

Desarrollo de una aplicación web para la predicción de la salud lumbar, aplicando técnicas de aprendizaje automático sobre las características biomecánicas de pacientes ortopédicos.

**Damaris Alarcón Vallejo**

Máster Universitario en Bioinformática y Bioestadística  
Inteligencia Artificial y Desarrollo web

**Romina Rebrij**

**Antoni Pérez Navarro**

24 de diciembre de 2021

Damaris Alarcón Vallejo

Damaris Alarcón Vallejo

Damaris Alarcón Vallejo



Esta obra está sujeta a una licencia de  
Reconocimiento-NoComercial-  
SinObraDerivada [3.0 España de Creative  
Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	<i>Desarrollo de una aplicación web para la predicción de la salud lumbar, aplicando técnicas de aprendizaje automático sobre las características biomecánicas de pacientes ortopédicos.</i>
<b>Nombre del autor:</b>	<i>Damaris Alarcón Vallejo</i>
<b>Nombre del consultor/a:</b>	<i>Romina Astrid Rebrij</i>
<b>Nombre del PRA:</b>	<i>Antoni Pérez Navarro</i>
<b>Fecha de entrega (mm/aaaa):</b>	12/2021
<b>Titulación:</b>	<i>Máster universitario en Bioinformática y Bioestadística</i>
<b>Área del Trabajo Final:</b>	<i>Inteligencia Artificial y desarrollo web</i>
<b>Idioma del trabajo:</b>	Castellano
<b>Número de créditos:</b>	15
<b>Palabras clave</b>	<i>Aprendizaje automático, Hernia discal, Espondilolistesis.</i>
<b>Resumen del Trabajo:</b>	<p>La columna vertebral es la estructura más importante del esqueleto humano y una de las partes fundamentales del sistema locomotor. La parte baja de la columna se denomina columna lumbar y está compuesta por vértebras más grandes, debido a que en esta zona se deposita la mayor parte del peso del cuerpo. El dolor en la región lumbar es una de las principales causas de consulta médica y se presenta con frecuencia en el 80-90% de la población adulta. La hernia de disco y la espondilolistesis son patologías degenerativas que afectan generalmente a la columna lumbar. La condición ortopédica de una persona se puede determinar a partir de sus características biomecánicas. El aprendizaje automático en el campo de la salud permite convertir datos clínicos, desde mediciones hasta imágenes, en conclusiones importantes para la toma de decisiones sobre el diagnóstico de enfermedades. Utilizando una base de datos de características biomecánicas de la columna, se evaluaron varios algoritmos de aprendizaje automático, el mejor modelo predictivo se obtuvo con el algoritmo de máquina de vectores de soporte (SVM), con una precisión del 85%. Utilizando este modelo se desarrolló una aplicación web con Shiny, en la que el usuario</p>

ingresa seis características biomecánicas de la columna y la aplicación devuelve el diagnóstico predicho por el modelo seleccionado. Esta aplicación fue desarrollada para facilitar a los médicos el diagnóstico de estas enfermedades, de manera que los pacientes puedan empezar de inmediato con el tratamiento adecuado.

**Abstract:**

The vertebral column is the most important structure of the human skeleton and one of the fundamental parts of the locomotor system. The lower part of the vertebral column is called the lumbar column and is made up of larger vertebrae, since most of the body's weight is deposited in this area. Pain in the lumbar region is one of the main causes of medical consultation and occurs frequently in 80-90% of the adult population. Herniated discs and spondylolisthesis are degenerative pathologies that generally affect the lumbar column. The orthopedic condition of a person can be determined from their biomechanical characteristics. Machine learning in the field of health allows converting clinical data, from measures to images, into important conclusions for decision making about the diagnosis of diseases. Using a database of biomechanical characteristics of the spine, several machine learning algorithms were evaluated, the best predictive model was obtained with the support vector machine (SVM) algorithm, with an accuracy of 85%. Using this model, a web application with Shiny was developed, in which the user enters six biomechanical characteristics of the vertebral column, and the application returns the diagnosis predicted by the selected model. This application was developed to make it easier for doctors to diagnose these diseases, so that patients can immediately start with the appropriate treatment.

