Repositorio de la Universidad Ricardo Palma. <https://repositorio.urp.edu.pe/handle/20.500.14138/4283>
- Serpa, Y. (2020, 19 de mayo). *What are classification metrics?* Towards Data Science. <https://towardsdatascience.com/the-illustrated-guide-to-classification-metrics-the-basics-cf3c2e9b89b2>
- Song, Y. X., Yang, X. D., Luo, Y. G., Ouyang, C. L., Yu, Y., Ma, Y. L., Li, H., Lou, J. S., Liu, Y. H., Chen, Y. Q., Cao, J. B., & Mi, W. D. (2023). Comparison of logistic regression and machine learning methods for predicting postoperative delirium

- in elderly patients: A retrospective study. *CNS Neuroscience & Therapeutics*, 29(1), 158–167. <https://doi.org/10.1111/cns.13991>
- Srimekha, B. C., Naveen Raj, R., & Mayya, V. (2022). A comprehensive machine learning based pipeline for an accurate early prediction of sepsis in ICU. *IEEE access: practical innovations, open solutions*, 10, 105120–105132. <https://doi.org/10.1109/access.2022.3210575>
- Statista. (2019). *El gasto en salud alrededor del mundo*. <https://es.statista.com/grafico/18967/gasto-en-salud-per-capita-en-el-mundo/>
- Statista. (s.f.). *Global total medtech growth per year 2010-2022*. <https://www.statista.com/statistics/329048/worldwide-medical-technology-growth-per-year/>
- Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Mueller, K. R. (2020). *Towards CRISP-ML(Q): A machine learning process model with quality assurance methodology*. arXiv. <https://doi.org/10.48550/arXiv.2003.05155>
- Tan, T. H., Hsu, C. C., Chen, C. J., Hsu, S. L., Liu, T. L., Lin, H. J., Wang, J. J., Liu, C. F. & Huang, C. C. (2021). Predicting outcomes in older ED patients with influenza in real time using a big data-driven and machine learning approach to the hospital information system. *BMC Geriatrics*, 21(1), 280. <https://doi.org/10.1186/s12877-021-02229-3>
- Torres.ai. (s.f.). *Datos y Overfitting*. <https://torres.ai/datos-y-overfitting-keras-tensorflow/>
- Vizcaíno Zúñiga, P. I., Cedeño Cedeño, R. J., & Maldonado Palacios, I. A. (2023). Metodología de la investigación científica: guía práctica. *Ciencia Latina Revista*

https://doi.org/10.37811/cl_rcm.v7i4.7658

Wang, Y., Zhu, K, Li, Ya., Lv, Q., Fu, G., & Zhang, W. (2020). A machine learning-based approach for the prediction of periprocedural myocardial infarction by using routine data. *Cardiovascular Diagnosis and Therapy*, 10(5), 1313-1324. <https://doi.org/10.21037/cdt-20-551>

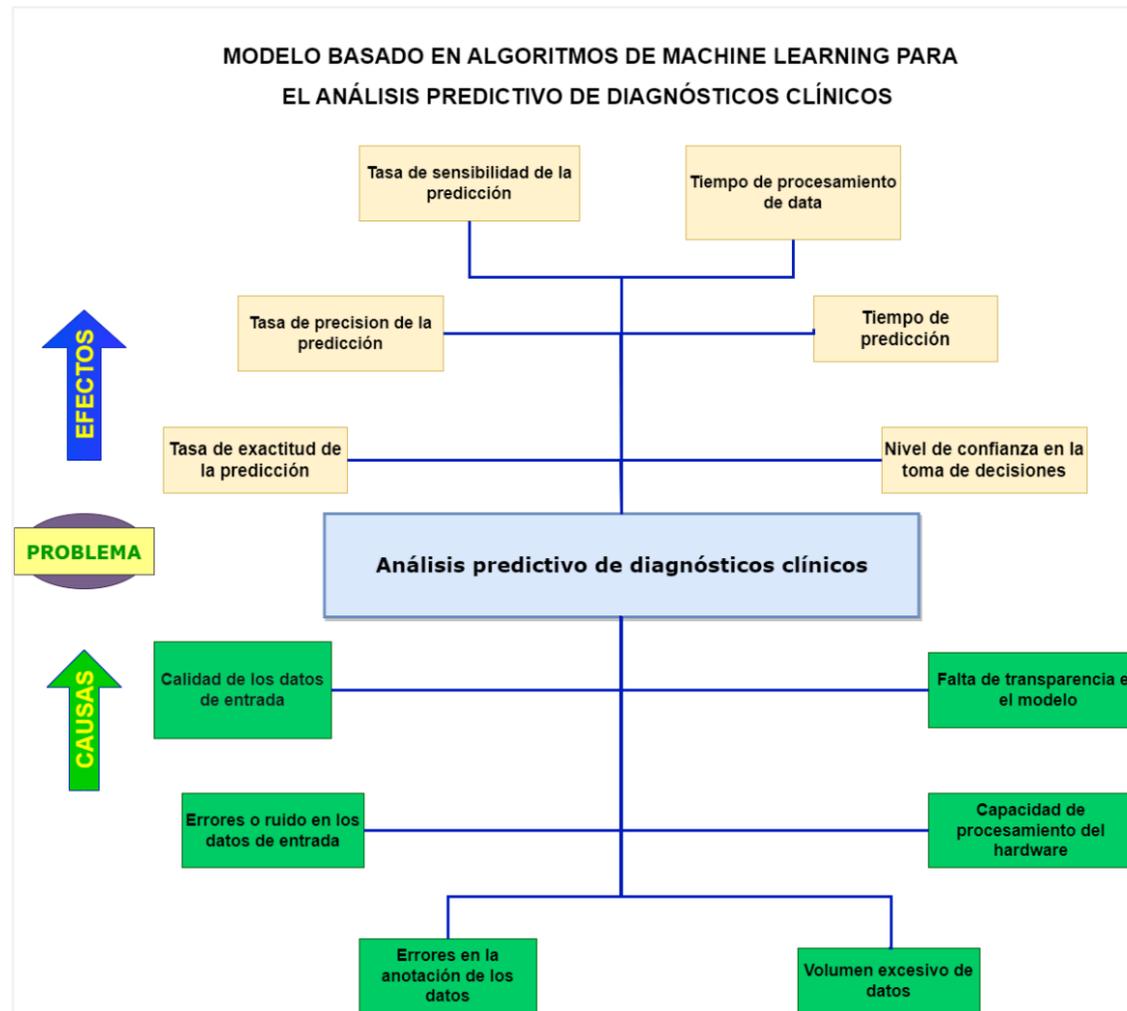
Yehoshua, R. (2023, 25 de marzo). *Random forests - Dr. Roi Yehoshua*. Medium. <https://medium.com/@roiyeho/random-forests-98892261dc49>

Zhang, D., Li, Y., Kalbaugh, C., Shi, L., Divers, J., Islam, S., & Annex, B. (2022). Machine learning approach to predict in-hospital mortality in patients admitted for peripheral artery disease in the united states. *Journal of the American Heart Association*, 11(20). <https://doi.org/10.1161/jaha.122.026987>

Zhang, P.-I., Hsu, C.-C., Kao, Y., Chen, C.-J., Kuo, Y.-W., Hsu, S.-L., Liu, T.-L., Lin, H.-J., Wang, J.-J., Liu, C.-F., & Huang, C.-C. (2020). Real-time AI prediction for major adverse cardiac events in emergency department patients with chest pain. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 28(1). <https://doi.org/10.1186/s13049-020-00786-x>

ANEXOS

Anexo 1: Árbol del problema



Anexo 2: Matriz de consistencia

Problemas	Objetivos	Hipótesis	Variables	Metodología
General			Independiente	
¿En qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora el análisis predictivo de diagnósticos clínicos en RedLab Perú S.A.C.?	Determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora el análisis predictivo de diagnósticos clínicos en RedLab Perú S.A.C.	La implementación de un modelo basado en algoritmos de Machine Learning mejora el análisis predictivo de diagnósticos clínicos en RedLab Perú S.A.C.	Modelo basado en algoritmos de Machine Learning (X)	Tipo de investigación: Aplicada Nivel de investigación: Explicativa Diseño de investigación: Pre-experimental Técnicas: - Observación - Encuesta Instrumentos: - Ficha de registro - Cuestionario
Específicos			Dependiente	
¿En qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de exactitud de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.?	Determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de exactitud de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.	La implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de exactitud de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.	Análisis predictivo de diagnósticos clínicos (Y)	
¿En qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de precisión de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.?	Determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de precisión de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.	La implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de precisión de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.		
¿En qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de sensibilidad de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.?	Determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de sensibilidad de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.	La implementación de un modelo basado en algoritmos de Machine Learning mejora la tasa de sensibilidad de las predicciones de los diagnósticos clínicos en RedLab Perú S.A.C.		
¿En qué medida la implementación de un modelo basado en algoritmos de Machine Learning reduce el tiempo requerido para procesar datos con fines diagnósticos en RedLab Perú S.A.C.?	Determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning reduce el tiempo requerido para procesar datos con fines diagnósticos en RedLab Perú S.A.C.	La implementación de un modelo basado en algoritmos de Machine Learning reduce el tiempo requerido para procesar datos con fines diagnósticos en RedLab Perú S.A.C.		
¿En qué medida la implementación de un modelo basado en algoritmos de Machine Learning disminuye el tiempo requerido para predecir diagnósticos clínicos en RedLab Perú S.A.C.?	Determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning disminuye el tiempo requerido para predecir diagnósticos clínicos en RedLab Perú S.A.C.	La implementación de un modelo basado en algoritmos de Machine Learning disminuye el tiempo requerido para predecir diagnósticos clínicos en RedLab Perú S.A.C.		
¿En qué medida la implementación de un modelo basado en algoritmos de Machine Learning incrementa el nivel de confianza en la toma de decisiones relacionadas con diagnósticos clínicos en RedLab Perú S.A.C.?	Determinar en qué medida la implementación de un modelo basado en algoritmos de Machine Learning incrementa el nivel de confianza en la toma de decisiones relacionadas con diagnósticos clínicos en RedLab Perú S.A.C.	La implementación de un modelo basado en algoritmos de Machine Learning incrementa el nivel de confianza en la toma de decisiones relacionadas con diagnósticos clínicos en RedLab Perú S.A.C.		

Anexo 3: Matriz de operacionalización

Variables (x,y)	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Tipo	Unidad de medida	Índice	Técnica	Instrumento	Fuente verificación
<p>Modelo basado en algoritmos de Machine Learning (X)</p> <p>Independiente</p>	<p>Un modelo sustentado en algoritmos de Machine Learning hace referencia a un sistema diseñado para analizar y aprender de grandes conjuntos de datos con el propósito de hacer predicciones precisas. Estos modelos pasan por diversas etapas críticas, como el preprocesamiento de datos, la identificación de características relevantes y su validación. Por ello, es fundamental que cada fase se ejecute meticulosamente para que el modelo pueda proporcionar resultados confiables y contribuir eficazmente a la toma de decisiones. (Wang, Y et al, 2020)</p>	<p>Cuando es NO, es porque no existe el modelo basado en algoritmos de Machine Learning en la empresa RedLab Perú S.A.C, y aún se estaría en la situación actual del problema.</p> <p>Cuando es SÍ es cuando se aplicó la solución (Modelo ML) y se espera obtener mejores resultados.</p>	PRESENCIA - AUSENCIA				SI-NO			
<p>Análisis predictivo de diagnósticos clínicos (Y)</p> <p>Dependiente</p>	<p>El análisis predictivo en el contexto clínico implica la utilización de algoritmos avanzados para estimar resultados asociados con la salud del paciente. Su finalidad principal es potenciar y respaldar las decisiones médicas. Para ello, se examinan conjuntos de datos clínicos que, tras ser procesados, son validados con base en diferentes métricas de rendimiento. Dichas métricas evalúan no solo la precisión de las predicciones, sino también la rapidez de la implementación de estas técnicas. Al integrar el análisis predictivo con sistemas de apoyo a decisiones, se incrementa su valor, permitiendo análisis más profundos que pueden orientar acciones médicas específicas. (Divyashree y Nandini, 2022)</p>	<p>La eficacia del modelo clínico se evalúa en base a tres métricas: exactitud, que debe superar el 85% para ser óptima; precisión; y sensibilidad.</p> <p>Además, se monitorizan dos tiempos críticos: el necesario para procesar los datos y el requerido para realizar la predicción diagnóstica. Ambos deben ser menores a 5 segundos.</p> <p>Finalmente, la confianza en el modelo se mide utilizando una escala de Likert, con respuestas que varían desde TCO (Totalmente conforme) hasta TDCO (Totalmente disconforme).</p>	<p>Métricas de rendimiento</p> <p>Tasa de exactitud de la predicción</p> <p>Tasa de precisión de la predicción</p> <p>Tasa de sensibilidad de la predicción</p>	<p>Proporción / Porcentaje (%)</p> <p>>85%</p>	<p>Cuantitativo</p>	<p>Segundos</p> <p>< 5s</p>	<p>Observación</p>	<p>Ficha de registro</p>	<p>Reportes de Software</p> <p>Cronómetro</p>	
			<p>Toma de decisiones</p>	<p>Nivel de confianza en la toma de decisiones</p>	<p>Ordinal</p>	<p>Escala de Likert</p>	<p>[Totalmente conforme, Bastante conforme, Moderadamente conforme, Poco conforme, Totalmente disconforme]</p>	<p>Encuesta</p>	<p>Cuestionario</p>	<p>Análisis de encuesta</p>

Anexo 4: Matriz de revisión

Mendeley Reference Manager

Mendeley Reference Manager File Edit Tools Help

Library | Notebook

Anderson Fuentes Fuentes

All References / Art. para sistematizar

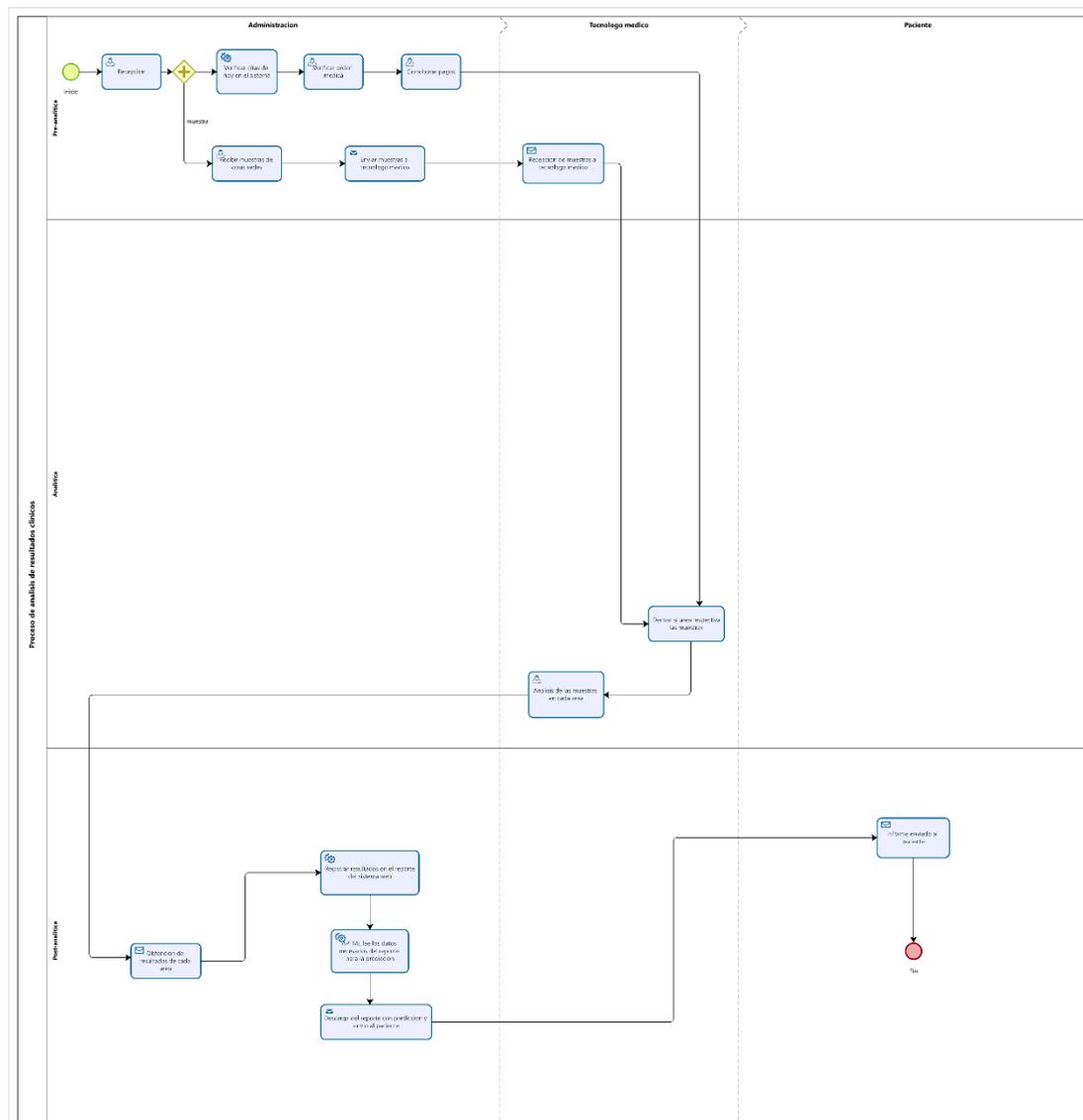
Search Filters

<input checked="" type="checkbox"/>	AUTHORS	YEAR	TITLE	SOURCE	ADDED	FILE
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Ding L, Zhang C, Wang K, ...	2023	A machine learning-based model for predicting the risk of early-stage inguinal lymph ...	Frontiers in Surgery	5/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Zhou S, Jv D, Meng X, Zha...	2023	Feasibility of machine learning-based modeling and prediction using multiple centers...	Annals of Medicine	5/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Alourani A, Tariq K, Tahir M...	2023	Patient Mortality Prediction and Analysis of Health Cloud Data Using a Deep Neural ...	Applied Sciences 20...	5/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> van Dijk L, Mohamed A, Ah...	2023	Head and neck cancer predictive risk estimator to determine control and therapeutic ...	European Journal of...	5/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Zhang Y, Yi X, Tang Z, Xie ...	2023	Using machine learning to predict lymph node metastasis in patients with renal cell c...	Frontiers in public h...	5/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Chen C, Chen B, Yang J, Li...	2023	Development and validation of a practical machine learning model to predict sepsis ...	Annals of Medicine	5/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Moradi H, Timothy Bunnell ...	2023	Assessing the effects of therapeutic combinations on SARS-CoV-2 infected patient o...	PLoS ONE	5/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Nghiem N, Atkinson J, Ngu...	2023	Predicting high health-cost users among people with cardiovascular disease using m...	Health Economics R...	5/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Abdellatif A, Abdellatef H, K...	2022	Improving the Heart Disease Detection and Patients' Survival Using Supervised Infin...	IEEE Access	1/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Liu P, Jin K, Jiao Y, He M, F...	2021	Prediction of Second Primary Lung Cancer Patient's Survivability Based on Improve...	IEEE Access	1/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Chicco D, Jurman G	2021	Arterial Disease Computational Prediction and Health Record Feature Ranking amo...	IEEE Access	1/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Nasim S, Almutairi M, Munir...	2022	A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learnin...	IEEE Access	1/5/2023	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Ishaq A, Sadiq S, Umer M, ...	2021	Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effect...	IEEE Access	1/5/2023	

101 references selected

Organize Mark as Export Delete

Anexo 6: Proceso TO-BE



Anexo 7: Carta de autorización de la empresa



CARTA N°: CRLP23 - 002

ASUNTO: **AUTORIZACIÓN PARA EL DESARROLLO DE TESIS.**

Lima, 20 de Junio del 2023

Señores

UNIVERSIDAD AUTÓNOMA DEL PERÚ

PRESENTE. -

De nuestra consideración

Es grato dirigirme a ustedes en calidad de Gerente General de RedLab Perú S.A.C con RUC N° 20601880564, para hacer de su conocimiento que los señores **Anderson Alfredo Fuentes Pinedo** identificado con DNI N° **75120338** y **Carlos Emmanuel Navarro Caraza** identificado con DNI N° **77352026**, estudiantes de la carrera profesional de Ingeniería de Sistemas de vuestra institución Universidad Autónoma del Perú, han sido admitidos para realizar la investigación de "Implementación de un modelo basado en algoritmos de Machine Learning para el análisis predictivo de diagnósticos clínicos en RedLab Peru S.A.C" en el área administrativa de nuestra empresa, con fines académicos para el desarrollo de tesis.

Se expide la presente carta de autorización, para los fines pertinentes.



RedLab PERÚ S.A.C.
RUC: 20601880564
PERCY URETA SIERRA
GERENTE GENERAL

Lic. Percy Ureta Sierra

Gerente General de RedLab Perú S.A.C.

DNI: 10420261

Anexo 8: Instrumento de observación - Métricas de rendimiento
Ficha de registro

I1: Tasa de exactitud de la predicción

Investigadores	Fuentes Pinedo, Anderson Alfredo
	Navarro Caraza, Carlos Enmanuel
Empresa	Redlab Perú S.A.C.
Dirección	JR. CAMANA NRO. 780 INT. 102B CERCADO DE LIMA - LIMA
Proceso	Métrica accuracy - Modelo elegido (Decision Tree - DT)
Indicador	Tasa de exactitud de la predicción
Fórmula	$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100$
Fecha	07/11/2023

Matriz de confusión:

		Predicción	
		Positivo	Negativo
Observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Positive	(TN) True Negative

Ítem	Prueba médica	Pre	TP	TN	FP	FN	Post
1	VB12 II	7.49%	9066	9066	0	0	100%
2	Anti-TPO	15.27%	4886	4886	0	0	100%
3	Anti-Tg	34.39%	4824	4824	0	0	100%
4	CEA	12.17%	5187	5187	0	0	100%
5	CA19-9	23.09%	4924	4924	0	0	100%
6	CA125	9.33%	5049	5049	0	0	100%
7	AFP	28.26%	5046	5046	0	0	100%
8	25-OH VitD	22.99%	5125	5125	7	7	100%
9	CA15-3	17.30%	4759	4759	0	0	100%

10	BNP	31.89%	4689	4689	0	0	100%
11	Anti-TP	45.32%	4870	4870	0	0	100%
12	Troponin-I	7.26%	4072	4072	11	11	100%
13	HIV Combo	16.55%	4960	4960	0	0	100%
14	HBsAg	55.22%	4872	4872	15	15	100%
15	ANA	12.49%	5427	5427	0	0	100%
16	Anti-HBc	31.37%	4954	4954	0	0	100%
17	Anti-HCV	19.63%	5196	10392	0	0	100%
18	Anti-CCP	39.18%	4752	9510	6	6	100%
19	E2	59.95%	5011	15615	291	291	97%
20	ACTH	15.30%	4036	24216	0	0	100%
21	Free PSA	24.66%	5337	5337	181	181	97%
22	TPSA	20.30%	5776	5776	236	236	96%
23	Ferritin	37.91%	6208	31100	15	15	100%
24	Testosterone	20.04	4385	13449	147	147	98%
25	HCG	14.36	4890	19578	6	6	100%
26	Prolactin	35.45	4960	20068	76	76	99%

KPI 1 - Tasa de Exactitud (Accuracy): En el PRE, se midió la exactitud de los resultados clínicos previos comparándolos con un resultado de referencia considerado correcto. Esto implicó revisar registros históricos para determinar cuántos de los resultados iniciales coincidían con los resultados de seguimiento o confirmaciones posteriores.

**Anexo 9: Instrumento de observación - Métricas de rendimiento
Ficha de registro**

I2: Tasa de precisión de la predicción

Investigadores	Fuentes Pinedo, Anderson Alfredo
	Navarro Caraza, Carlos Enmanuel
Empresa	Redlab Perú S.A.C.
Dirección	JR. CAMANA NRO. 780 INT. 102B CERCADO DE LIMA - LIMA
Proceso	Métrica de precisión - Modelo elegido (Decision Tree - DT)
Indicador	Tasa de precisión de la predicción
Fórmula	$pre = \left(\frac{TP}{TP + FP} \right) \times 100$
Fecha	07/11/2023

Matriz de confusión:

		Predicción	
		Positivo	Negativo
Observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Positive	(TN) True Negative

Ítem	Prueba médica	Pre	TP	FP	Post
1	VB12 II	8.70%	9066	0	100%
2	Anti-TPO	36.57%	4886	0	100%
3	Anti-Tg	45.54%	4824	0	100%
4	CEA	16.10%	5187	0	100%
5	CA19-9	32.59%	4924	0	100%
6	CA125	3.55%	5049	0	100%
7	AFP	26.80%	5046	0	100%
8	25-OH VitD	19.97%	5125	7	100%
9	CA15-3	20.30%	4759	0	100%

10	BNP	43.58%	4689	0	100%
11	Anti-TP	4.00%	4870	0	100%
12	Troponin-I	12.26%	4072	11	100%
13	HIV Combo	14.06%	4960	0	100%
14	HBsAg	82.12%	4872	15	100%
15	ANA	4.65%	5427	0	100%
16	Anti-HBc	60.29%	4954	0	100%
17	Anti-HCV	12.86%	5196	0	100%
18	Anti-CCP	34.54%	4752	6	100%
19	E2	58.56%	5011	291	95%
20	ACTH	19.71%	4036	0	100%
21	Free PSA	34.14%	5337	181	97%
22	TPSA	27.10%	5776	236	96%
23	Ferritin	38.68%	6208	15	100%
24	Testosterone	33.93%	4385	147	97%
25	HCG	39.93%	4890	6	100%
26	Prolactin	68.45%	4960	76	98%

KPI 2 - Tasa de Precisión: De manera similar al KPI 1, se analizó la proporción de resultados clínicos previos que, ya fueran positivos o negativos, efectivamente correspondieron a casos confirmados, en comparación con la totalidad de resultados entregados.

Anexo 10: Instrumento de observación - Métricas de rendimiento
Ficha de registro

I3: Tasa de sensibilidad de la predicción

Investigadores	Fuentes Pinedo, Anderson Alfredo
	Navarro Caraza, Carlos Enmanuel
Empresa	Redlab Perú S.A.C.
Dirección	JR. CAMANA NRO. 780 INT. 102B CERCADO DE LIMA - LIMA
Proceso	Métrica sensibilidad - Modelo elegido (Decision Tree - DT)
Indicador	Tasa de sensibilidad de la predicción
Fórmula	$TPR = \left(\frac{TP}{TP + FN} \right) \times 100$
Fecha	07/11/2023

Matriz de confusión:

		Predicción	
		Positivo	Negativo
Observación	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Positive	(TN) True Negative

Ítem	Prueba médica	Pre	TP	FN	Post
1	VB12 II	4.82%	9066	0	100%
2	Anti-TPO	18.18%	4886	0	100%
3	Anti-Tg	40.67%	4824	0	100%
4	CEA	21.59%	5187	0	100%
5	CA19-9	26.59%	4924	0	100%
6	CA125	4.90%	5049	0	100%
7	AFP	46.67%	5046	0	100%
8	25-OH VitD	58.06%	5125	7	100%
9	CA15-3	32.73%	4759	0	100%

10	BNP	25.95%	4689	0	100%
11	Anti-TP	92.47%	4870	0	100%
12	Troponin-I	6.37%	4072	11	100%
13	HIV Combo	45.42%	4960	0	100%
14	HBsAg	50.47%	4872	15	100%
15	ANA	3.51%	5427	0	100%
16	Anti-HBc	35.29%	4954	0	100%
17	Anti-HCV	83.33%	5196	0	100%
18	Anti-CCP	88.02%	4752	6	100%
19	E2	85.08%	5011	291	95%
20	ACTH	19.34%	4036	0	100%
21	Free PSA	22.41%	5337	181	97%
22	TPSA	23.84%	5776	236	96%
23	Ferritin	46.47%	6208	15	100%
24	Testosterone	26.62%	4385	147	97%
25	HCG	12.21%	4890	6	100%
26	Prolactin	32.90%	4960	76	98%

KPI 3 - Tasa de Sensibilidad: Para el PRE, se evaluó cuántos de los casos reales (resultados clínicos) fueron correctamente clasificados por los tecnólogos médicos, utilizando las pruebas médicas y rangos referenciales disponibles en ese momento.

**Anexo 11: Instrumento de observación - Tiempo
Ficha de registro**

I4: Tiempo requerido para procesar data

Investigadores	Fuentes Pinedo, Anderson Alfredo
	Navarro Caraza, Carlos Enmanuel
Empresa	Redlab Perú S.A.C.
Dirección	JR. CAMANA NRO. 780 INT. 102B CERCADO DE LIMA - LIMA
Proceso	Procesamiento de datos en el entrenamiento con el modelo elegido (Decision Tree - DT)
Indicador	Duración del procesamiento de datos
Fórmula	$T_{final} - T_{inicial}$
Periodo - Fecha	07/11/2023

Ítem	Prueba médica	Pre	T_inicial	T_final	Post
1	VB12 II	578	03:42:09	03:42:10.	1.00111
2	Anti-TPO	579	03:42:11.	03:42:12.	1.00111
3	Anti-Tg	606	03:42:12.	03:42:13.	1.00113
4	CEA	872	03:42:30	03:42:31	1.00108
5	CA19-9	369	03:42:31.	03:42:32.	1.00110
6	CA125	458	03:42:32.	03:42:33.	1.00081
7	AFP	587	03:42:33.	03:42:34.	1.00018
8	25-OH VitD	472	03:42:42.	03:42:43.	1.00111
9	CA15-3	392	03:42:47.	03:42:48.	1.00018
10	BNP	878	03:42:48.	03:42:49.	1.00114
11	Anti-TP	542	03:42:50.	03:42:51.	1.00114
12	Troponin-I	816	03:42:39.	03:42:40.	1.00113
13	HIV Combo	603	03:42:08.	03:42:09.	1.00110
14	HBsAg	439	03:42:19.	03:42:20.	1.00062

15	ANA	701	03:42:20.	03:42:21.	1.00113
16	Anti-HBc	623	03:42:28.	03:42:29.	1.00111
17	Anti-HCV	451	03:42:29.	03:42:30.	1.00106
18	Anti-CCP	508	03:42:37.	03:42:38.	1.00114
19	E2	765	03:42:13.	03:42:14.	1.00115
20	ACTH	471	03:42:46.	03:42:47.	1.00079
21	Free PSA	567	03:42:05.	03:42:06.	1.00113
22	TPSA	633	03:42:35.	03:42:36.	1.00058
23	Ferritin	387	03:42:17.	03:42:18.	1.00112
24	Testosterone	414	03:42:22.	03:42:23.	1.00112
25	HCG	409	03:42:18	03:42:19	1.00111
26	Prolactin	649	03:42:21	03:42:22	1.00055

KPI 4 - Tiempo de Procesamiento de Datos: En el KPI 4 se midió el tiempo promedio que le tomaba al personal correspondiente (por ejemplo, tecnólogos médicos) recopilar, organizar y procesar los datos de las pruebas médicas antes de llegar a una conclusión de resultado clínico.

**Anexo 12: Instrumento de observación - Tiempo
Ficha de registro**

I5: Tiempo requerido para predecir

Investigadores	Fuentes Pinedo, Anderson Alfredo
	Navarro Caraza, Carlos Enmanuel
Empresa	Redlab Perú S.A.C.
Dirección	JR. CAMANA NRO. 780 INT. 102B CERCADO DE LIMA - LIMA
Proceso	Predicción de diagnósticos del modelo elegido (Decision Tree - DT)
Indicador	Duración de la predicción
Fórmula	$T_{final} - T_{inicial}$
Periodo - Fecha	07/11/2023

Ítem	Prueba médica	Pre	T_inicial	T_final	Post
1	VB12 II	425	04:33:59.954	04:33:59.957	0.003
2	Anti-TPO	120	04:33:59.973	04:33:59.978	0.005
3	Anti-Tg	152	04:33:59.984	04:33:59.990	0.006
4	CEA	324	04:34:00.118	04:34:00.123	0.005
5	CA19-9	486	04:34:00.127	04:34:00.131	0.004
6	CA125	407	04:34:00.135	04:34:00.141	0.006
7	AFP	337	04:34:00.145	04:34:00.149	0.004
8	25-OH VitD	587	04:34:00.210	04:34:00.213	0.003
9	CA15-3	297	04:34:00.244	04:34:00.247	0.003
10	BNP	582	04:34:00.250	04:34:00.253	0.003
11	Anti-TP	367	04:34:00.264	04:34:00.266	0.002
12	Troponin-I	162	04:34:00.196	04:34:00.198	0.002
13	HIV Combo	581	04:33:59.946	04:33:59.949	0.003

14	HBsAg	403	04:34:00.035	04:34:00.038	0.003
15	ANA	316	04:34:00.042	04:34:00.045	0.003
16	Anti-HBc	346	04:34:00.100	04:34:00.105	0.005
17	Anti-HCV	249	04:34:00.109	04:34:00.114	0.004
18	Anti-CCP	135	04:34:00.181	04:34:00.186	0.005
19	E2	541	04:33:59.994	04:33:59.997	0.003
20	ACTH	449	04:34:00.237	04:34:00.240	0.003
21	Free PSA	218	04:33:59.926	04:33:59.930	0.003
22	TPSA	359	04:34:00.167	04:34:00.170	0.003
23	Ferritin	548	04:34:00.022	04:34:00.025	0.003
24	Testosterone	233	04:34:00.056	04:34:00.059	0.003
25	HCG	381	04:34:00.029	04:34:00.032	0.003
26	Prolactin	283	04:34:00.049	04:34:00.052	0.003

KPI 5 - Tiempo de Predicción: Este fue el tiempo que le tomaba al tecnólogo médico interpretar cada prueba médica y llegar a un resultado clínico. Esto incluía desde el momento en que la prueba se completaba hasta que se entregaba el resultado clínico final.

Anexo 13: Cuestionario - Toma de decisiones clínicas

I6: Nivel de confianza en la toma de decisiones

Investigadores	Fuentes Pinedo, Anderson Alfredo
	Navarro Caraza, Carlos Enmanuel
Empresa	Redlab Perú S.A.C.
Dirección	JR. CAMANA NRO. 780 INT. 102B CERCADO DE LIMA - LIMA
Indicador	Nivel de confianza en la toma de decisiones
Nombres y apellidos	
Fecha	

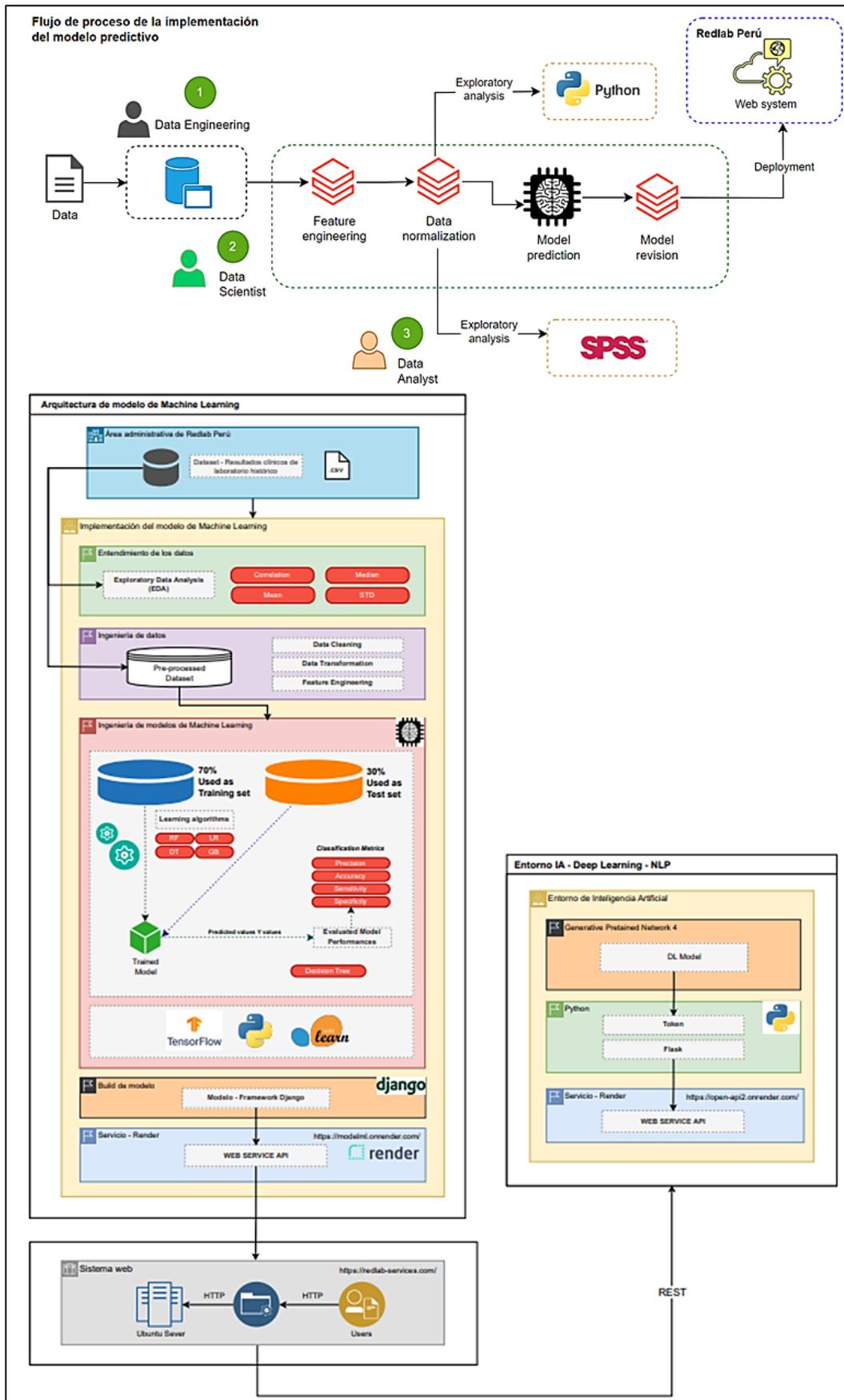
Puntaje	(1)	(2)	(3)	(4)	(5)
	Totalmente disconforme	Poco conforme	Moderadamente conforme	Bastante conforme	Totalmente conforme

Califique su nivel de confianza en relación a la toma de decisiones, sobre los procesos utilizados para la determinación de los diagnósticos clínicos en los informes emitidos a los pacientes.

Nro.	Preguntas	Valor				
		1	2	3	4	5
Dimensión - Métricas de rendimiento						
1	¿Qué tan conforme se siente con la exactitud de los diagnósticos actuales en el laboratorio clínico?					
2	¿Qué tan conforme se siente con la precisión de los resultados obtenidos de los diagnósticos clínicos?					
3	Basándose en su experiencia, ¿qué tan conforme se siente con la sensibilidad (habilidad del diagnóstico para identificar acertadamente las condiciones reales) al comunicar los hallazgos clínicos?					
Dimensión - Tiempo						
4	¿Qué tan conforme se siente con el tiempo que generalmente se toma para preparar y procesar					

	los datos antes de un diagnóstico?					
5	¿Qué tan conforme se siente con el tiempo que se dedica a la recopilación y preparación de datos para el diagnóstico?					
6	¿Qué tan conforme se siente con el intervalo de tiempo entre el procesamiento de datos y la obtención de un diagnóstico?					
7	¿Qué tan conforme se siente con el tiempo que se tarda desde que se inicia el análisis hasta que se comunica un diagnóstico clínico?					
Dimensión - Toma de decisiones						
8	¿Qué tan conforme se siente con la claridad de los diagnósticos clínicos que obtiene en el laboratorio, en cuanto a su influencia en su nivel de confianza para tomar decisiones?					
9	Al revisar los resultados de laboratorio, ¿qué tan conforme se siente con la información proporcionada para la toma de decisiones?					
10	Basado en los resultados que obtiene del laboratorio, ¿qué tan conforme se siente respecto a su capacidad de influir en decisiones posteriores?					