# Índice

<b>1. Resumen</b>	<b>1</b>
<b>2. Introducción</b>	<b>1</b>
2.1 Contexto y justificación del Trabajo	1
2.2 Objetivos del Trabajo	3
2.3 Enfoque y método seguido	3
2.4 Planificación del Trabajo	5
2.5 Breve resumen de contribuciones y productos obtenidos	8
2.6 Breve descripción de los otros capítulos de la memoria	8
<b>3. Aprendizaje automático para el diagnóstico de enfermedades</b>	<b>8</b>
3.1 Diagnóstico de hernia de disco y espondilolistesis	10
3.2 Medidas biomecánicas de la columna	10
3.3 Algoritmos de aprendizaje automático	11
<b>4. Metodología</b>	<b>13</b>
4.1 Implementación de algoritmos de Aprendizaje Automático en RStudio	13
4.2 Desarrollo de la aplicación web en Shiny	16
<b>5. Resultados</b>	<b>17</b>
5.1 Métricas de los modelos desarrollados	17
5.2 Aplicación en Shiny	18
<b>6. Discusión</b>	<b>20</b>
<b>7. Conclusiones</b>	<b>21</b>
7.1 Conclusiones	21
7.2 Líneas de futuro	21
7.3 Seguimiento de la planificación	22
<b>8. Glosario</b>	<b>23</b>
<b>9. Bibliografía</b>	<b>24</b>

### **Lista de figuras**

Figura 1.- Algoritmos de aprendizaje automático.	4
Figura 2.- Funcionamiento de aplicaciones Shiny.	5
Figura 3.- Diagrama de Gantt planificado	7
Figura 4.- Medidas biomecánicas para el diagnóstico de hernia y espondilolistesis.	10
Figura 5.- Boxplot de las medidas biomecánicas.	14
Figura 6.- Número de observaciones por diagnóstico.	14
Figura 7.- Datos de entrenamiento balanceados.	15
Figura 8.- Imagen de la interfaz con valores erróneos.	19
Figura 9.- Imagen de la interfaz con el diagnóstico predicho.	20

## **Lista de tablas**

Tabla 1.- Planificación con hitos y temporización.	7
Tabla 2.- Ejemplos de aplicaciones de aprendizaje automático en datos médicos.	9
Tabla 3.- Resumen de las métricas de las variables.	13
Tabla 4.- Resumen de las métricas obtenidas.	18

# 1. Resumen

La columna vertebral es la estructura más importante del esqueleto humano y una de las partes fundamentales del sistema locomotor. La parte baja de la columna se denomina columna lumbar y está compuesta por vértebras más grandes, debido a que en esta zona se deposita la mayor parte del peso del cuerpo. El dolor en la región lumbar es una de las principales causas de consulta médica y se presenta con frecuencia en el 80-90% de la población adulta. La hernia de disco y la espondilolistesis son patologías degenerativas que afectan generalmente a la columna lumbar. La condición ortopédica de una persona se puede determinar a partir de sus características biomecánicas. El aprendizaje automático en el campo de la salud permite convertir datos clínicos, desde mediciones hasta imágenes, en conclusiones importantes para la toma de decisiones sobre el diagnóstico de enfermedades. Utilizando una base de datos de características biomecánicas de la columna, se evaluaron varios algoritmos de aprendizaje automático, el mejor modelo predictivo se obtuvo con el algoritmo de máquina de vectores de soporte (SVM), con una precisión del 85%. Utilizando este modelo se desarrolló una aplicación web con Shiny, en la que el usuario ingresa seis características biomecánicas de la columna y la aplicación devuelve el diagnóstico predicho por el modelo seleccionado. Esta aplicación fue desarrollada para facilitar a los médicos el diagnóstico de estas enfermedades, de manera que los pacientes puedan empezar de inmediato con el tratamiento adecuado.

## 2. Introducción

### 2.1 Contexto y justificación del Trabajo

La columna vertebral es la estructura más importante del esqueleto humano y una de las partes fundamentales del sistema locomotor [1]. Está formada por 33 vértebras, entre las cuales se encuentran los discos fibrocartilaginosos. Las vértebras se clasifican según su ubicación en: 7 cervicales, 12 torácicas, 5 lumbares, 5 sacras y 3-4 coccígeas [2]. Entre las funciones principales de la columna vertebral están: dar soporte al cuerpo, permitir la estabilidad y proteger a la médula espinal, compuesta por tejido nervioso sensible [1]. La parte baja de la columna se denomina región lumbar, está compuesta por vértebras más grandes y gruesas, debido a que en esta zona se deposita la mayor parte del peso del cuerpo, lo que hace que esta zona sea más propensa a la aparición de dolores y afecciones [2].

El dolor en la región lumbar es una de las principales causas de consulta médica y se presenta con frecuencia en el 80-90% de la población adulta [3]. La hernia de disco y la espondilolistesis son patologías degenerativas que afectan

generalmente a la columna lumbar, siendo algunas de las posibles causas del dolor en esta zona. La hernia de disco consiste en la salida del material gelatinoso contenido en el disco intervertebral hacia el canal raquídeo [4]. Por otro lado, la espondilolistesis es el desplazamiento de una vértebra, por lo general hacia adelante, de la vértebra que se encuentra debajo [5]. Ambas patologías generan dolor a causa de la compresión de las raíces nerviosas.