Anexo 14: Arquitectura de aplicación



Anexo 15: Validación de instrumentos - Experto 1

CERTIFICADO DE VALIDEZ DE CONTENIDO DE INSTRUMENTOS A TRAVÉS DE JUICIO DE EXPERTO

Título de la investigación	IMPLEMENTACIÓN DE UN MODELO BASADO EN ALGORITMOS DE MACHINE LEARNING PARA EL ANÁLISIS PREDICTIVO DE DIAGNÓSTICOS CLÍNICOS EN REDLAB PERÚ S.A.C
Nombre(s) del(os) instrumento(s)	<ul style="list-style-type: none"> Fichas de registro Cuestionario
Autor(es) del instrumento	<ul style="list-style-type: none"> Fuentes Pinedo, Anderson Alfredo Navarro Caraza, Carlos Enmanuel

N°	DIMENSIONES / Indicadores	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		SI	NO	SI	NO	SI	NO	
DIMENSIÓN 1: MÉTRICAS DE RENDIMIENTO								
1	Tasa de exactitud de la predicción	X		X		X		
2	Tasa de precisión de la predicción	X		X		X		
3	Tasa de sensibilidad de la predicción	X		X		X		
DIMENSIÓN 2: TIEMPO								
1	Tiempo requerido para procesar datos	X		X		X		
2	Tiempo requerido para predecir el diagnóstico clínico	X		X		X		
DIMENSIÓN 3: TOMA DE DECISIONES								
1	¿Qué tan conforme se siente con la exactitud de los diagnósticos actuales en el laboratorio clínico?	X		X		X		
2	¿Qué tan conforme se siente con la precisión de los resultados obtenidos de los diagnósticos clínicos?	X		X		X		
3	Basándose en su experiencia, ¿qué tan conforme se siente con la sensibilidad (habilidad del diagnóstico para identificar acertadamente las condiciones reales) al comunicar los hallazgos clínicos?	X		X		X		
4	¿Qué tan conforme se siente con el tiempo que generalmente se toma para preparar y procesar los datos antes de un diagnóstico?	X		X		X		
5	¿Qué tan conforme se siente con el tiempo que se dedica a la recopilación y preparación de datos para el diagnóstico?	X		X		X		
6	¿Qué tan conforme se siente con el intervalo de tiempo entre el procesamiento de datos y la obtención de un diagnóstico?	X		X		X		
7	¿Qué tan conforme se siente con el tiempo que se tarda desde que se inicia el análisis hasta que se comunica un diagnóstico clínico?	X		X		X		
8	¿Qué tan conforme se siente con la claridad de los diagnósticos clínicos que obtiene en el laboratorio, en cuanto a su influencia en su nivel de confianza para tomar decisiones?	X		X		X		
9	Al revisar los resultados de laboratorio, ¿qué tan conforme se siente con la información proporcionada para la toma de decisiones?	X		X		X		
10	Basado en los resultados que obtiene del laboratorio, ¿qué tan conforme se siente respecto a su capacidad de influir en decisiones posteriores?	X		X		X		

Observaciones (precisar si hay suficiencia): Si existe suficiencia

Opinión de aplicabilidad: Aplicable Aplicable después de corregir No aplicable

Apellidos y nombres del juez validador: ANICAMA SILVA JOSE CARLOS DNI: 44053189

Especialidad del validador: Temático Metodológico Estadístico

Lima, 22 de septiembre de 2023

¹**Pertinencia:** El ítem corresponde al concepto teórico formulado.

²**Relevancia:** El ítem es apropiado para representar al componente o dimensión específica del constructo.

³**Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo.

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión.



Firma del experto informante

Anexo 16: Verificación de grados y títulos de SUNEDU - Experto 1



PERÚ

Ministerio de Educación

Superintendencia Nacional de
Educación Superior Universitaria

Dirección de Documentación e
Información Universitaria y
Registro de Grados y Títulos

REGISTRO NACIONAL DE GRADOS ACADÉMICOS Y TÍTULOS PROFESIONALES

Graduado	Grado o Título	Institución
ANICAMA SILVA, JOSE CARLOS DNI 44053189	BACHILLER EN INGENIERIA DE COMPUTACION Y SISTEMAS Fecha de diploma: 11/11/16 Modalidad de estudios: PRESENCIAL Fecha matrícula: 20/02/2014 Fecha egreso: 26/08/2016	UNIVERSIDAD DE SAN MARTÍN DE PORRES <i>PERU</i>
ANICAMA SILVA, JOSE CARLOS DNI 44053189	INGENIERO DE COMPUTACIÓN Y SISTEMAS Fecha de diploma: 06/04/18 Modalidad de estudios: PRESENCIAL	UNIVERSIDAD DE SAN MARTÍN DE PORRES <i>PERU</i>
ANICAMA SILVA, JOSE CARLOS DNI 44053189	MAESTRO EN EDUCACIÓN CON MENCIÓN EN INFORMÁTICA Y TECNOLOGÍA EDUCATIVA Fecha de diploma: 11/12/20 Modalidad de estudios: SEMIPRESENCIAL Fecha matrícula: 27/03/2018 Fecha egreso: 04/02/2019	UNIVERSIDAD DE SAN MARTÍN DE PORRES <i>PERU</i>

Anexo 17: Validación de instrumentos - Experto 2

CERTIFICADO DE VALIDEZ DE CONTENIDO DE INSTRUMENTOS A TRAVÉS DE JUICIO DE EXPERTO

Título de la investigación	IMPLEMENTACIÓN DE UN MODELO BASADO EN ALGORITMOS DE MACHINE LEARNING PARA EL ANÁLISIS PREDICTIVO DE DIAGNÓSTICOS CLÍNICOS EN REDLAB PERÚ S.A.C
Nombre(s) del(os) instrumento(s)	<ul style="list-style-type: none"> Fichas de registro Cuestionario
Autor(es) del instrumento	<ul style="list-style-type: none"> Fuentes Pinedo, Anderson Alfredo Navarro Caraza, Carlos Enmanuel

N°	DIMENSIONES / Indicadores	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		SI	NO	SI	NO	SI	NO	
DIMENSIÓN 1: MÉTRICAS DE RENDIMIENTO								
1	Tasa de exactitud de la predicción	X		X		X		
2	Tasa de precisión de la predicción	X		X		X		
3	Tasa de sensibilidad de la predicción	X		X		X		
DIMENSIÓN 2: TIEMPO								
1	Tiempo requerido para procesar datos	X		X		X		
2	Tiempo requerido para predecir el diagnóstico clínico	X		X		X		
DIMENSIÓN 3: TOMA DE DECISIONES								
1	¿Qué tan conforme se siente con la exactitud de los diagnósticos actuales en el laboratorio clínico?	X		X		X		
2	¿Qué tan conforme se siente con la precisión de los resultados obtenidos de los diagnósticos clínicos?	X		X		X		
3	Basándose en su experiencia, ¿qué tan conforme se siente con la sensibilidad (habilidad del diagnóstico para identificar acertadamente las condiciones reales) al comunicar los hallazgos clínicos?	X		X		X		
4	¿Qué tan conforme se siente con el tiempo que generalmente se toma para preparar y procesar los datos antes de un diagnóstico?	X		X		X		
5	¿Qué tan conforme se siente con el tiempo que se dedica a la recopilación y preparación de datos para el diagnóstico?	X		X		X		
6	¿Qué tan conforme se siente con el intervalo de tiempo entre el procesamiento de datos y la obtención de un diagnóstico?	X		X		X		
7	¿Qué tan conforme se siente con el tiempo que se tarda desde que se inicia el análisis hasta que se comunica un diagnóstico clínico?	X		X		X		
8	¿Qué tan conforme se siente con la claridad de los diagnósticos clínicos que obtiene en el laboratorio, en cuanto a su influencia en su nivel de confianza para tomar decisiones?	X		X		X		
9	Al revisar los resultados de laboratorio, ¿qué tan conforme se siente con la información proporcionada para la toma de decisiones?	X		X		X		
10	Basado en los resultados que obtiene del laboratorio, ¿qué tan conforme se siente respecto a su capacidad de influir en decisiones posteriores?	X		X		X		

Observaciones (precisar si hay suficiencia): Si hay suficiencia

Opinión de aplicabilidad: Aplicable Aplicable después de corregir No aplicable

Apellidos y nombres del juez validador: PETRLIK AZABACHE IVAN CARLO DNI: 10140461

Especialidad del validador: Temático Metodológico Estadístico

¹**Pertinencia:** El ítem corresponde al concepto teórico formulado.

²**Relevancia:** El ítem es apropiado para representar al componente o dimensión específica del constructo.

³**Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo.

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión.



Lima, 23 de septiembre de 2023

Firma del experto informante

Anexo 18: Verificación de grados y títulos de SUNEDU - Experto 2



PERÚ

Ministerio de Educación

Superintendencia Nacional de
Educación Superior Universitaria

Dirección de Documentación e
Información Universitaria y
Registro de Grados y Títulos

REGISTRO NACIONAL DE GRADOS ACADÉMICOS Y TÍTULOS PROFESIONALES

Graduado	Grado o Título	Institución
PETRLIK AZABACHE, IVAN CARLO DNI 10140461	INGENIERO DE COMPUTACION Y SISTEMAS Fecha de diploma: 01/09/2006 Modalidad de estudios: -	UNIVERSIDAD DE SAN MARTÍN DE PORRES <i>PERU</i>
PETRLIK AZABACHE, IVAN CARLO DNI 10140461	BACHILLER EN INGENIERIA DE COMPUTACION Y SISTEMAS Fecha de diploma: 20/08/2004 Modalidad de estudios: - Fecha matrícula: Sin información (***) Fecha egreso: Sin información (***)	UNIVERSIDAD DE SAN MARTÍN DE PORRES <i>PERU</i>
PETRLIK AZABACHE, IVÁN CARLO DNI 10140461	MAESTRO EN INGENIERÍA DE SISTEMAS Fecha de diploma: 08/05/17 Modalidad de estudios: PRESENCIAL Fecha matrícula: 04/10/2007 Fecha egreso: 07/02/2008	UNIVERSIDAD NACIONAL FEDERICO VILLARREAL <i>PERU</i>
PETRLIK AZABACHE, IVÁN CARLO DNI 10140461	DOCTOR EN INGENIERÍA DE SISTEMAS Fecha de diploma: 10/08/20 Modalidad de estudios: PRESENCIAL Fecha matrícula: 14/06/2016 Fecha egreso: 10/12/2019	UNIVERSIDAD NACIONAL FEDERICO VILLARREAL <i>PERU</i>

Anexo 19: Validación de instrumentos - Experto 3

CERTIFICADO DE VALIDEZ DE CONTENIDO DE INSTRUMENTOS A TRAVÉS DE JUICIO DE EXPERTO

Título de la investigación	IMPLEMENTACIÓN DE UN MODELO BASADO EN ALGORITMOS DE MACHINE LEARNING PARA EL ANÁLISIS PREDICTIVO DE DIAGNÓSTICOS CLÍNICOS EN REDLAB PERÚ S.A.C
Nombre(s) del(los) instrumento(s)	<ul style="list-style-type: none"> Fichas de registro Cuestionario
Autor(es) del instrumento	<ul style="list-style-type: none"> Fuentes Pinedo, Anderson Alfredo Navarro Caraza, Carlos Enmanuel

N°	DIMENSIONES / Indicadores	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		SI	NO	SI	NO	SI	NO	
DIMENSIÓN 1: MÉTRICAS DE RENDIMIENTO								
1	Tasa de exactitud de la predicción	X		X		X		
2	Tasa de precisión de la predicción	X		X		X		
3	Tasa de sensibilidad de la predicción	X		X		X		
DIMENSIÓN 2: TIEMPO								
1	Tiempo requerido para procesar datos	X		X		X		
2	Tiempo requerido para predecir el diagnóstico clínico	X		X		X		
DIMENSIÓN 3: TOMA DE DECISIONES								
1	¿Qué tan conforme se siente con la exactitud de los diagnósticos actuales en el laboratorio clínico?	X		X		X		
2	¿Qué tan conforme se siente con la precisión de los resultados obtenidos de los diagnósticos clínicos?	X		X		X		
3	Basándose en su experiencia, ¿qué tan conforme se siente con la sensibilidad (habilidad del diagnóstico para identificar acertadamente las condiciones reales) al comunicar los hallazgos clínicos?	X		X		X		
4	¿Qué tan conforme se siente con el tiempo que generalmente se toma para preparar y procesar los datos antes de un diagnóstico?	X		X		X		
5	¿Qué tan conforme se siente con el tiempo que se dedica a la recopilación y preparación de datos para el diagnóstico?	X		X		X		
6	¿Qué tan conforme se siente con el intervalo de tiempo entre el procesamiento de datos y la obtención de un diagnóstico?	X		X		X		
7	¿Qué tan conforme se siente con el tiempo que se tarda desde que se inicia el análisis hasta que se comunica un diagnóstico clínico?	X		X		X		
8	¿Qué tan conforme se siente con la claridad de los diagnósticos clínicos que obtiene en el laboratorio, en cuanto a su influencia en su nivel de confianza para tomar decisiones?	X		X		X		
9	Al revisar los resultados de laboratorio, ¿qué tan conforme se siente con la información proporcionada para la toma de decisiones?	X		X		X		
10	Basado en los resultados que obtiene del laboratorio, ¿qué tan conforme se siente respecto a su capacidad de influir en decisiones posteriores?	X		X		X		

Observaciones (precisar si hay suficiencia):

Opinión de aplicabilidad: Aplicable [X] Aplicable después de corregir [] No aplicable []

Apellidos y nombres del juez validador: Lon Kan Prado Carlos Alberto DNI: 15595507

Especialidad del validador: Temático [X] Metodológico [] Estadístico []

Lima, 28 de septiembre de 2023

¹**Pertinencia:** El ítem corresponde al concepto teórico formulado.

²**Relevancia:** El ítem es apropiado para representar al componente o dimensión específica del constructo.

³**Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo.

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión.

Firma del experto informante

Anexo 20: Verificación de grados y títulos de SUNEDU - Experto 3



PERÚ

Ministerio de Educación

Superintendencia Nacional de
Educación Superior Universitaria

Dirección de Documentación e
Información Universitaria y
Registro de Grados y Títulos

REGISTRO NACIONAL DE GRADOS ACADÉMICOS Y TÍTULOS PROFESIONALES

Graduado	Grado o Título	Institución
LON KAN PRADO, CARLOS ALBERTO DNI 15595507	INGENIERO ELECTRONICO Fecha de diploma: 31/01/2003 Modalidad de estudios: -	UNIVERSIDAD NACIONAL DE INGENIERÍA <i>PERU</i>
LON KAN PRADO, CARLOS ALBERTO DNI 15595507	MAESTRO EN INGENIERÍA INDUSTRIAL CON MENCIÓN EN GESTIÓN DE OPERACIONES Y PRODUCTIVIDAD Fecha de diploma: 30/04/15 Modalidad de estudios: PRESENCIAL Fecha matrícula: Sin información (***) Fecha egreso: Sin información (***)	UNIVERSIDAD NACIONAL FEDERICO VILLARREAL <i>PERU</i>
LON KAN PRADO, CARLOS ALBERTO DNI 15595507	BACHILLER EN CIENCIAS CON MENCION EN INGENIERIA ELECTRONICA Fecha de diploma: 06/12/90 Modalidad de estudios: PRESENCIAL Fecha matrícula: Sin información (***) Fecha egreso: Sin información (***)	UNIVERSIDAD NACIONAL DE INGENIERÍA <i>PERU</i>

Anexo 21: Catálogo de pruebas médicas disponibles por RedLab Perú



ID	PRUEBAS MEDICAS	DESCRIPCION
01	Free PSA	PSA libre, cáncer de próstata
02	TPSA	PSA total, enfermedades prostáticas
03	CoV-2 IgG	Anticuerpos IgG, COVID-19
04	HIV Combo	VIH, anticuerpos y antígenos
05	HBsAg	Hepatitis B, antígeno de superficie
06	ANA	Autoinmunidad, anticuerpos antinucleares
07	Anti-HBc	Hepatitis B, anticuerpos core
08	Anti-HCV	Hepatitis C, anticuerpos
09	Anti-CCP	Artritis reumatoide, anticuerpos
10	CoV-2 NAb	COVID-19, anticuerpos neutralizantes
11	dsDNA IgG	Lupus, anticuerpos ADN
12	VB12 II	Vitamina B12, niveles en sangre
13	TSH	Función tiroidea, hormona estimulante
14	Anti-TPO	Tiroides, anticuerpos peroxidasa
15	Anti-Tg	Tiroides, anticuerpos tiroglobulina
16	Insulin	Insulina, niveles en sangre
17	FT4	Tiroxina libre, metabolismo
18	CEA	Cáncer, antígeno carcinoembrionario
19	CA19-9	Cáncer pancreático, antígeno
20	CA125	Cáncer de ovario, antígeno
21	AFP	Cáncer hepático, antígeno
22	T4	Tiroxina total, metabolismo
23	PCT	Infección, procalcitonina
24	25-OH VitD	Triyodotironina libre, metabolismo
25	FT3	Cáncer de tiroides, tiroglobulina
26	Tg	Cáncer de tiroides, tiroglobulina
27	CA15-3	Cáncer de mama, antígeno
28	BNP	Insuficiencia cardíaca, péptido
29	CYFRA21-1	Cáncer de pulmón, citoqueratina
30	Anti-TP	Sífilis, anticuerpos

31	E2	Hormona esteroidea
32	Ferritin	Almacén de hierro
33	HCG	Embarazo, hormona
34	Prolactin	Hormona lactógena, niveles en sangre
35	Testosterone	Hormona androgénica
36	Progesterone	Hormona, ciclo menstrual
37	DHEA-S	Hormona, función adrenal
38	LH	Reproducción, niveles en sangre
39	Cortisol	Estrés, hormona adrenal.
40	ACTH	Estimula cortisol
41	FSH	Reproducción, niveles en sangre
42	CoV-2 IgM	Anticuerpos IgM, COVID-19
43	Folate	Vitamina B9, niveles en sangre
44	Troponin-I	Infarto, marcador cardíaco
45	T3	Hormona tiroidea, metabolismo
46	Hemoglobina A1c (HbA1c)	Control de glucosa, diabetes
47	Glucosa plasmática en ayunas (FPG)	Niveles de glucosa, diabetes
48	Creatinina sérica	Función renal, niveles en sangre
49	Nitrógeno ureico en sangre (BUN)	Indicador de función renal y metabolismo proteico
50	Ácido úrico sérico	Metabolismo de purinas
51	Colesterol total sérico	Lípidos en sangre
52	Lipoproteína de alta densidad (HDL)	Transporte reverso de colesterol
53	Lipoproteína de baja densidad (LDL)	Transporte de colesterol a tejidos
54	Triglicéridos séricos	Lípidos en sangre
55	Bilirrubina total sérica	Metabolismo hepático
56	Bilirrubina directa sérica	Fracción conjugada
57	Bilirrubina indirecta sérica	Fracción no conjugada
58	Fosfatasa alcalina (ALP)	Función hepática y ósea



59	Alanina aminotransferasa (ALT)	Enzima hepática
60	Aspartato aminotransferasa (AST)	Enzima hepática, función celular
61	Gamma-glutamil transferasa (GGT)	Enzima hepática, metabolismo de glutatión
62	Amilasa sérica	Enzima pancreática
63	Lipasa sérica	Enzima pancreática crucial en la digestión de lípidos
64	Proteína C reactiva (CRP)	Marcador de inflamación
65	Albumina sérica	Proteína principal del suero
66	Proteínas totales séricas	Medida total de proteínas en el suero
67	Globulinas séricas	Grupo de proteínas en el suero
68	Relación albúmina/globulina (A/G)	Cociente entre albúmina y globulinas
69	Hierro sérico	Concentración de hierro en el suero
70	Transferrina sérica	Proteína transportadora de hierro
71	Saturación de transferrina	Porcentaje de capacidad de unión de la transferrina ocupada por hierro
72	Hemoglobina sérica	Proteína que transporta oxígeno en los glóbulos rojos
73	Hematocrito	Porcentaje de volumen de glóbulos rojos en la sangre total
74	Conteo de leucocitos (WBC)	Número total de glóbulos blancos
75	Conteo de eritrocitos (RBC)	Número total de glóbulos rojos
76	Conteo de plaquetas (PLT)	Número de plaquetas en la sangre
77	Volumen corpuscular medio (MCV)	Tamaño promedio de los glóbulos rojos
78	Hemoglobina corpuscular media (MCH)	Cantidad promedio de hemoglobina en los glóbulos rojos
79	Concentración de hemoglobina corpuscular media (MCHC)	Concentración promedio de hemoglobina en los glóbulos rojos
80	Amplitud de distribución de eritrocitos (RDW)	Variación en el tamaño de los glóbulos rojos
81	Conteo de reticulocitos	Número de reticulocitos en la sangre
82	Velocidad de sedimentación globular (ESR)	Velocidad de sedimentación de los glóbulos rojos en la sangre
83	Fibrinógeno plasmático	Proteína de la coagulación
84	Dímero-D	Fragmento de la fibrina
85	Tiempo de protrombina (PT)	Tiempo de coagulación de la sangre
86	Tiempo de tromboplastina parcial activada (aPTT)	Tiempo de coagulación en un ensayo específico
87	Relación internacional normalizada (INR)	Estándar internacional para la medición del tiempo de protrombina
88	Calcio sérico	Nivel de calcio en el suero
89	Magnesio sérico	Nivel de magnesio en el suero
90	Fosfato sérico	Nivel de fosfato en el suero
91	Sodio sérico	Nivel de sodio en el suero

92	Potasio sérico	Concentración de potasio en el suero
93	Cloruro sérico	Concentración de cloruro en el suero
94	Bicarbonato sérico	Concentración de bicarbonato en el suero
95	Gases arteriales (ABG)	Medición de oxígeno, dióxido de carbono y pH en la sangre arterial
96	pH arterial	Medida de la acidez o alcalinidad de la sangre arterial
97	Lactato sérico	Nivel de lactato en el suero, indicativo de metabolismo anaeróbico
98	Osmolaridad plasmática	Concentración de partículas osmóticamente activas en el plasma
99	Microalbuminuria	Presencia elevada de albúmina en la orina
100	Creatinina en orina	Medición de creatinina en la orina
101	Electrolitos urinarios	Concentración de electrolitos en la orina
102	Ácido láctico plasmático	Nivel de ácido láctico en el plasma
103	Homocisteína plasmática	Nivel de homocisteína en el plasma
104	Cistatina C sérica	Marcador de función renal
105	Proteína C reactiva de alta sensibilidad (PCR hs)	Marcador de función renal
106	Troponina T ultrasensible	Marcador cardíaco para detectar daño al miocardio
107	NT-proBNP (N-terminal pro-BNP)	Biomarcador de insuficiencia cardíaca
108	Deshidrogenasa láctica (LDH)	Enzima celular
109	Mioglobina sérica	Proteína muscular
110	Péptido natriurético tipo B (BNP)	Biomarcador de estrés en el corazón
111	Proteína S funcional	Proteína implicada en la coagulación sanguínea
112	Antitrombina III	Proteína anticoagulante
113	Anticuerpos antiendomiso (EMA)	Anticuerpos contra tejido intestinal
114	Proteína C funcional	Proteína que regula la coagulación
115	Hemoglobina glicosilada (A1c)	Medida del control glucémico en diabéticos
116	Prueba de Coombs directa	Detección de anticuerpos en glóbulos rojo
117	Prueba de Coombs indirecta	Detección de anticuerpos en suero
118	Electrolitos urinarios	Concentración de electrolitos en la orina
119	Anticuerpos antitransglutaminasa tisular (tTG)	Anticuerpos contra la enzima tisular
120	Anticuerpos anti gliadina (AGA)	Anticuerpos contra la gliadina

Anexo 22: Desarrollo de la implementación del modelo de ML

Se presenta la implementación de las fases de la metodología CRISP-ML(Q) para el desarrollo de la solución:

Fase 1: Compresión del negocio y los datos

Objetivo del negocio:

El proyecto de implementación de un modelo basado en algoritmos de Machine Learning en Redlab Perú S.A.C, un laboratorio clínico especializado en la realización de pruebas médicas a pacientes, tiene como objetivo principal optimizar y mejorar la precisión de los resultados clínicos a través del análisis avanzado de datos. Al emplear técnicas de Machine Learning, se busca identificar patrones y correlaciones en los resultados de las pruebas médicas, proporcionando a los profesionales tecnólogos herramientas más precisas para la interpretación y diagnóstico clínico. Este enfoque no solo mejorará la eficiencia operativa del laboratorio, sino que también elevará la calidad de los diagnósticos ofrecidos a los pacientes.

Objetivo de los datos:

Para alcanzar los objetivos del negocio, se necesita un análisis exhaustivo de los datos recopilados por el laboratorio. Específicamente, se busca:

1. Identificar variables críticas y su relación con diferentes condiciones médicas.
2. Desarrollar modelos predictivos que puedan anticipar resultados clínicos anómalos o diagnósticos basados en combinaciones de pruebas médicas.
3. Mejorar la precisión y rapidez en la interpretación de resultados clínicos.

Recolección de datos:

El dataset utilizado en este proyecto contiene 11,771 registros de pruebas médicas realizadas a pacientes. Cada registro incluye resultados detallados de diversas pruebas médicas esenciales para el diagnóstico y seguimiento de diferentes condiciones de salud. Las pruebas médicas consideradas en este estudio son:

Tabla 53

Pruebas médicas del dataset

ID	Prueba médica	Descripción
001	VB12 II	Relacionado con la vitamina B12
002	Anti-TPO	Anti-TPO, ligados a enfermedades tiroideas
003	Anti-Tg	Anticuerpos contra la tiroglobulina
004	CEA	Antígeno carcinoembrionario, utilizado como marcador tumoral.
005	CA19-9	Marcador tumoral del cáncer de páncreas
006	CA125	Marcador tumoral asociado con el cáncer de ovario
007	AFP	Alfa-fetoproteína, un marcador tumoral
008	25-OH VitD	25-hidroxivitamina D, un marcador de la vitamina D en el cuerpo
009	CA15-3	Marcador tumoral para cáncer de mama
010	BNP	Péptido cerebral ligado a insuficiencia cardíaca
011	Anti-TP	Anticuerpos contra Treponema pallidum, causante de sífilis
012	Troponin-I	Una proteína del músculo cardíaco, indicador de daño cardíaco
013	HIV Combo	Prueba combinada para la detección del VIH
014	HBsAg	Antígeno de superficie del virus de la hepatitis B
015	ANA	Anticuerpos antinucleares, ligados a autoinmunidad
016	Anti-HBc	Anticuerpos contra el núcleo del virus de la hepatitis B
017	Anti-HCV	Anticuerpos contra el virus de la hepatitis C
018	Anti-CCP	Anticuerpos anti-CCP, ligados a artritis reumatoide
019	E2	Estradiol, una hormona esteroide sexual femenina
020	ACTH	ACTH, estimula el cortisol

021	Free PSA	PSA, diagnóstico de problemas prostáticos
022	TPSA	Área de superficie polar total, común en química
023	Ferritin	Una proteína que almacena hierro en el cuerpo
024	Testosterone	Testosterona, una hormona sexual masculina
025	HCG	Gonadotropina coriónica humana, asociada con el embarazo
026	Prolactin	Prolactina, una hormona producida por la hipófisis

Nota. Las pruebas médicas consideradas forman parte del catálogo del laboratorio clínico. Cada una se describe junto con su utilidad en diagnósticos específicos y condiciones clínicas relevantes.

Calidad de los datos:

La calidad de los datos constituyó un aspecto esencial para el éxito del modelo de Machine Learning desarrollado en esta investigación. Como parte del proceso, se llevó a cabo un análisis riguroso para garantizar la integridad y consistencia de los datos recolectados. Esto incluyó la identificación de valores faltantes, la corrección de errores en los registros y la resolución de inconsistencias que pudieron comprometer el análisis posterior. Este esfuerzo permitió construir una base sólida para el modelo, asegurando que los datos utilizados reflejaran adecuadamente las características del fenómeno estudiado y contribuyeran al desarrollo de un modelo robusto.

Asimismo, se aplicaron técnicas avanzadas para mejorar la calidad y cantidad del dataset. Durante la investigación, se utilizó la síntesis de datos con el objetivo de generar información adicional basada en patrones presentes en los datos recolectados, y se implementó el aumento de datos, lo que permitió enriquecer el conjunto mediante modificaciones controladas. También se incorporaron métodos para balancear las clases dentro del dataset, garantizando una distribución equitativa de las categorías analizadas. Estas estrategias optimizaron el dataset disponible, mejorando su representatividad y robustez, asegurando que los modelos desarrollados fueran no solo precisos, sino también confiables y aplicables a diversos escenarios relacionados con el problema de estudio.

Análisis exploratorio de datos (EDA):

Figura 26

Análisis exploratorio de datos de pruebas médicas



Nota. Elaborado mediante herramientas de visualización avanzadas en Python.

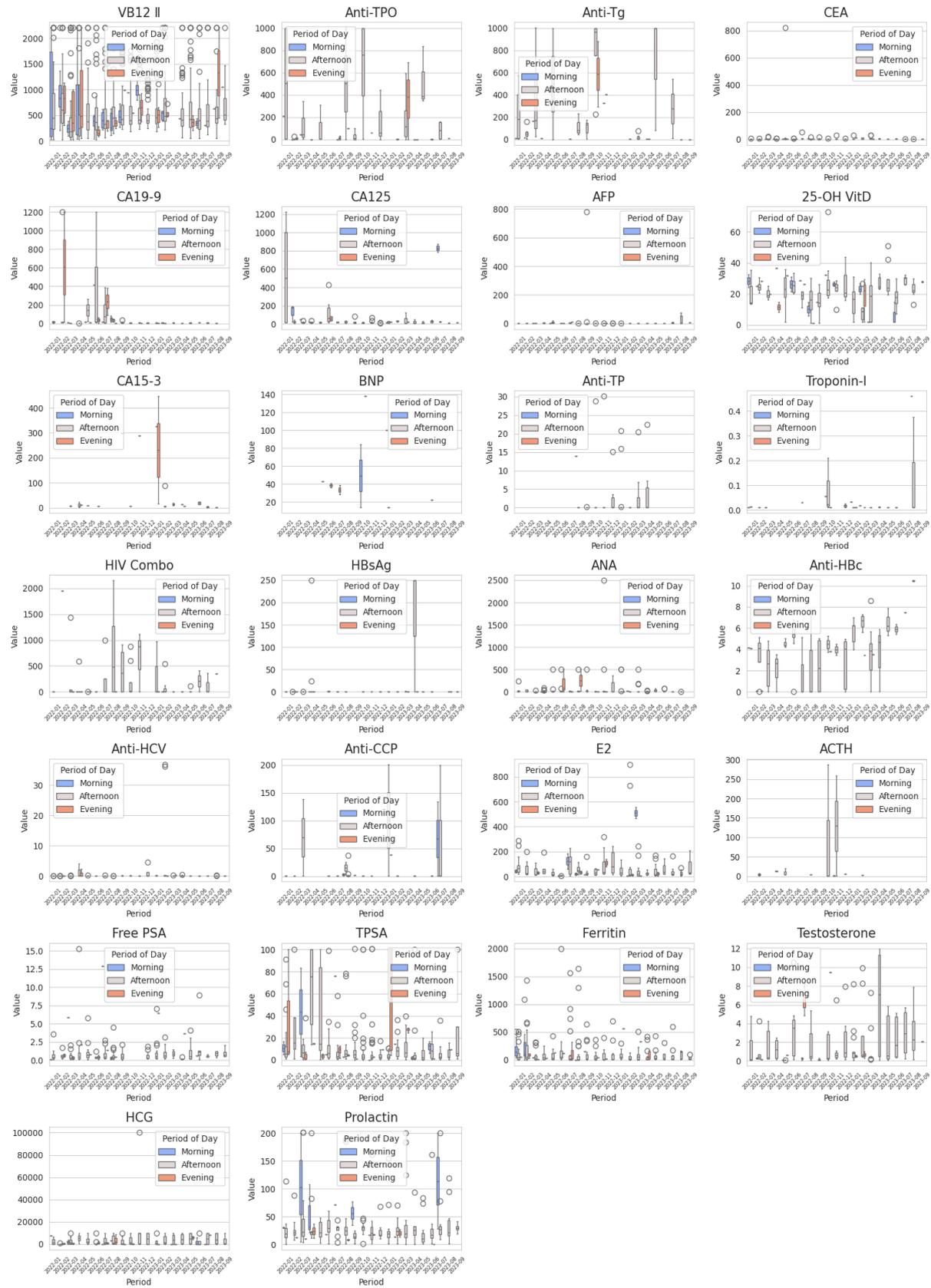
La exploración de datos (EDA) es crucial para comprender el dataset y preparar los datos para el modelado. En este estudio, se analizaron pruebas médicas de pacientes de Redlab Perú S.A.C., incluyendo VB12 II , Anti-TPO, Anti-Tg, CEA, CA19-9, CA125, y AFP. Cada prueba fue evaluada con estadísticas descriptivas como media, percentiles, desviación estándar, mínimo y máximo, para entender la distribución de los datos. Estas estadísticas ayudan a identificar la centralidad y dispersión de los datos, revelando posibles anomalías y relaciones entre pruebas. La visualización mediante gráficos de barras permite una comprensión clara y accesible de los resultados, destacando valores faltantes y asegurando una representación completa de la información recolectada.

Además de las estadísticas descriptivas, se llevó a cabo un análisis detallado de la frecuencia de las pruebas médicas utilizando histogramas, los cuales reflejan la variación de los resultados según el periodo del día (mañana, tarde y noche). Este análisis fue fundamental para identificar patrones temporales en los resultados de las pruebas, proporcionando una visión más profunda y detallada de cómo los valores de estas pruebas pueden variar a lo largo del día.

Los histogramas fueron diseñados para resaltar las diferencias en frecuencia y valores según el periodo del día, facilitando una interpretación más precisa de los datos recolectados. Esta visualización gráfica revela tendencias significativas y posibles anomalías en los resultados. En conjunto, este análisis de frecuencia muestra cómo los factores temporales pueden afectar los resultados de las pruebas médicas, enriqueciendo el análisis exploratorio de datos y proporcionando información valiosa para los profesionales de la salud, ayudándoles a tomar decisiones más informadas y a comprender mejor el comportamiento de los biomarcadores en diferentes momentos del día.

Figura 27

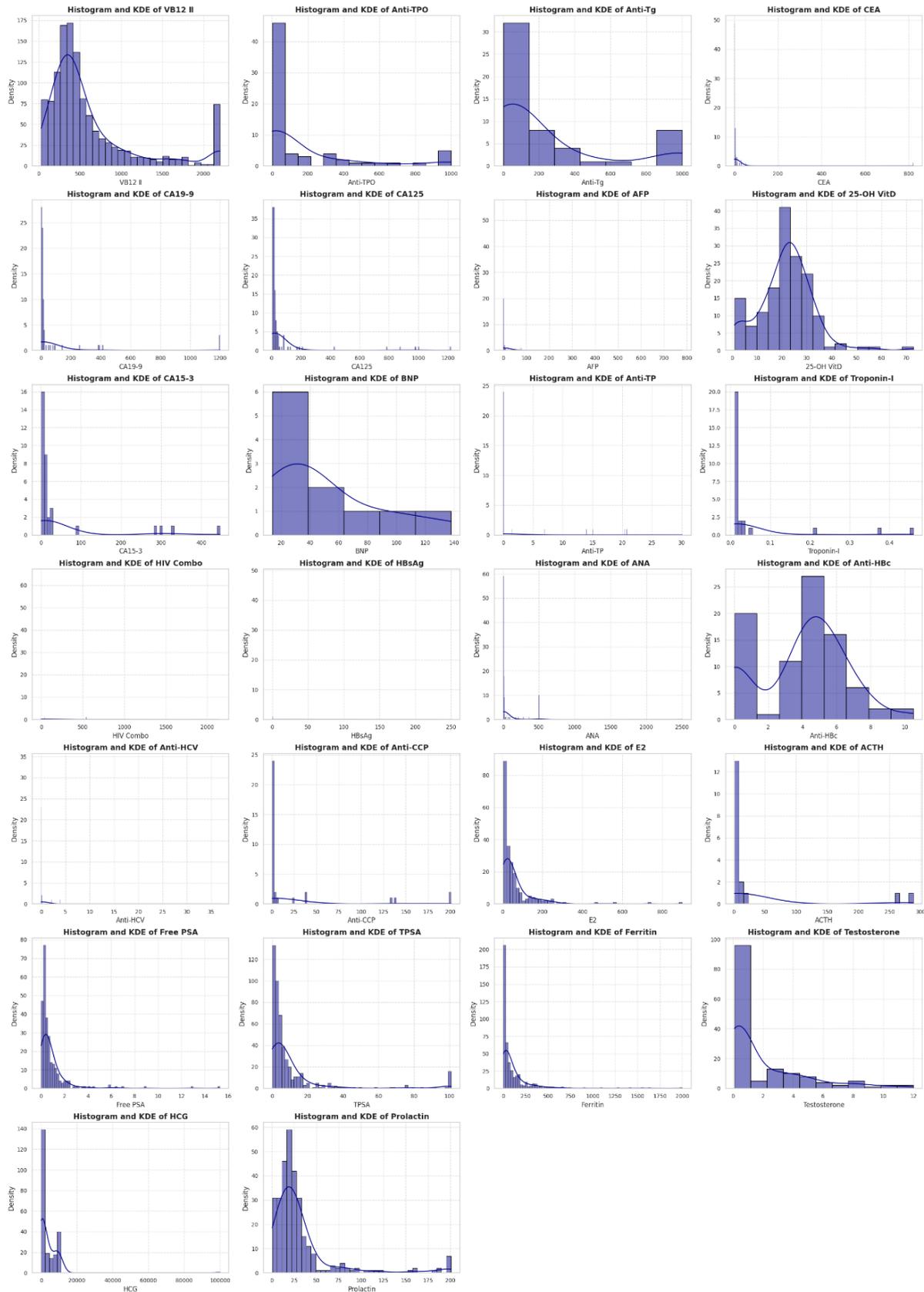
Histogramas de frecuencia de pruebas médicas por periodo del día



Nota. Elaborado mediante análisis de datos y herramientas de visualización avanzadas, utilizando metodologías estadísticas para la representación gráfica.

Figura 28

Histogramas y KDE de las pruebas médicas



Nota. Elaborado mediante técnicas de estimación de densidad de Kernel (KDE) y análisis estadístico avanzado, representando la distribución de las pruebas médicas de forma visual y detallada.

Fase 2: Preparación de los datos

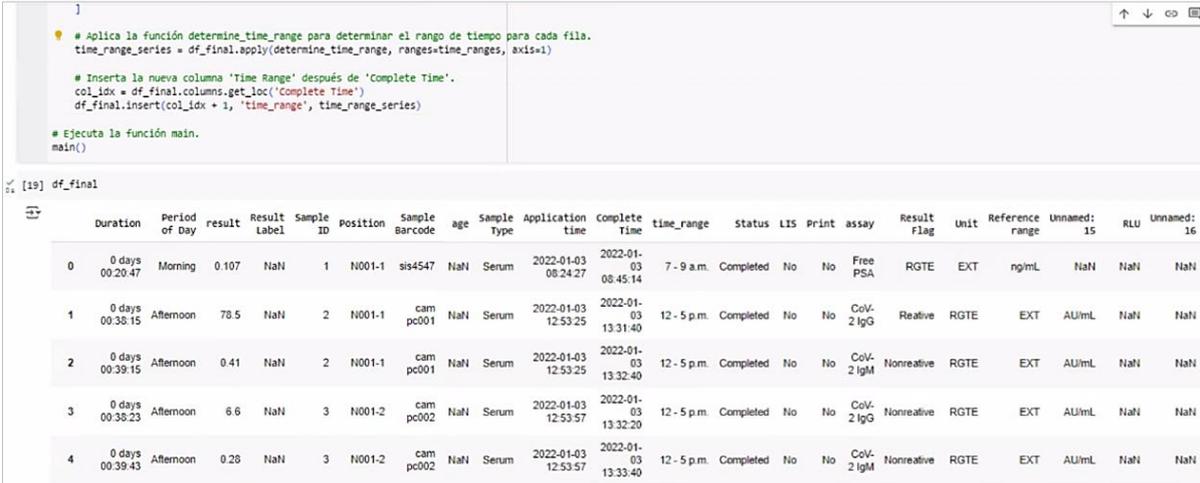
La fase de preparación de los datos en el proceso CRISP-ML(Q) es fundamental para transformar y limpiar el dataset, asegurando su adecuación para el análisis y modelado de Machine Learning. A continuación, se detallan las actividades realizadas en esta fase.

Selección e integración de datos:

En esta etapa, se seleccionaron las variables y registros relevantes para el análisis, eliminando columnas irrelevantes o redundantes y seleccionando un subconjunto representativo de los datos. Además, se integraron datos provenientes de múltiples fuentes para crear un dataset consolidado, asegurando cohesión y homogeneidad en la información utilizada para el análisis.