La condición de salud ortopédica de una persona frente a estas y otras patologías se puede determinar a partir de sus características biomecánicas. La biomecánica es la ciencia que analiza la mecánica de los movimientos corporales, desde su origen hasta sus efectos, explicando las razones por las que un cuerpo se mueve de la manera en que lo hace y las características de los huesos involucrados [21]. Las características biomecánicas derivadas de la forma y orientación de la columna lumbar y de la pelvis, describen la relación de la columna con respecto al piso y con respecto a la pelvis [18]. Estas características influyen en el comportamiento del balance sagital en condiciones normales y patológicas de la columna [22]. El balance sagital se refiere al equilibrio que debe mantenerse entre la columna y la pelvis, para lograr una postura correcta de bipedestación y marcha, alineando adecuadamente la cabeza, hombros, pelvis y pies con el menor gasto energético posible y el menor impacto sobre articulaciones, músculos y ligamentos [18].

Las curvaturas normales de la columna y su alineación con la pelvis permiten la distribución equitativa de diferentes fuerzas y movimientos de la columna. El desequilibrio de estas fuerzas se da generalmente por deformaciones que van presentándose con el paso de los años y por patologías que afectan a la columna vertebral como la hernia de disco y la espondilolistesis [23]. Este desequilibrio provoca cambios adaptativos en la pelvis y extremidades inferiores, alterando los parámetros biomecánicos normales; se altera la angulación de la pelvis, forzando las caderas y rodillas a flexionarse en diferentes grados, con el fin de compensar los cambios que se producen en la alineación de la columna. Estos cambios adaptativos implican un mayor gasto de energía y aumentan el esfuerzo muscular [24]. Analizar las alteraciones en estos parámetros biomecánicos sirve como línea de base para la evaluación de condiciones patológicas [25].

Muchos clínicos encuentran difícil diferenciar entre los síntomas causados por las diversas afecciones de la columna, ya que la sintomatología referida por los pacientes es únicamente el dolor en la región lumbar [6]. En este sentido, el desarrollo de software basado en aprendizaje automático es de gran importancia en la actualidad. Varios estudios han desarrollado herramientas que optimizan el diagnóstico de enfermedades, mejoran la medicina preventiva, facultan la medicina personalizada, además de disminuir los costos y aumentar la productividad de los profesionales de la salud [7].

Con el fin de optimizar el diagnóstico tanto de hernia de disco como de espondilolistesis y diferenciarlos de pacientes sanos, se pueden aplicar varias

técnicas de aprendizaje automático a las características de la pelvis y columna, para determinar un modelo que clasifique automáticamente a pacientes ortopédicos con este tipo de afecciones. Emplear dicho modelo para el desarrollo de una aplicación web interactiva, permite interpretar rápidamente las características biomecánicas de nuevos pacientes y facilita el diagnóstico para los médicos, de manera que los pacientes puedan empezar de inmediato con el tratamiento necesario.

## 2.2 Objetivos del Trabajo

### 2.2.1 Objetivo General:

Desarrollar una aplicación web que permita predecir dos patologías de la salud lumbar en base a características biomecánicas de la pelvis y columna lumbar.

### 2.2.2 Objetivos específicos:

- Evaluar diversas técnicas de clasificación para determinar el mejor modelo predictivo.
- Optimizar el modelo seleccionado para lograr una precisión mínima del 75%.
- Desarrollar una aplicación web interactiva en base al modelo seleccionado.

## 2.3 Enfoque y método seguido

El aprendizaje automático es ampliamente utilizado en el campo de la salud con el objetivo de procesar datos y obtener información relevante a partir de los mismos. Para el desarrollo del presente trabajo se utilizó la base de datos "column3Cweka.csv" [8] disponible en la página de UCI Machine Learning Repository. Se trata de un conjunto de datos biomecánicos construido por el Dr. Henrique da Mota durante un periodo de residencia médica en el grupo de investigación aplicada en ortopedia, GARO por sus siglas en inglés, del Centro Médico Quirúrgico de Readaptación des Massues en Lyon-Francia. La base de datos consta de 310 pacientes, de los cuales se consideran 6 características biomecánicas, provenientes de la forma y orientación de la pelvis y de la columna lumbar. Se clasifican en tres clases según el diagnóstico de cada paciente; Normal, Hernia de disco y Espondilolistesis [8].