Figura 29

Integración de datos en dataset consolidado



```
]
# Aplica la función determine_time_range para determinar el rango de tiempo para cada fila.
time_range_series = df_final.apply(determine_time_range, ranges=time_ranges, axis=1)

# Inserta la nueva columna 'Time Range' después de 'Complete Time'.
col_idx = df_final.columns.get_loc('Complete Time')
df_final.insert(col_idx + 1, 'time_range', time_range_series)

# Ejecuta la función main.
main()
```

	Duration	Period of Day	result	Result Label	Sample ID	Position	Sample Barcode	age	Sample Type	Application time	Complete Time	time_range	Status	LIS	Print	assay	Result Flag	Unit	Reference range	Unnamed: 15	RLU	Unnamed: 16
0	0 days 00:20:47	Morning	0.107	NaN	1	N001-1	sis4547	NaN	Serum	2022-01-03 08:24:27	2022-01-03 08:45:14	7 - 9 a.m.	Completed	No	No	Free PSA	RGTE	EXT	ng/mL	NaN	NaN	NaN
1	0 days 00:38:15	Afternoon	78.5	NaN	2	N001-1	cam pc001	NaN	Serum	2022-01-03 12:53:25	2022-01-03 13:31:40	12 - 5 p.m.	Completed	No	No	CoV-2 IgG	Reactive	RGTE	EXT	AU/mL	NaN	NaN
2	0 days 00:39:15	Afternoon	0.41	NaN	2	N001-1	cam pc001	NaN	Serum	2022-01-03 12:53:25	2022-01-03 13:32:40	12 - 5 p.m.	Completed	No	No	CoV-2 IgM	Nonreactive	RGTE	EXT	AU/mL	NaN	NaN
3	0 days 00:38:23	Afternoon	6.6	NaN	3	N001-2	cam pc002	NaN	Serum	2022-01-03 12:53:57	2022-01-03 13:32:29	12 - 5 p.m.	Completed	No	No	CoV-2 IgG	Nonreactive	RGTE	EXT	AU/mL	NaN	NaN
4	0 days 00:39:43	Afternoon	0.28	NaN	3	N001-2	cam pc002	NaN	Serum	2022-01-03 12:53:57	2022-01-03 13:33:40	12 - 5 p.m.	Completed	No	No	CoV-2 IgM	Nonreactive	RGTE	EXT	AU/mL	NaN	NaN

Nota. Integración de datos consolidada mediante herramientas avanzadas en Python.

Limpieza de datos:

La limpieza de datos es esencial para garantizar la integridad y consistencia del dataset. Se identificaron y corrigieron valores faltantes, errores y outliers mediante técnicas como la imputación de valores faltantes, la corrección de errores tipográficos y la transformación o eliminación de outliers. En el contexto de Redlab Perú S.A.C.,

se abordaron 11,771 registros de pruebas médicas para asegurar que los datos fueran precisos y completos. Se aplicaron técnicas de imputación de datos para aumentar el volumen del dataset, utilizando algoritmos para determinar valores faltantes en variables críticas como la edad y género de los pacientes. Adicionalmente, se estandarizaron las variables categóricas, como el tipo de muestra y el rango de referencia, para facilitar su análisis y comparación. También se realizaron transformaciones en variables numéricas para normalizar la distribución de los datos y mejorar la aplicabilidad de los modelos estadísticos. Todo este proceso no solo mejoró la calidad del dataset, sino que también permitió obtener insights más fiables y representativos para la toma de decisiones estratégicas en el ámbito médico.

Figura 30

Incorporación de datos

	Duration	Period of Day	result	Result Label	Sample ID	Position	Sample Barcode	Patient_Label	age	gender	Sample Type	Application time	Complete Time	time_range	Status	LIS	Print	assay	Result Flag	Unit	Reference range	Unnamed: 15	RLU
0	0 days 00:20:47	Morning	0.107	NaN	1.0	N001-1	sis4547	Paciente-1	51	F	Serum	2022-01-03 08:24:27	2022-01-03 08:45:14	7 - 9 a.m.	Completed	No	No	Free PSA	RGTE	EXT	ng/mL	NaN	0.0
1	0 days 00:38:15	Afternoon	78.5	NaN	2.0	N001-1	cam pc001	Paciente-2	76	F	Serum	2022-01-03 12:53:25	2022-01-03 13:31:40	12 - 5 p.m.	Completed	No	No	CoV-2 IgG	Reactive	RGTE	EXT	AU/mL	0.0
2	0 days 00:39:15	Afternoon	0.41	NaN	2.0	N001-1	cam pc001	Paciente-2	57	F	Serum	2022-01-03 12:53:25	2022-01-03 13:32:40	12 - 5 p.m.	Completed	No	No	CoV-2 IgM	Nonreactive	RGTE	EXT	AU/mL	0.0
3	0 days 00:38:23	Afternoon	6.6	NaN	3.0	N001-2	cam pc002	Paciente-3	61	M	Serum	2022-01-03 12:53:57	2022-01-03 13:32:29	12 - 5 p.m.	Completed	No	No	CoV-2 IgG	Nonreactive	RGTE	EXT	AU/mL	0.0
4	0 days 00:39:43	Afternoon	0.28	NaN	3.0	N001-2	cam pc002	Paciente-3	62	M	Serum	2022-01-03 12:53:57	2022-01-03 13:33:40	12 - 5 p.m.	Completed	No	No	CoV-2 IgM	Nonreactive	RGTE	EXT	AU/mL	0.0
...
68303	NaT	NaN	3182650.500088	NaN	NaN	NaN	NaN	Paciente-0	22	M	NaN	NaT	NaT	Out of Range	NaN	NaN	NaN	T3	NaN	NaN	NaN	NaN	NaN
68304	NaT	NaN	429072.093523	NaN	NaN	NaN	NaN	Paciente-0	23	F	NaN	NaT	NaT	8 - 12 a.m.	NaN	NaN	NaN	T3	NaN	NaN	NaN	NaN	NaN
68305	NaT	NaN	3793856.795874	NaN	NaN	NaN	NaN	Paciente-0	52	F	NaN	NaT	NaT	3 - 7 p.m.	NaN	NaN	NaN	T3	NaN	NaN	NaN	NaN	NaN
68306	NaT	NaN	2908605.510553	NaN	NaN	NaN	NaN	Paciente-0	46	F	NaN	NaT	NaT	7 - 9 a.m.	NaN	NaN	NaN	T3	NaN	NaN	NaN	NaN	NaN
68307	NaT	NaN	2781121.597237	NaN	NaN	NaN	NaN	Paciente-0	21	F	NaN	NaT	NaT	12 - 5 p.m.	NaN	NaN	NaN	T3	NaN	NaN	NaN	NaN	NaN

68308 rows x 31 columns

Nota. Consolidación de datos realizada mediante procesos avanzados de manipulación y análisis.

Transformación de datos:

Se transformaron los datos para adecuarlos a los requisitos del modelo de Machine Learning, incluyendo la normalización, estandarización y codificación de variables categóricas, así como la generación de nuevas variables derivadas para enriquecer el dataset. Además, se llevaron a cabo pruebas de consistencia y validación cruzada para garantizar que las transformaciones realizadas no introdujeran sesgos ni errores en el conjunto de datos final. Estas transformaciones

permitieron optimizar el desempeño de los algoritmos de aprendizaje automático, mejorando la precisión de las predicciones y la capacidad del modelo para generalizar sobre nuevos datos. Asimismo, se implementaron pipelines automatizados para garantizar la reproducibilidad del proceso y facilitar futuras actualizaciones del dataset con nuevas observaciones.

Figura 31

Transformación de valores en el dataset

```

Returns:
- pd.DataFrame: DataFrame with the added 'Patient_Label' column.
"""

# Create unique numerical identifiers for each name
df['Patient_ID'] = pd.factorize(df['Name'])[0] + 1

# Convert these identifiers to the desired format
df['Patient_Label'] = 'Paciente-' + df['Patient_ID'].astype(str)

# Reorder the columns to place 'Patient_Label' right after 'Name'
name_idx = df.columns.get_loc('Name')
df.insert(name_idx + 1, 'Patient_Label', df.pop('Patient_ID'))

# Drop the numerical 'Patient_ID' column and 'Name' column as they're not needed anymore
df.drop(columns=['Patient_ID', 'Name'], inplace=True)

return df

df_final = create_patient_labels(expanded_df)
df_final

```

	Duration	Period of Day	result	Result Label	Sample ID	Position	Sample Barcode	Patient_Label	age	gender	Sample Type	Application time	Complete Time	time_range	Status	LIS
0	0 days 00:20:47	Morning	0.107	NaN	1.0	N001-1	sis4547	Paciente-1	51	F	Serum	2022-01-03 08:24:27	2022-01-03 08:45:14	7 - 9 a.m.	Completed	No
1	0 days 00:38:15	Afternoon	78.5	NaN	2.0	N001-1	cam pc001	Paciente-2	76	F	Serum	2022-01-03 12:53:25	2022-01-03 13:31:40	12 - 5 p.m.	Completed	No
2	0 days 00:39:15	Afternoon	0.41	NaN	2.0	N001-1	cam pc001	Paciente-2	57	F	Serum	2022-01-03 12:53:25	2022-01-03 13:32:40	12 - 5 p.m.	Completed	No

✓ 0 s completado a las 18:53

Nota. Transformación y manipulación avanzada para garantizar la usabilidad del dataset.

Ingeniería de características:

Se crearon nuevas características basadas en el conocimiento del dominio y técnicas automatizadas de ingeniería de características, con el objetivo de enriquecer el dataset y mejorar el rendimiento del modelo. Se derivaron nuevas variables a partir de las existentes para captar información adicional y proporcionar mayor contexto al modelo de Machine Learning. Además, se evaluó cuidadosamente la importancia de cada característica utilizando métodos como la selección de características basada en importancia y análisis de correlación, asegurando que solo se incluyeran las variables más relevantes para evitar redundancias y reducir la complejidad del modelo.

Figura 32

Preparación de campo nuevo en el dataset de resultados clínicos

```
[ ] #JSON
# Crear un diccionario a partir de un DataFrame previamente definido llamado df_config_dict
sheets_dict = df_config_dict

# Funciones auxiliares
# Esta función recupera las configuraciones específicas para una prueba de ensayo dada.
def get_config_data(assay):
    # Filtrar las configuraciones que coinciden con el ensayo dado
    config_data = sheets_dict['config'][sheets_dict['config']['assay'] == assay]
    if config_data.empty:
        return []
    tipo_config_id_values = config_data.iloc[0]['tipo_config_id']
    return tipo_config_id_values.split(",") if isinstance(tipo_config_id_values, str) else [tipo_config_id_values]

# Esta función recupera las configuraciones que coinciden con un tipo_config_id y una fila específicos.
def get_matching_configs(tipo_config_id, row):
    tipo_config_rows = sheets_dict['tipo_config'][sheets_dict['tipo_config']['id'] == float(tipo_config_id)]
    for _, tipo_config_row in tipo_config_rows.iterrows():
        expected_entries = tipo_config_row['name'].split(",")
        if any(entry in row.to_string() for entry in expected_entries):
            config_sheet_name = tipo_config_row['dataframe_name']
            filtered_configs = sheets_dict[config_sheet_name][sheets_dict[config_sheet_name]['assay'] == row['assay']]
            # Filtrar por género y rango de tiempo si están disponibles en las configuraciones y en la fila
            if 'gender' in filtered_configs.columns and 'gender' in row:
                filtered_configs = filtered_configs[filtered_configs['gender'] == row['gender']]
            if 'time_range' in filtered_configs.columns and 'time_range' in row:
                filtered_configs = filtered_configs[filtered_configs['time_range'] == row['time_range']]
            return filtered_configs
    return pd.DataFrame()

# Esta función recupera las entradas utilizadas para un tipo_config_id específico.
def get_used_entries(tipo_config_id):
    tipo_config_rows = sheets_dict['tipo_config'][sheets_dict['tipo_config']['id'] == float(tipo_config_id)]
    if not tipo_config_rows.empty:
        return tipo_config_rows.iloc[0]['name']
```

Nota. Configuración y preparación de datos realizada mediante funciones personalizadas para la integración y estandarización de resultados clínicos en el dataset.

Figura 33

Incorporación de nueva característica en el dataset en formato JSON

	Duration	Period of Day	result	Used_Entries	Final_Result	Result Label	Sample ID	Position	Sample Barcode	Patient_Label	age	gender	Sample Type	Application time	Complete Time
0	0 days 00:20:47	Morning	0.107	result, age, assay	{diagnostic_result: "Free PSA adecuado, se s...	NaN	1.0	N001-1	sis4547	Paciente-1	51	F	Serum	2022-01-03 08:24:27	2022-01-03 08:45:14
1	0 days 00:38:15	Afternoon	78.5	result, assay	{diagnostic_result: "El valor es reactivo pa...	NaN	2.0	N001-1	cam pc001	Paciente-2	76	F	Serum	2022-01-03 12:53:25	2022-01-03 13:31:40
2	0 days 00:39:15	Afternoon	0.41	result, assay	{diagnostic_result: "El valor es no reactivo...	NaN	2.0	N001-1	cam pc001	Paciente-2	57	F	Serum	2022-01-03 12:53:25	2022-01-03 13:32:40
3	0 days 00:38:23	Afternoon	6.6	result, assay	{diagnostic_result: "El valor es no reactivo...	NaN	3.0	N001-2	cam pc002	Paciente-3	61	M	Serum	2022-01-03 12:53:57	2022-01-03 13:32:20
4	0 days 00:39:43	Afternoon	0.28	result, assay	{diagnostic_result: "El valor es no reactivo...	NaN	3.0	N001-2	cam pc002	Paciente-3	62	M	Serum	2022-01-03 12:53:57	2022-01-03 13:33:40

Nota. Incorporación de una nueva característica en formato JSON para enriquecer el dataset y facilitar su interpretación y análisis.

Fase 3: Modelado – Ingeniería de Modelos

La Fase 3 del proceso CRISP-ML(Q) se enfoca en la creación, evaluación y selección de los modelos de Machine Learning que mejor se adapten a los objetivos del proyecto. A continuación, se detallan las actividades realizadas durante esta fase,

con un enfoque especial en el análisis comparativo de modelos y la elección del modelo de Machine Learning.

Análisis comparativo de Modelos de Machine Learning:

Los modelos entrenados fueron evaluados utilizando el conjunto de prueba. Se emplearon métricas de rendimiento como Accuracy, Recall, y Precision para comparar los diferentes modelos. Este análisis comparativo permitió identificar cuál de los modelos ofrecía el mejor equilibrio entre sesgo y varianza. Se utilizaron gráficos comparativos y tablas de métricas para visualizar y analizar las diferencias en el rendimiento de los modelos. Además, se aplicaron técnicas de validación cruzada para garantizar la robustez de los resultados y minimizar el riesgo de sobreajuste en los modelos seleccionados.

1. Logistic Regression: Es un modelo de ML que predice la probabilidad de un resultado binario. Transforma la combinación lineal de las características de entrada en una probabilidad usando la función logística (sigmoide), resultando en valores entre 0 y 1.

a. Importación y preparación de datos

Figura 34

Importación de librerías de Machine Learning – Logistic Regression

```
# Logistic Regression
# Este bloque ha sido entrenado con éxito
# Importando las bibliotecas necesarias
import pandas as pd
import json
import joblib
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, recall_score, precision_score
from sklearn.preprocessing import LabelEncoder, StandardScaler
import matplotlib.pyplot as plt
import seaborn as sns
```

Figura 35

Importación y conversión de datos JSON - Logistic Regression

```
# Convertir la columna 'Final_Result' en un diccionario
df_final['Final_Result_Dict'] = df_final['Final_Result'].apply(safe_json_or_dict)
df_final['Diagnostic_Result'] = df_final['Final_Result_Dict'].apply(lambda x: x.get('diagnostic_result', ''))
```

Figura 36

Codificación de etiquetas - Logistic Regression

```
# Codificar los resultados diagnósticos
label_encoder = LabelEncoder()
df_final['Diagnostic_Result_Coded'] = label_encoder.fit_transform(df_final['Diagnostic_Result'].astype(str))
```

b. Codificación y estandarización de variables

Figura 37

Codificación one-hot - Logistic Regression

```
# Codificación one-hot para las columnas relevantes
columns_to_encode = ['time_range', 'gender', 'assay']
encoded_features = pd.get_dummies(df_final[columns_to_encode])
```

Figura 38

Estandarización - Logistic Regression

```
# Estandarizar las columnas 'age' y 'result'
age_scaler = StandardScaler().fit(df_final[['age']])
result_scaler = StandardScaler().fit(df_final[['result']])
scaled_age = age_scaler.transform(df_final[['age']])
scaled_result = result_scaler.transform(df_final[['result']])
```

Figura 39

Concatenación de datos - Logistic Regression

```
# Crear conjunto de datos final
X = pd.concat([encoded_features, pd.DataFrame(scaled_age, columns=['age']), pd.DataFrame(scaled_result, columns=['result'])], axis=1)
y = df_final['Diagnostic_Result_Coded']
```

La codificación y estandarización de variables optimizan el preprocesamiento en regresión logística. Primero, la codificación one-hot (Figura 37) convierte datos categóricos en valores numéricos, eliminando la ordinalidad implícita y facilitando su interpretación por el modelo. Mientras tanto, la estandarización (Figura 38) normaliza las variables numéricas, asegurando una escala comparable que mejora la convergencia y estabilidad del modelo. Finalmente, la concatenación (Figura 39) integra todos los datos en un conjunto unificado, garantizando que las características relevantes estén alineadas y listas para ser procesadas. Este flujo asegura que el modelo aproveche al máximo la información disponible, reduciendo sesgos y mejorando su capacidad predictiva, mientras se valida el pipeline de transformación para garantizar la reproducibilidad en nuevos datos.

c. División del dataset y entrenamiento del modelo

Figura 40

División de datos de entrenamiento y prueba - Logistic Regression

```
# Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Entrenar el modelo de Regresión Logística
log_reg = LogisticRegression(max_iter=10000, random_state=42)
log_reg.fit(X_train, y_train)

# Hacer predicciones
y_pred = log_reg.predict(X_test)
```

d. Evaluación del modelo

Figura 41

Evaluación del modelo con métricas de rendimiento - Logistic Regression

```
# Calcular métricas
accuracy = accuracy_score(y_test, y_pred)
recall = recall_score(y_test, y_pred, average='weighted')
precision = precision_score(y_test, y_pred, average='weighted')

# Mostrar métricas
print(f"Accuracy: {accuracy}")
print(f"Recall: {recall}")
print(f"Precision: {precision}")

# Visualización de métricas
sns.set_theme(style="whitegrid")
plt.figure(figsize=(10, 7))
ax = sns.barplot(x=['Accuracy', 'Recall', 'Precision'], y=[accuracy, recall, precision], palette="Blues_d")
plt.ylim(0, 1)
plt.title('Model Metrics Overview')
plt.ylabel('Score')
for p in ax.patches:
    width = p.get_width()
    height = p.get_height()
    ax.text(p.get_x() + width/2., height + 0.02, '{:1.2f}'.format(height), ha="center")
plt.show()
```

e. Visualización de métricas

Tabla 54

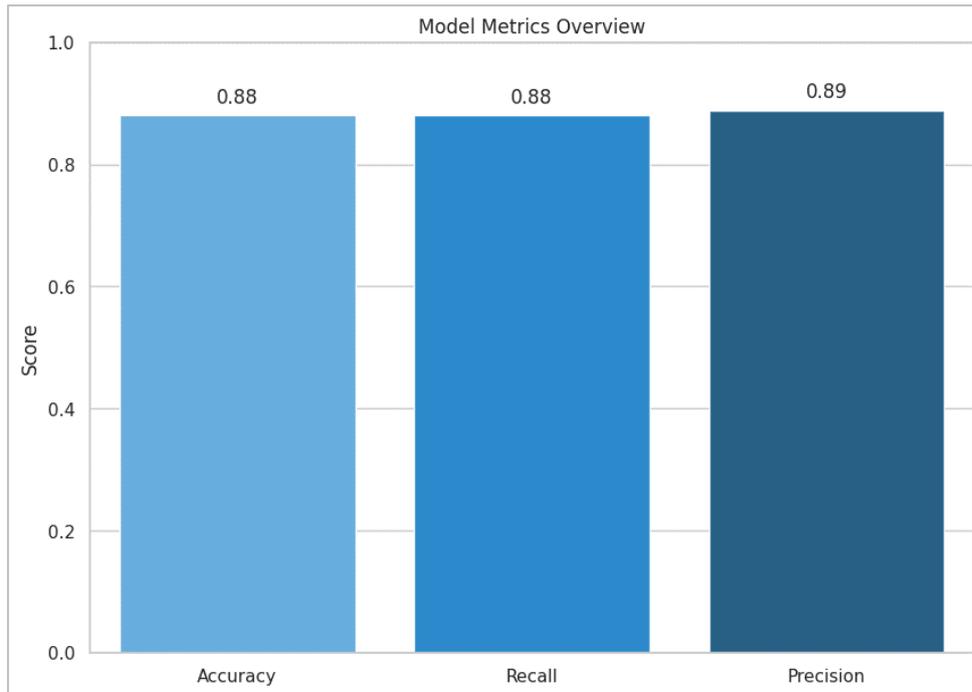
Métricas obtenidas del modelo - Logistic Regression

ID	Métrica de rendimiento	Valor obtenido
01	Accuracy	0.880541
02	Sensitivity	0.880541
03	Precision	0.888802

Nota. Métricas calculadas mediante Python utilizando bibliotecas como scikitLearn para la evaluación de modelos de Machine Learning.

Figura 42

Gráfico de barras de las métricas obtenidas - Logistic Regression



Nota. Representación gráfica de las métricas de rendimiento obtenidas para el modelo de regresión logística, facilitando la comparación visual de Accuracy, Recall y Precision.

f. Función de predicción para pruebas médicas

Figura 43

Función para predicción desde archivos Pickle ML - Logistic Regression

```
[ ] # Método para cargar el modelo y hacer predicciones desde archivos .pkl
def updated_predict_diagnosis_lr_from_pkl(**kwargs):
    # Cargar el modelo y demás objetos desde los archivos .pkl
    log_reg = joblib.load('{}\build\LR\log_reg_model.pkl'.format(dir))
    label_encoder = joblib.load('{}\build\LR\label_encoder.pkl'.format(dir))
    age_scaler = joblib.load('{}\build\LR\age_scaler.pkl'.format(dir))
    result_scaler = joblib.load('{}\build\LR\result_scaler.pkl'.format(dir))
    encoded_features_columns = joblib.load('{}\build\LR\encoded_features_columns.pkl'.format(dir))

    data = pd.DataFrame([0] * len(encoded_features_columns), index=encoded_features_columns).transpose()
    for key, value in kwargs.items():
        if key + "_" + str(value) in data.columns:
            data[key + "_" + str(value)] = 1
        elif key in data.columns:
            data[key] = value
    if 'age' in kwargs:
        data['age'] = age_scaler.transform([[kwargs['age']]])[0][0]
    if 'result' in kwargs:
        data['result'] = result_scaler.transform([[kwargs['result']]])[0][0]

    predicted_code = log_reg.predict(data)
    predicted_label = label_encoder.inverse_transform(predicted_code)
    return {
        'diagnostic_dict_output': predicted_label[0]
    }
```

Nota. Implementación de función en Python para realizar predicciones utilizando archivos Pickle que almacenan el modelo entrenado y sus objetos auxiliares, como codificadores y escaladores.

g. Predicción de pruebas médicas

Figura 44

Predicciones realizadas en pruebas médicas - Logistic Regression

```
[ ] # Obtener la predicción usando el modelo cargado desde archivos .pkl
prediction_from_pkl = updated_predict_diagnosis_lr_from_pkl(assay='ANA', result=500)

# Convertir la predicción en un DataFrame
df_prediction_from_pkl = pd.DataFrame([prediction_from_pkl], columns=prediction_from_pkl.keys())
print(df_prediction_from_pkl)

↔ diagnostic_dict_output
0 El valor es reactivo para ANA

[ ] #Probando con assay, result
print(updated_predict_diagnosis_lr_from_pkl(assay='VB12 II', result=500))
# Obtener la predicción
#Probando con assay, age y result
print(updated_predict_diagnosis_lr_from_pkl(assay='Free PSA', age=20, result=0.1))

↔ {'diagnostic_dict_output': 'Valor apto para VB12 II'}
{'diagnostic_dict_output': 'Free PSA no adecuado'}

▶ # Obtener la predicción
prediction = updated_predict_diagnosis_lr_from_pkl(assay='E2', gender="M", age=50, result=70)
print(prediction)

↔ {'diagnostic_dict_output': 'Valor no optimo para E2'}

[ ] # Obtener la predicción
prediction = updated_predict_diagnosis_lr_from_pkl(assay='Free PSA', age=20, result=0.1)
print(prediction)

↔ {'diagnostic_dict_output': 'Free PSA no adecuado'}
```

Nota. Predicciones en pruebas médicas realizadas con un modelo de regresión logística desde archivos Pickle.

2. Random Forest: Es un modelo de ML que emplea múltiples árboles de decisión para mejorar la precisión y evitar el sobreajuste. Combina las predicciones de varios árboles, entrenados con diferentes subconjuntos de datos, logrando un resultado más robusto y preciso.

a. Importación y preparación de datos

Figura 45

Importación de librerías de Machine Learning – Random Forest

```
[ ] # Random Forest con las funcionalidades
# Importando las bibliotecas necesarias
import pandas as pd
import json
import joblib
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score
from sklearn.preprocessing import LabelEncoder, StandardScaler
import matplotlib.pyplot as plt
import seaborn as sns
```

Figura 46

Importación y conversión de datos JSON – Random Forest

```
# Convertir la columna 'Final_Result' en un diccionario
df_final['Final_Result_Dict'] = df_final['Final_Result'].apply(safe_json_or_dict)
df_final['Diagnostic_Result'] = df_final['Final_Result_Dict'].apply(lambda x: x.get('diagnostic_result', ''))
```

Figura 47

Codificación de etiquetas - Random Forest

```
# Codificar los resultados diagnósticos
label_encoder = LabelEncoder()
df_final['Diagnostic_Result_Coded'] = label_encoder.fit_transform(df_final['Diagnostic_Result'].astype(str))
```

b. Codificación y estandarización de variables

Figura 48

Codificación one-hot, estandarización y concatenación - Random Forest

```
# Codificación one-hot para las columnas relevantes
columns_to_encode = ['time_range', 'gender', 'assay']
encoded_features = pd.get_dummies(df_final[columns_to_encode])

# Estandarizar las columnas 'age' y 'result'
age_scaler = StandardScaler().fit(df_final[['age']])
result_scaler = StandardScaler().fit(df_final[['result']])
scaled_age = age_scaler.transform(df_final[['age']])
scaled_result = result_scaler.transform(df_final[['result']])

# Crear conjunto de datos final
X = pd.concat([encoded_features, pd.DataFrame(scaled_age, columns=['age']), pd.DataFrame(scaled_result, columns=['result'])], axis=1)
y = df_final['Diagnostic_Result_Coded']
```

Nota. Proceso de preprocesamiento de datos mediante codificación one-hot, estandarización de variables numéricas y concatenación de características, optimizando el dataset para su uso en el modelo Random Forest.

c. División del dataset y entrenamiento del modelo

Figura 49

División de datos de entrenamiento y prueba - Random Forest

```
# Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Entrenar el modelo de Random Forest
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Hacer predicciones
y_pred = rf_model.predict(X_test)
```

Nota. Proceso completo de preparación de datos, incluyendo la división en conjuntos de entrenamiento y prueba, seguido del entrenamiento del modelo Random Forest con hiperparámetros definidos, y la generación de predicciones para evaluar su rendimiento en análisis supervisado.

d. Evaluación del modelo

Figura 50

Evaluación del modelo con métricas de rendimiento - Random Forest

```
# Calcular métricas
accuracy = accuracy_score(y_test, y_pred)
recall = recall_score(y_test, y_pred, average='weighted')
precision = precision_score(y_test, y_pred, average='weighted')

# Mostrar métricas
print(f"Accuracy: {accuracy}")
print(f"Recall: {recall}")
print(f"Precision: {precision}")

# Visualización de métricas
sns.set_theme(style="whitegrid")
plt.figure(figsize=(10, 7))
ax = sns.barplot(x=['Accuracy', 'Recall', 'Precision'], y=[accuracy, recall, precision], palette="Blues_d")
plt.ylim(0, 1)
plt.title('Model Metrics Overview')
plt.ylabel('Score')
for p in ax.patches:
    width = p.get_width()
    height = p.get_height()
    ax.text(p.get_x() + width/2., height + 0.02, '{:1.2f}'.format(height), ha="center")
plt.show()
```

e. Visualización de métricas

Tabla 55

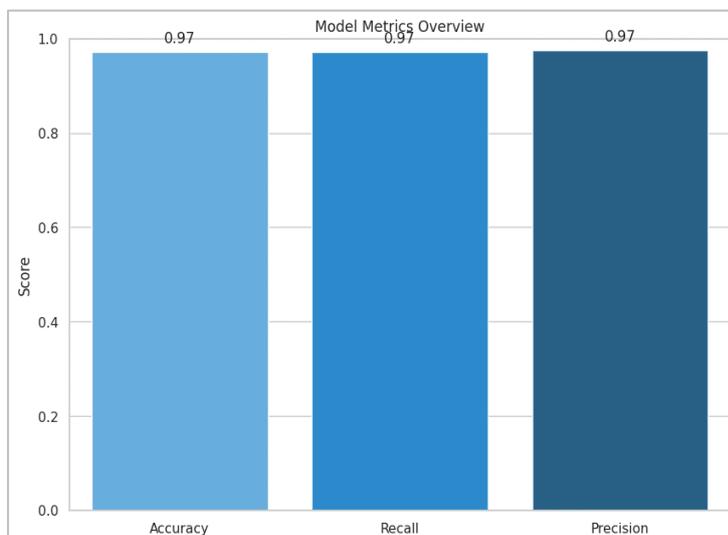
Métricas obtenidas del modelo – Random Forest

ID	Métrica de rendimiento	Valor obtenido
01	Accuracy	0.972066
02	Sensitivity	0.972066
03	Precision	0.974677

Nota. Métricas calculadas mediante Python utilizando bibliotecas como scikitLearn para la evaluación de modelos de Machine Learning.

Figura 51

Gráfico de barras de las métricas obtenidas - Random Forest



Nota. Representación gráfica de las métricas de rendimiento obtenidas para el modelo de Random Forest, facilitando la comparación visual de Accuracy, Recall y Precision.

f. Función de predicción para pruebas médicas

Figura 52

Función para predicción desde archivos Pickle ML - Random Forest

```
[ ] # Método para cargar el modelo y hacer predicciones desde archivos .pkl
def updated_predict_diagnosis_rf_from_pkl(dir, **kwargs):
    # Cargar el modelo y demás objetos desde los archivos .pkl
    rf_model = joblib.load('{}\build\RF\rf_model.pkl'.format(dir))
    label_encoder = joblib.load('{}\build\RF\label_encoder.pkl'.format(dir))
    age_scaler = joblib.load('{}\build\RF\age_scaler.pkl'.format(dir))
    result_scaler = joblib.load('{}\build\RF\result_scaler.pkl'.format(dir))
    encoded_features_columns = joblib.load('{}\build\RF\encoded_features_columns.pkl'.format(dir))

    data = pd.DataFrame([0] * len(encoded_features_columns), index=encoded_features_columns.transpose())
    for key, value in kwargs.items():
        if key + "_" + str(value) in data.columns:
            data[key + "_" + str(value)] = 1
        elif key in data.columns:
            data[key] = value
    if 'age' in kwargs:
        data['age'] = age_scaler.transform([[kwargs['age']]])[0][0]
    if 'result' in kwargs:
        data['result'] = result_scaler.transform([[kwargs['result']]])[0][0]

    predicted_code = rf_model.predict(data)
    predicted_label = label_encoder.inverse_transform(predicted_code)
    return {
        'diagnostic_dict_output': predicted_label[0]
    }
```

Nota. Implementación de una función en Python para realizar predicciones con el modelo Random Forest almacenado en archivos Pickle, integrando codificadores y escaladores para procesar y transformar los datos de entrada.

g. Predicción de pruebas médicas

Figura 53

Predicciones realizadas en pruebas médicas - Random Forest

```
[ ] #Probando con assay y result
print(updated_predict_diagnosis_rf(assay='CoV-2 IgG', result=11))

{'diagnostic_dict_output': 'El valor es no reactivo CoV-2 IgG'}
```

```
import pandas as pd

# Obtener la predicción
prediction = updated_predict_diagnosis_rf(assay='VB12 II', result=500)

# Convertir la predicción en un DataFrame
df_prediction = pd.DataFrame([prediction], columns=prediction.keys())

# Estilizar el DataFrame utilizando la función
styled_df = style_dataframe(df_prediction)

# Mostrar el DataFrame estilizado
display(styled_df)
```

```
diagnostic_dict_output
Valor apto para VB12 II
```

Nota. Predicciones realizadas en pruebas médicas utilizando el modelo Random Forest, con resultados organizados y presentados en un DataFrame para facilitar su interpretación y análisis.

3. Gradient Boosting: Es un modelo de ML que combina varios modelos débiles, generalmente árboles de decisión, para crear un predictor robusto. Se construye de forma secuencial, con cada nuevo modelo corrigiendo los errores de los anteriores, guiado por el gradiente del error. Esto resulta en un modelo final preciso y eficiente, ampliamente utilizado en tareas de clasificación y regresión debido a su capacidad para manejar datos complejos y minimizar el sobreajuste mediante técnicas de regularización. Además, su enfoque iterativo permite una optimización progresiva del rendimiento, garantizando adaptabilidad a distintos conjuntos de datos y escenarios analíticos. Su versatilidad lo convierte en una herramienta clave en problemas que requieren soluciones precisas y escalables.

a. Importación y preparación de datos

Figura 54

Importación de librerías de Machine Learning – Gradient Boosting

```
# Gradient Boosting
# Importando las bibliotecas necesarias
import pandas as pd
import json
from sklearn.model_selection import train_test_split
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score
from sklearn.preprocessing import LabelEncoder, StandardScaler
import matplotlib.pyplot as plt
import seaborn as sns
```

Figura 55

Importación y conversión de datos JSON – Gradient Boosting

```
# Convertir la columna 'Final_Result' en un diccionario
df_final['Final_Result_Dict'] = df_final['Final_Result'].apply(safe_json_or_dict)
df_final['Diagnostic_Result'] = df_final['Final_Result_Dict'].apply(lambda x: x.get('diagnostic_result', ''))
```

Figura 56

Codificación de etiquetas - Gradient Boosting

```
# Codificar los resultados diagnósticos
label_encoder = LabelEncoder()
df_final['Diagnostic_Result_Coded'] = label_encoder.fit_transform(df_final['Diagnostic_Result'].astype(str))
```

b. Codificación y estandarización de variables

Figura 57

Codificación one-hot, estandarización y concatenación - Gradient Boosting

```
# Codificación one-hot para las columnas relevantes
columns_to_encode = ['time_range', 'gender', 'assay']
encoded_features = pd.get_dummies(df_final[columns_to_encode])

# Estandarizar las columnas 'age' y 'result'
age_scaler = StandardScaler().fit(df_final[['age']])
result_scaler = StandardScaler().fit(df_final[['result']])
scaled_age = age_scaler.transform(df_final[['age']])
scaled_result = result_scaler.transform(df_final[['result']])

# Crear conjunto de datos final
X = pd.concat([encoded_features, pd.DataFrame(scaled_age, columns=['age']), pd.DataFrame(scaled_result, columns=['result'])], axis=1)
y = df_final['Diagnostic_Result_Coded']
```

c. División del dataset y entrenamiento del modelo

Figura 58

División de datos de entrenamiento y prueba - Gradient Boosting

```
# Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Entrenar el modelo de Gradient Boosting
gb_model = GradientBoostingClassifier(n_estimators=100, random_state=42)
gb_model.fit(X_train, y_train)

# Hacer predicciones
y_pred = gb_model.predict(X_test)
```

d. Evaluación del modelo

Figura 59

Evaluación del modelo con métricas de rendimiento - Gradient Boosting

```
# Calcular métricas
accuracy = accuracy_score(y_test, y_pred)
recall = recall_score(y_test, y_pred, average='weighted')
precision = precision_score(y_test, y_pred, average='weighted')

# Mostrar métricas
print(f"Accuracy: {accuracy}")
print(f"Recall: {recall}")
print(f"Precision: {precision}")

# Visualización de métricas
sns.set_theme(style="whitegrid")
plt.figure(figsize=(10, 7))
ax = sns.barplot(x=['Accuracy', 'Recall', 'Precision'], y=[accuracy, recall, precision], palette="Blues_d")
plt.ylim(0, 1)
plt.title('Model Metrics Overview')
plt.ylabel('Score')
for p in ax.patches:
    width = p.get_width()
    height = p.get_height()
    ax.text(p.get_x() + width/2., height + 0.02, '{:1.2f}'.format(height), ha="center")
plt.show()
```

Nota. Evaluación del modelo Gradient Boosting mediante métricas de rendimiento como Accuracy, Recall y Precision, calculadas y visualizadas en Python para analizar la eficacia del modelo.

e. Visualización de métricas

Tabla 56

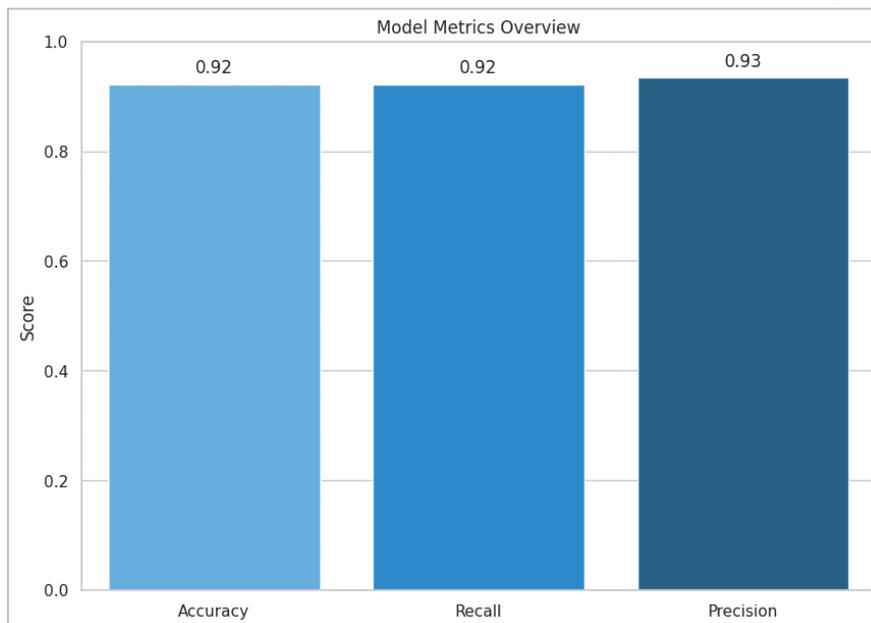
Métricas obtenidas del modelo – Gradient Boosting

ID	Métrica de rendimiento	Valor obtenido
01	Accuracy	0.922577
02	Sensitivity	0.922577
03	Precision	0.934006

Nota. Métricas calculadas mediante Python utilizando bibliotecas como scikitLearn para la evaluación de modelos de Machine Learning.

Figura 60

Gráfico de barras de las métricas obtenidas – Gradient Boosting



Nota. Representación gráfica de las métricas de rendimiento obtenidas para el modelo de Gradient Boosting, facilitando la comparación visual de Accuracy, Recall y Precision.

La Tabla 56 muestra las métricas del modelo Gradient Boosting, con un Accuracy y Recall de 92.26% y una Precision de 93.40%, destacando su capacidad para clasificar correctamente y minimizar errores. La Figura 60, mediante un gráfico de barras, permite visualizar y comparar estas métricas, evidenciando el desempeño equilibrado y eficiente del modelo. Este resultado refleja la robustez del modelo al manejar datos complejos y su eficacia en tareas de clasificación. Además, la

representación conjunta de la tabla y el gráfico facilita una interpretación clara y rápida de las métricas obtenidas.

f. Función de predicción para pruebas médicas

Figura 61

Función para predicción desde archivos Pickle ML – Gradient Boosting

```
[ ] # Método para cargar el modelo y hacer predicciones desde archivos .pkl
def updated_predict_diagnosis_gb_from_pkl(dir, **kwargs):
    # Cargar el modelo y demás objetos desde los archivos .pkl
    gb_model = joblib.load('{}build/GB/gb_model.pkl'.format(dir))
    label_encoder = joblib.load('{}build/GB/label_encoder.pkl'.format(dir))
    age_scaler = joblib.load('{}build/GB/age_scaler.pkl'.format(dir))
    result_scaler = joblib.load('{}build/GB/result_scaler.pkl'.format(dir))
    encoded_features_columns = joblib.load('{}build/GB/encoded_features_columns.pkl'.format(dir))

    data = pd.DataFrame([0] * len(encoded_features_columns), index=encoded_features_columns).transpose()
    for key, value in kwargs.items():
        if key + "_" + str(value) in data.columns:
            data[key + "_" + str(value)] = 1
        elif key in data.columns:
            data[key] = value
    if 'age' in kwargs:
        data['age'] = age_scaler.transform([[kwargs['age']])[0][0]
    if 'result' in kwargs:
        data['result'] = result_scaler.transform([[kwargs['result']])[0][0]

    predicted_code = gb_model.predict(data)
    predicted_label = label_encoder.inverse_transform(predicted_code)
    return {
        'diagnostic_dict_output': predicted_label[0]
    }
```

Nota. Función en Python para predicciones con Gradient Boosting desde archivos Pickle, integrando codificadores y escaladores.

g. Predicción de pruebas médicas

Figura 62

Predicciones realizadas en pruebas médicas – Gradient Boosting

```
[ ] #Probando con assay, age y result
print(updated_predict_diagnosis_gb(assay='Free PSA', age=20, result=8))

{'diagnostic_dict_output': 'Free PSA no adecuado'}

import pandas as pd

# Obtener la predicción
prediction = updated_predict_diagnosis_gb(assay='Free PSA', age=20, result=8)

# Convertir la predicción en un DataFrame
df_prediction = pd.DataFrame([prediction], columns=prediction.keys())

# Estilizar el DataFrame utilizando la función
styled_df = style_dataframe(df_prediction)

# Mostrar el DataFrame estilizado
display(styled_df)

diagnostic_dict_output
Free PSA no adecuado

[ ] #Probando con assay, gender, age y result
print(updated_predict_diagnosis_gb(assay='E2', gender="M", age=50, result=70))

{'diagnostic_dict_output': 'Valor no optimo para E2'}
```

4. Decision Tree: Es un modelo de ML que utiliza un árbol de decisiones para predecir el valor de una variable objetivo. Divide los datos en subconjuntos basados en características, creando nodos y ramas que representan decisiones lógicas, hasta alcanzar una predicción en las hojas del árbol.

a. Importación y preparación de datos

Figura 63

Importación de librerías de Machine Learning – Decision Tree

```
[ ] import pandas as pd
import numpy as np
import json
from sklearn.model_selection import train_test_split, GridSearchCV, cross_validate, RepeatedStratifiedKFold
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score, make_scorer
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.utils import class_weight
import ast
```

Figura 64

Importación y conversión de datos JSON – Decision Tree

```
# Convertir la columna 'Final_Result' en un diccionario
df_final['Final_Result_Dict'] = df_final['Final_Result'].apply(safe_json_or_dict)
df_final['Diagnostic_Result'] = df_final['Final_Result_Dict'].apply(lambda x: x.get('diagnostic_result', ''))
```

Figura 65

Codificación de etiquetas – Decision Tree

```
# Codificar los resultados diagnósticos
label_encoder = LabelEncoder()
df_final['Diagnostic_Result_Coded'] = label_encoder.fit_transform(df_final['Diagnostic_Result'].astype(str))
```

b. Codificación y estandarización de variables

Figura 66

Codificación one-hot, estandarización y concatenación de datos – Decision Tree

```
# Codificación one-hot para las columnas categóricas
columns_to_encode = ['time_range', 'gender', 'assay']
encoded_features = pd.get_dummies(df_final[columns_to_encode])

# Estandarizar las columnas numéricas
scaler = StandardScaler()
scaled_numerical = scaler.fit_transform(df_final[['age', 'result']])
scaled_numerical_df = pd.DataFrame(scaled_numerical, columns=['age', 'result'])

# Crear conjunto de datos final, asegurándote de que esté bien mezclado para evitar cualquier orden sesgado
X = pd.concat([encoded_features, scaled_numerical_df], axis=1).sample(frac=1, random_state=42)
y = df_final['Diagnostic_Result_Coded'].sample(frac=1, random_state=42) # Asegurarnos de usar el mismo random_state
```

c. División del dataset y entrenamiento del modelo

Figura 67

División de datos de entrenamiento y prueba – Decision Tree

```
# Dividir los datos en conjuntos de entrenamiento y prueba con estratificación (si es posible)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, stratify=stratify_param, random_state=42)

assert X.shape[0] == y.shape[0], "Las longitudes de X y y no coinciden."
# Asegurarnos de que todas las clases tienen al menos dos instancias para la estratificación
y = df_final['Diagnostic_Result_Coded']
if (y.value_counts() >= 2).all():
    stratify_param = y
else:
    stratify_param = None

# Definir parámetros para GridSearchCV, ajustando el rango de ccp_alpha
params = {
    'max_depth': [None, 5, 10, 15],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
    'criterion': ['gini', 'entropy'],
    'ccp_alpha': [0.001, 0.01, 0.1] # Valores para la poda
}

# Instancia del modelo DecisionTreeClassifier
dt_model = DecisionTreeClassifier(random_state=42, class_weight='balanced')

# Instancia de RepeatedStratifiedKFold para la validación cruzada
cv = RepeatedStratifiedKFold(n_splits=5, n_repeats=3, random_state=42)
grid_search = GridSearchCV(estimator=dt_model, param_grid=params, cv=cv, n_jobs=-1, scoring='accuracy')

# Ajustar GridSearchCV con validación cruzada
grid_search.fit(X_train, y_train)
```

Nota. División de datos con estratificación, entrenamiento del modelo Decision Tree y ajuste de hiperparámetros mediante GridSearchCV con validación cruzada para optimizar el rendimiento.

d. Evaluación del modelo

Figura 68

Evaluación del modelo con métricas de rendimiento – Decision Tree

```
# Mejor modelo encontrado
best_dt_model = grid_search.best_estimator_

# Usar cross_validate para una evaluación más robusta con múltiples métricas
scoring = {
    'accuracy': make_scorer(accuracy_score),
    'recall': make_scorer(recall_score, average='weighted'),
    'precision': make_scorer(precision_score, average='weighted'),
    'f1': make_scorer(f1_score, average='weighted')
}

scores = cross_validate(best_dt_model, X_train, y_train, cv=cv, scoring=scoring)

# Mostrar métricas promedio de la validación cruzada
print(f"Cross-Validated Accuracy: {np.mean(scores['test_accuracy'])}")
print(f"Cross-Validated Recall: {np.mean(scores['test_recall'])}")
print(f"Cross-Validated Precision: {np.mean(scores['test_precision'])}")
print(f"Cross-Validated F1 Score: {np.mean(scores['test_f1'])}")
```

Nota. Evaluación del modelo Decision Tree utilizando validación cruzada con múltiples métricas de rendimiento, como Accuracy, Recall, Precision y F1-Score, para garantizar un análisis robusto y confiable.

e. Visualización de métricas

Tabla 57

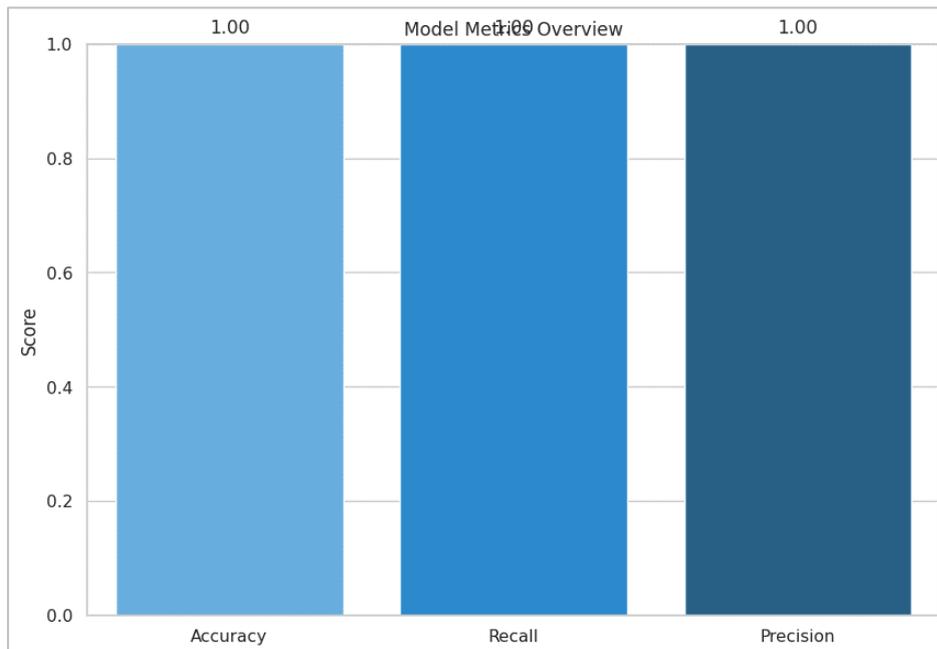
Métricas obtenidas del modelo – Decision Tree

ID	Métrica de rendimiento	Valor obtenido
01	Accuracy	0.998777
02	Sensitivity	0.998777
03	Precision	0.998792

Nota. Métricas calculadas mediante Python utilizando bibliotecas como scikitLearn para la evaluación de modelos de Machine Learning.

Figura 69

Gráfico de barras de las métricas obtenidas – Decision Tree



Nota. Representación gráfica de las métricas de rendimiento obtenidas para el modelo de Decision Tree, facilitando la comparación visual de Accuracy, Recall y Precision.

La Tabla 57 y la Figura 69 presentan las métricas de rendimiento del modelo Decision Tree, identificado como el mejor modelo por su excelente desempeño en Accuracy (99.88%), Recall (99.88%) y Precision (99.87%). Estas métricas destacan su capacidad para clasificar correctamente casi todas las instancias del conjunto de prueba, minimizando errores. La Figura 69, con un gráfico de barras, evidencia el

equilibrio y la consistencia del modelo, confirmando su robustez y eficacia para la tarea en cuestión.

f. Función de predicción para pruebas médicas

Figura 70

Métricas obtenidas del modelo – Decision Tree

```
[ ] # Método para cargar el modelo y hacer predicciones desde archivos .pkl
def updated_predict_diagnosis_dt_from_pkl(dir, **kwargs):
    # Cargar el modelo y demás objetos desde los archivos .pkl
    dt_model = joblib.load('{}\build\DT\dt_model.pkl'.format(dir))
    label_encoder = joblib.load('{}\build\DT\label_encoder.pkl'.format(dir))
    age_scaler = joblib.load('{}\build\DT\age_scaler.pkl'.format(dir))
    result_scaler = joblib.load('{}\build\DT\result_scaler.pkl'.format(dir))
    encoded_features_columns = joblib.load('{}\build\DT\encoded_features_columns.pkl'.format(dir))

    data = pd.DataFrame([0] * len(encoded_features_columns), index=encoded_features_columns).transpose()
    for key, value in kwargs.items():
        if key + "_" + str(value) in data.columns:
            data[key + "_" + str(value)] = 1
        elif key in data.columns:
            data[key] = value
    if 'age' in kwargs:
        data['age'] = age_scaler.transform([[kwargs['age']])[0])[0]
    if 'result' in kwargs:
        data['result'] = result_scaler.transform([[kwargs['result']])[0])[0]

    predicted_code = dt_model.predict(data)
    predicted_label = label_encoder.inverse_transform(predicted_code)
    return {
        'diagnostic_dict_output': predicted_label[0]
    }
```

Nota. Función en Python para predicciones con Decision Tree desde archivos Pickle, integrando codificadores y escaladores.

g. Predicción de pruebas médicas

Figura 71

Predicciones realizadas en pruebas médicas – Decision Tree

```
[ ] #Probando con assay, age y result
print(updated_predict_diagnosis_dt(assay='Free PSA', age=40, result=10))

{'diagnostic_dict_output': 'Free PSA no adecuado'}

import pandas as pd

# Obtener la predicción
prediction = updated_predict_diagnosis_dt(assay='Free PSA', age=40, result=10)

# Convertir la predicción en un DataFrame
df_prediction = pd.DataFrame([prediction], columns=prediction.keys())

# Estilizar el DataFrame utilizando la función
styled_df = style_dataframe(df_prediction)

# Mostrar el DataFrame estilizado
display(styled_df)

diagnostic_dict_output
Free PSA no adecuado

[ ] #Probando con assay, gender, age y result
print(updated_predict_diagnosis_dt(assay='E2', gender="M", age=50, result=50))

{'diagnostic_dict_output': 'Valor optimo para E2'}
```

Nota. Función en Python para predicciones con Decision Tree desde archivos Pickle, integrando codificadores y escaladores.

Fase 4: Evaluación

La Fase 4 del proceso CRISP-ML(Q) se centra en la validación exhaustiva del modelo seleccionado para garantizar que cumple con los objetivos del proyecto y las expectativas del negocio. En esta fase se realizan las siguientes actividades clave:

Revisión de métricas de rendimiento

Se revisan y comparan nuevamente las métricas de rendimiento del modelo seleccionado, tales como Accuracy, Recall (Sensivity) y Precision. Este paso es crucial para asegurar que el modelo sigue siendo el más adecuado según los criterios de evaluación definidos en la fase de modelado. Además, se utilizan técnicas de validación cruzada y pruebas adicionales con datos no vistos para confirmar que el modelo no esté sobreajustado y pueda generalizarse bien a nuevos datos.

Tabla 58

Métricas del análisis comparativo de modelos de Machine Learning

ID	Modelo evaluado	Exactitud	Precisión	Sensibilidad
01	Logistic Regression	0.880541	0.888802	0.880541
02	Random Forest	0.972066	0.974677	0.972066
03	Gradient Boosting	0.922577	0.934006	0.922577
04	Decision Tree	0.998777	0.998792	0.998777

Nota. Valores obtenidos mediante Python, utilizando técnicas de evaluación comparativa para analizar el rendimiento de los modelos Logistic Regression, Random Forest, Gradient Boosting y Decision Tree.

El modelo de Decision Tree presenta las mejores métricas de rendimiento con una exactitud de 0.998777, una precisión de 0.998792 y una sensibilidad de 0.998777, destacándose sobre los otros modelos evaluados.

Validación del modelo

Una vez seleccionado el modelo de Decision Tree durante la revisión de métricas de rendimiento, se lleva a cabo una validación exhaustiva para confirmar su desempeño. Se utilizan técnicas de validación cruzada y pruebas adicionales con

datos no vistos para garantizar que el modelo no esté sobreajustado y pueda generalizarse bien a nuevos datos.

Figura 72

Aplicación de validación cruzada sobre el modelo Decision Tree

```
# Mejor modelo encontrado
best_dt_model = grid_search.best_estimator_

# Usar cross_validate para una evaluación más robusta con múltiples métricas
scoring = {
    'accuracy': make_scorer(accuracy_score),
    'recall': make_scorer(recall_score, average='weighted'),
    'precision': make_scorer(precision_score, average='weighted'),
    'f1': make_scorer(f1_score, average='weighted')
}

scores = cross_validate(best_dt_model, X_train, y_train, cv=cv, scoring=scoring)

# Mostrar métricas promedio de la validación cruzada
print(f"Cross-Validated Accuracy: {np.mean(scores['test_accuracy'])}")
print(f"Cross-Validated Recall: {np.mean(scores['test_recall'])}")
print(f"Cross-Validated Precision: {np.mean(scores['test_precision'])}")
print(f"Cross-Validated F1 Score: {np.mean(scores['test_f1'])}")
```

↔ Cross-Validated Accuracy: 0.9675060325880526
Cross-Validated Recall: 0.9675060325880526
Cross-Validated Precision: 0.9751831575523835
Cross-Validated F1 Score: 0.967001326989288

Figura 73

Evaluación del modelo Decision Tree con el conjunto de prueba

```
[ ] # Evaluar el mejor modelo en el conjunto de prueba
test_predictions = best_dt_model.predict(X_test)
test_accuracy = accuracy_score(y_test, test_predictions)
test_recall = recall_score(y_test, test_predictions, average='weighted')
test_precision = precision_score(y_test, test_predictions, average='weighted')
test_f1 = f1_score(y_test, test_predictions, average='weighted')

# Mostrar las métricas de evaluación
print(f"Test Accuracy: {test_accuracy}")
print(f"Test Recall: {test_recall}")
print(f"Test Precision: {test_precision}")
print(f"Test F1 Score: {test_f1}")
```

↔ Test Accuracy: 0.9662307241850479
Test Recall: 0.9662307241850479
Test Precision: 0.9743504352130188
Test F1 Score: 0.9655113792407918

Según la Figura 72, los resultados de la validación cruzada muestran que el modelo de Decision Tree alcanzó una exactitud promedio de 0.967506, una sensibilidad de 0.967506, una precisión de 0.971581, y un puntaje F1 de 0.967001. Similarmente, según la Figura 73, en la evaluación con el conjunto de prueba, el

modelo obtuvo una exactitud de 0.966230, una sensibilidad de 0.973405, una precisión de 0.965513 y un puntaje F1 de 0.965513. Estos resultados indican que el modelo generaliza bien y tiene un alto rendimiento en términos de precisión y sensibilidad.

Figura 74

Obtención de métricas y elementos por prueba médica – Decision Tree

```
import numpy as np
from sklearn.metrics import confusion_matrix

# Asumimos que 'y_test', 'test_predictions' y 'X_test_with_assay' están definidos como se espera.
X_test_with_assay = X_test.join(df_final['assay'], how='left')

unique_assays = X_test_with_assay['assay'].unique()
metrics_per_assay = {}

# Esta función calcula los elementos de la matriz de confusión y devuelve los totales por clase.
def calculate_confusion_matrix_totals(conf_matrix):
    FP = conf_matrix.sum(axis=0) - np.diag(conf_matrix)
    FN = conf_matrix.sum(axis=1) - np.diag(conf_matrix)
    TP = np.diag(conf_matrix)
    TN = conf_matrix.sum() - (FP + FN + TP)
    return TP.sum(), TN.sum(), FP.sum(), FN.sum()

# Calcula las métricas agregadas para cada ensayo.
for assay in unique_assays:
    mask = X_test_with_assay['assay'] == assay
    y_true_assay = y_test[mask]
    y_pred_assay = test_predictions[mask]

    conf_matrix = confusion_matrix(y_true_assay, y_pred_assay)
    TP, TN, FP, FN = calculate_confusion_matrix_totals(conf_matrix)

    # Calcula las métricas globales para cada ensayo.
    accuracy = (TP + TN) / (TP + FP + FN + TN)
    precision = TP / (TP + FP) if TP + FP > 0 else 0
    recall = TP / (TP + FN) if TP + FN > 0 else 0
    f1_score = 2 * (precision * recall) / (precision + recall) if (precision + recall) > 0 else 0

    # Almacena las métricas calculadas en el diccionario.
    metrics_per_assay[assay] = {
        'TP': TP,
        'TN': TN,
        'FP': FP,
        'FN': FN,
        'Accuracy': accuracy,
        'Precision': precision,
        'Recall': recall,
        'F1 Score': f1_score
    }

# Imprime las métricas para cada ensayo.
for assay, metrics in metrics_per_assay.items():
    print(f"Metrics for assay '{assay}':")
    for metric_name, metric_value in metrics.items():
        if metric_name in ['Accuracy', 'Precision', 'Recall', 'F1 Score']:
            print(f" {metric_name}: {metric_value:.2f}")
        else:
            print(f" {metric_name}: {metric_value}")
    print("-" * 30)
```

Nota. Cálculo de métricas de rendimiento y elementos de la matriz de confusión por ensayo utilizando Python, evaluando el modelo Decision Tree para pruebas médicas.

La Figura 74 muestra el cálculo de métricas de rendimiento del modelo Decision Tree por prueba médica utilizando la matriz de confusión. Se calculan los verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN), y se derivan las métricas de exactitud, precisión, sensibilidad y F1 para cada prueba. Estas métricas se almacenan en un diccionario, proporcionando una evaluación detallada del modelo en diferentes pruebas médicas.

Figura 75

Resultados de métricas de rendimiento de determinadas pruebas médicas

```
-----  
Metrics for assay 'CA19-9':  
TP: 370  
TN: 370  
FP: 0  
FN: 0  
Accuracy: 1.00  
Precision: 1.00  
Recall: 1.00  
F1 Score: 1.00  
-----  
Metrics for assay 'CA125':  
TP: 388  
TN: 388  
FP: 0  
FN: 0  
Accuracy: 1.00  
Precision: 1.00  
Recall: 1.00  
F1 Score: 1.00  
-----  
Metrics for assay 'CA15-3':  
TP: 346  
TN: 346  
FP: 0  
FN: 0  
Accuracy: 1.00  
Precision: 1.00  
Recall: 1.00  
F1 Score: 1.00  
-----
```

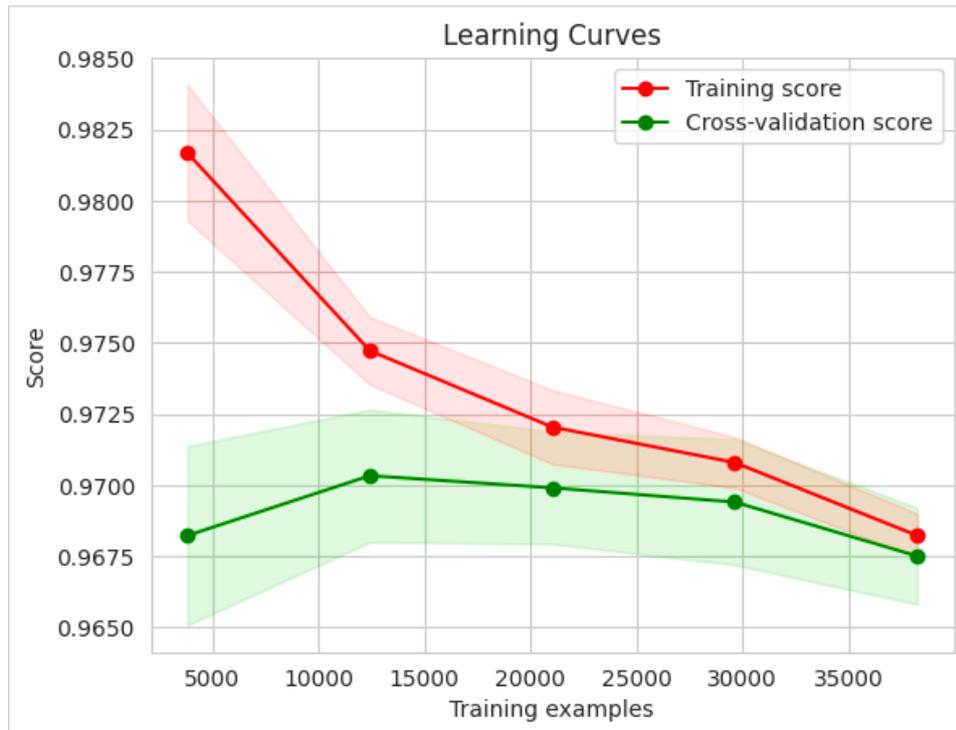
Figura 76

Verificación de sobreajuste mediante curvas de aprendizaje

```
from sklearn.model_selection import learning_curve  
import matplotlib.pyplot as plt  
  
train_sizes, train_scores, test_scores = learning_curve(  
    best_dt_model, X_train, y_train, cv=cv, n_jobs=-1,  
    train_sizes=np.linspace(.1, 1.0, 5))  
  
# Calcular media y desviación estándar del error de entrenamiento y prueba  
train_scores_mean = np.mean(train_scores, axis=1)  
train_scores_std = np.std(train_scores, axis=1)  
test_scores_mean = np.mean(test_scores, axis=1)  
test_scores_std = np.std(test_scores, axis=1)  
  
plt.fill_between(train_sizes, train_scores_mean - train_scores_std,  
                train_scores_mean + train_scores_std, alpha=0.1, color="r")  
plt.fill_between(train_sizes, test_scores_mean - test_scores_std,  
                test_scores_mean + test_scores_std, alpha=0.1, color="g")  
  
plt.plot(train_sizes, train_scores_mean, 'o-', color="r", label="Training score")  
plt.plot(train_sizes, test_scores_mean, 'o-', color="g", label="Test score")  
  
plt.title("Learning Curves")  
plt.xlabel("Training examples")  
plt.ylabel("Score")  
plt.legend(loc="best")  
plt.show()
```

Figura 77

Curvas de aprendizaje del Modelo Decision Tree



Nota. Las curvas de aprendizaje del modelo Decision Tree muestran la relación entre el tamaño del conjunto de entrenamiento y las puntuaciones de rendimiento, tanto en el conjunto de entrenamiento como en la validación cruzada, permitiendo evaluar el comportamiento del modelo y detectar posibles problemas de sobreajuste o subajuste.

La Figura 77 muestra que el puntaje de entrenamiento del modelo Decision Tree disminuye ligeramente con más datos, mientras que el puntaje de validación se mantiene estable, indicando buena generalización y ausencia de sobreajuste significativo.

Evaluación de impacto en el negocio

Una vez seleccionado el modelo de Decision Tree durante la fase de modelado, se lleva a cabo una evaluación exhaustiva para confirmar su desempeño y garantizar su aplicabilidad en escenarios reales. Se utilizan técnicas de validación cruzada y pruebas adicionales con datos no vistos para asegurar que el modelo no esté sobreajustado y pueda generalizarse bien a nuevos datos, manteniendo su robustez. Los resultados, como se muestra en las Figuras 72 y 73, indican que el modelo tiene una excelente capacidad de generalización y un alto rendimiento,

superando consistentemente los umbrales esperados de precisión. Además, las curvas de aprendizaje en la Figura 77 confirman que el modelo mantiene su estabilidad con diferentes tamaños de conjuntos de entrenamiento, lo que refuerza la confianza en que su implementación será confiable, efectiva y capaz de aportar un valor significativo para el negocio.

Elección final

Después de seleccionar el modelo de Decision Tree, se validó su desempeño con técnicas de validación cruzada y datos no vistos. Los resultados confirmaron su alta capacidad de generalización y rendimiento, consolidándolo como la opción para su implementación en producción. Además, se analizaron las curvas de aprendizaje para garantizar un equilibrio adecuado entre sobreajuste y subajuste, lo que respaldó su robustez en diversos escenarios. Se ajustaron los hiperparámetros clave para optimizar su precisión y sensibilidad, maximizando su eficacia en el contexto de las pruebas realizadas. Finalmente, su implementación fue evaluada en un entorno de prueba para asegurar la estabilidad y consistencia de los resultados en producción.

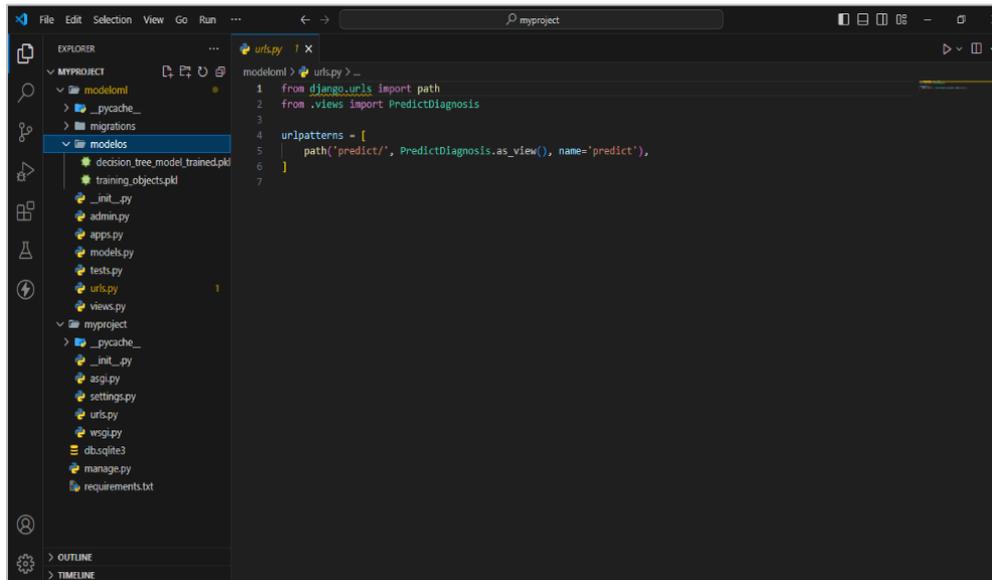
Fase 5: Despliegue

La Fase 5, con respecto a la metodología CRISP-ML(Q), se destaca por el despliegue del modelo de Machine Learning en producción, utilizando diversos entornos como:

Django: Es un framework de desarrollo web orientado al backend. Se seleccionó Django debido a su escalabilidad, su sólido respaldo comunitario y sus avanzadas características de seguridad. En este framework se integró el modelo de Machine Learning y se desarrolló una API que permite realizar solicitudes y consumir el modelo a través de Postman, facilitando su implementación en el sistema de laboratorio clínico.

Figura 78

Entorno Django donde se consume el Modelo ML

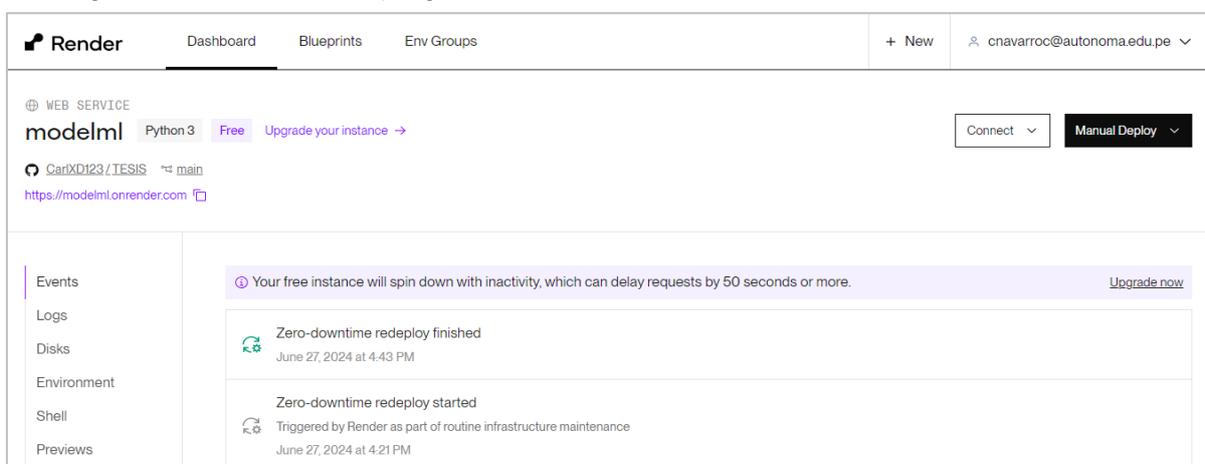


Nota. Configuración del entorno Django utilizado para consumir el modelo de Machine Learning, con la definición de rutas para gestionar predicciones a través de una API.

Render: Es una plataforma de hosting que permite desplegar sistemas tanto de frontend como de backend. En este caso, se utilizó Render para alojar la aplicación desarrollada en Django, habilitándola como una API funcional en el entorno de producción.

Figura 79

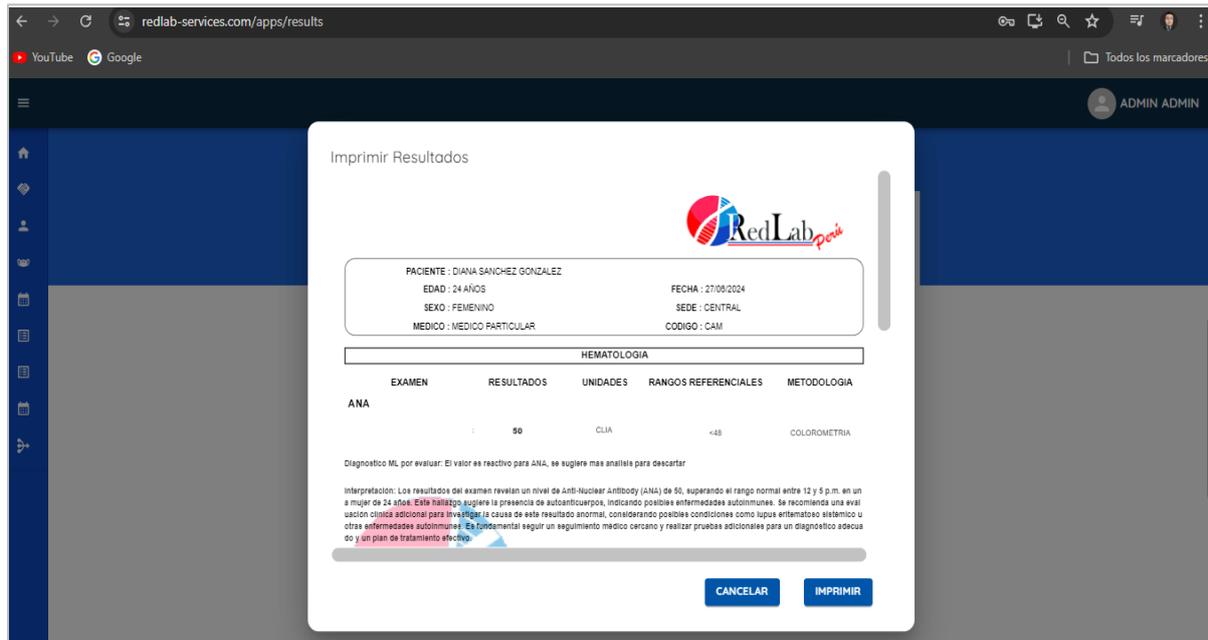
Hosting donde sube la API Django



Sistema Laboratorio: Es el sistema alojado en el servidor web en el cual se hará uso del modelo ML entrenado para realizar las predicciones.

Figura 80

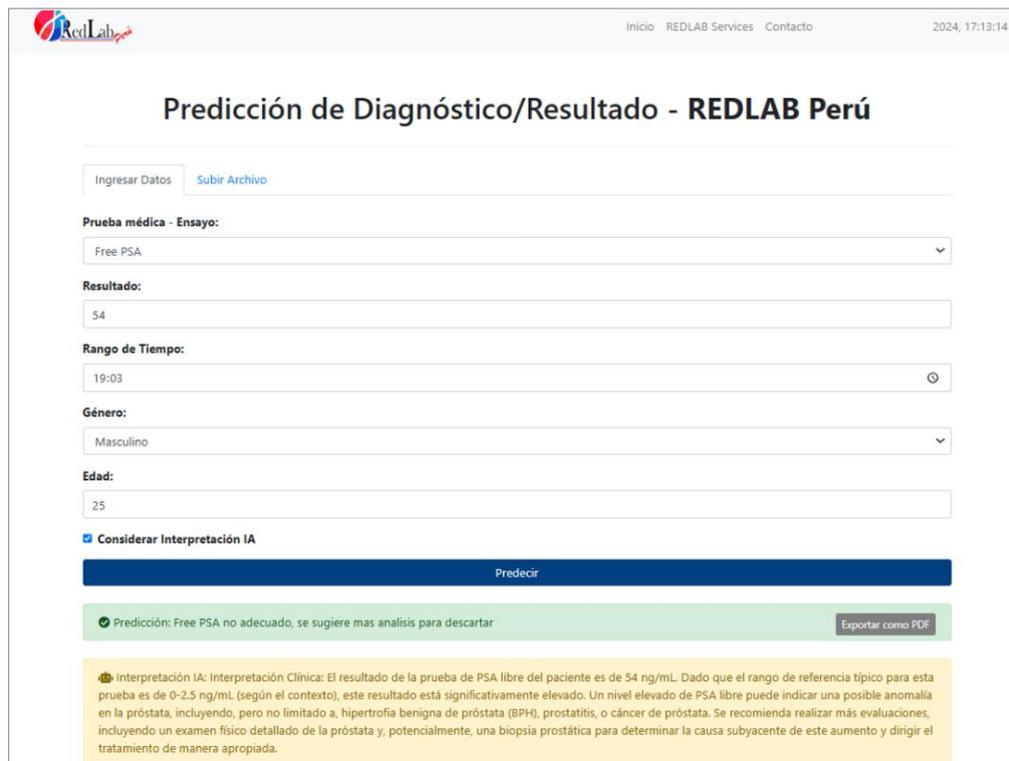
Informe de laboratorio de paciente con predicción de resultado clínico



Nota. Informe generado por el sistema, integrando resultados clínicos y predicciones del modelo de Machine Learning, presentado en un formato detallado y listo para impresión.

Figura 81

Predicción de diagnóstico/resultados de pruebas médicas



Nota. Interfaz web diseñada para la predicción de resultados médicos utilizando Machine Learning, permitiendo ingresar datos clínicos, generar diagnósticos automáticos y exportar resultados en formato PDF.

La Figura 80 muestra el informe generado para un paciente, el cual integra los resultados clínicos con predicciones obtenidas a través del modelo de Machine Learning. Este informe incluye información detallada como el tipo de examen, resultados obtenidos, unidades, rangos referenciales y metodología empleada. Además, se presenta un diagnóstico automatizado que proporciona una interpretación basada en los datos ingresados, todo en un formato claro, profesional y listo para impresión, facilitando su uso en entornos clínicos.

Por otro lado, la Figura 81 presenta la interfaz web utilizada para ingresar datos clínicos y realizar predicciones automáticas de diagnósticos médicos mediante Machine Learning. La interfaz permite al usuario ingresar manualmente valores como ensayo, resultado, rango de tiempo, género y edad, o subir un archivo con datos. Tras procesar los datos, el sistema genera un diagnóstico detallado con interpretaciones asociadas, que pueden exportarse en formato PDF. Este flujo asegura un análisis eficiente y preciso, con una salida adecuada para la toma de decisiones médicas.

Figura 82

Predicción de diagnóstico/resultado de pruebas médicas desde archivo

Paciente	Status	Gender	Age	Assay	Result	Result Flag	Interpretation	Time Prediction
Paciente 1	Completed	Male	55	Ferritin	75.60	El valor esta en el rango optimo, se sugiere mas analisis para descartar	El resultado del examen de ferritina para el paciente es de 75.6 ng/mL. Según los valores estándar de referencia, los niveles de ferritina oscilan entre 20-500 ng/mL en hombres y 20-200 ng/mL en mujeres. Por ende, el nivel de ferritina del paciente se encuentra dentro del rango normal. Esto indica una adecuada reserva de hierro, lo que sugiere que actualmente no hay evidencia de deficiencia ni sobrecarga de hierro. No se requieren acciones adicionales específicas con respecto a este resultado de ferritina, pero se recomienda continuar con el seguimiento rutinario y evaluaciones periódicas para monitorear el estado del hierro en el futuro.	ML Time: 0.0026 seconds DL Time: 4.4323 seconds
Paciente 2	Completed	Female	43	HIV Combo	0.04	El valor es no reactivo HIV Combo, se sugiere mas analisis para descartar	El examen médico realizado, específicamente la prueba HIV Combo, ha resultado en un valor de 0.04. Si asumimos que el rango de referencia considerado negativo es inferior a 1.00, este resultado indica que no se detectaron anticuerpos ni antígenos para VIH, sugiriendo que el paciente actualmente no está infectado con el virus de inmunodeficiencia humana. Es importante considerar que el periodo ventana debe ser tenido en cuenta, por lo que, dependiendo de la posible fecha de exposición, podría ser recomendable realizar pruebas adicionales después de un periodo apropiado. En ausencia de otros indicadores de infección, este resultado se considera tranquilizador, pero debe ser interpretado dentro del contexto clínico completo del paciente.	ML Time: 0.0033 seconds DL Time: 6.1229 seconds

Nota. Interfaz web que permite la carga de archivos para la predicción masiva de resultados médicos, generando diagnósticos automáticos y detallados a partir de datos ingresados en formato Excel.

Fase 6: Monitoreo y mantenimiento

En la Fase 6, se empleó Postman como herramienta clave para la validación y prueba de solicitudes hacia la API. Postman permite realizar pruebas detalladas que aseguran el correcto funcionamiento y rendimiento de la API en el entorno de producción. Además de las pruebas iniciales, Postman resulta fundamental para el monitoreo continuo, facilitando la detección proactiva y resolución de problemas. Esta herramienta permite verificar la integridad de los datos y asegura que el sistema funcione de manera eficiente y confiable, lo que contribuye al mantenimiento y actualización efectiva de la API.

El monitoreo continuo es fundamental para asegurar que el modelo de Machine Learning mantenga su rendimiento y precisión con el tiempo. Se implementaron alertas y análisis de logs para identificar cualquier anomalía o disminución en el desempeño del modelo. Además, se programaron revisiones periódicas del modelo y la API para confirmar que sigan cumpliendo con los requisitos del negocio y se ajusten a cambios en los datos o en las necesidades operativas. Este enfoque proactivo en mantenimiento y monitoreo garantiza la durabilidad y eficacia del sistema implementado. También se incluyó el análisis de métricas clave como la precisión y el recall en tiempo real, para detectar posibles desviaciones en el desempeño. Esto permite tomar medidas correctivas de manera inmediata, maximizando la eficiencia del sistema. Adicionalmente, se establecieron procedimientos claros para la actualización del modelo, asegurando que pueda adaptarse a nuevas necesidades sin interrumpir las operaciones.

También se implementaron mecanismos automatizados para la detección de inconsistencias en los datos y pruebas adicionales para garantizar la escalabilidad de la API. Este proceso asegura que el sistema esté preparado para manejar un

aumento en la carga de solicitudes y cambios inesperados en el entorno operativo, reforzando la confiabilidad y robustez del sistema en el tiempo. Asimismo, se realizaron simulaciones de escenarios críticos para evaluar la respuesta del sistema ante posibles fallos y puntos de saturación, lo que permitió optimizar su arquitectura. Finalmente, se estableció un protocolo de soporte técnico continuo, que incluye diagnósticos rápidos y acciones correctivas, para garantizar el funcionamiento ininterrumpido del sistema en un entorno dinámico.