Debido a las características de la base de datos, se aplicó el enfoque de clasificación mediante aprendizaje supervisado, ya que se conoce la etiqueta de diagnóstico de los pacientes. El aprendizaje supervisado intenta optimizar una función (modelo) para encontrar la combinación adecuada de los valores de las características que dan como resultado una determinada etiqueta [9]. Existen varios algoritmos de aprendizaje supervisado:

Model	Learning task
<b>Supervised Learning Algorithms</b>	
Nearest Neighbor	Classification
Naive Bayes	Classification
Decision Trees	Classification
Classification Rule Learners	Classification
Linear Regression	Numeric prediction
Regression Trees	Numeric prediction
Model Trees	Numeric prediction
Neural Networks	Dual use
Support Vector Machines	Dual use

Figura 1.- Algoritmos de aprendizaje automático.  
[9]

El objetivo es clasificar a los pacientes ortopédicos, por lo que los algoritmos de clasificación son los indicados para este caso. Para la implementación de este tipo de algoritmos en general se siguen los siguientes pasos [9]:

1. *Recopilar datos*, se recolectan los datos necesarios para el estudio y se analiza su composición, así como la fuente de los mismos.
2. *Explorar y preprocesar datos*, se analiza las características de las variables y lo que representa cada parámetro, se determina la clase y se aplican técnicas para modificar los datos, según sea necesario, antes de utilizarlos para el entrenamiento de los modelos. También se divide la base de datos en dos grupos; entrenamiento y evaluación.
3. *Construir y entrenar un modelo en los datos*, se implementa los diferentes algoritmos de aprendizaje automático al grupo de datos de entrenamiento, es aquí donde el modelo “aprende” sobre estos datos, se establecen parámetros para evaluar datos futuros.
4. *Evaluar el rendimiento del modelo*, para esto se utiliza el grupo de datos de evaluación, se comparan las etiquetas devueltas por el modelo con las etiquetas reales de estos datos.
5. *Mejorar el rendimiento del modelo*, se modifican parámetros utilizados por cada algoritmo, con el objetivo de obtener un mejor rendimiento.
6. *Evaluar el modelo mejorado*, se evalúa nuevamente el rendimiento del modelo con el grupo de datos de evaluación.

Se aplicaron todos estos pasos para el desarrollo de los modelos, los algoritmos seleccionados se evaluaron en el entorno RStudio. Después de analizar los resultados obtenidos, se seleccionó el mejor modelo predictivo para desarrollar la aplicación web empleando el paquete Shiny de R.

Shiny es un paquete de R que facilita el desarrollo de aplicaciones web interactivas, adicionalmente Shiny cuenta con su propio servidor donde estas aplicaciones pueden ser publicadas [27]. Desarrollar una aplicación web con Shiny requiere generar un fichero *app.R*, el cual debe contener dos secciones bien definidas: *ui* y *server*. La sección *ui* define el diseño de la interfaz de la aplicación, mientras que la sección *server* contiene las instrucciones o cálculos que la aplicación debe realizar con los datos ingresados [28].

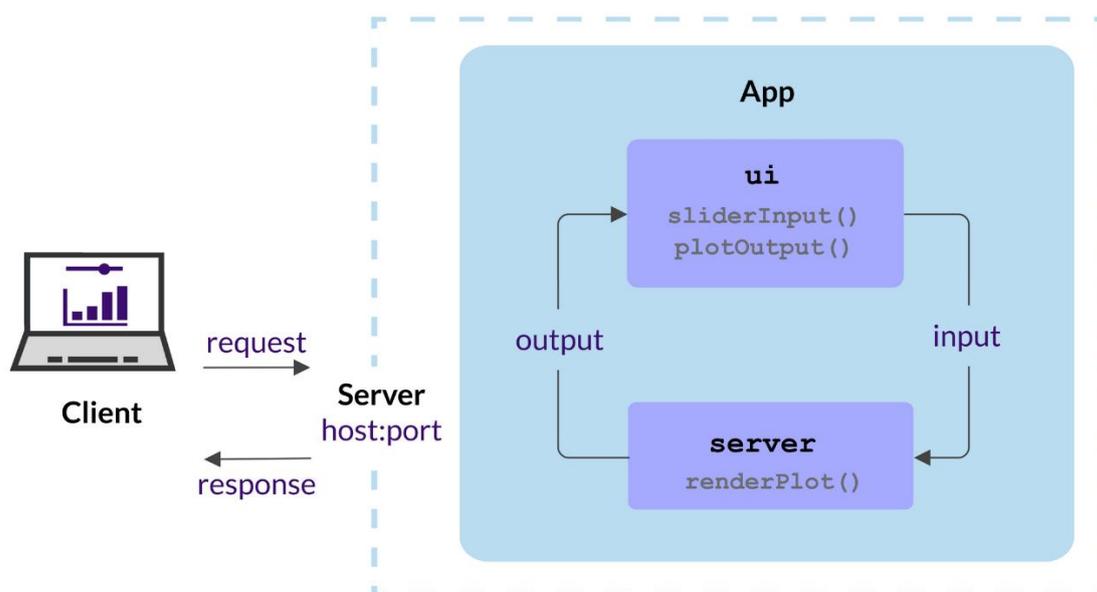


Figura 2.- Funcionamiento de aplicaciones Shiny.  
[28]

#### 2.4 Planificación del Trabajo

Descripción de tareas y fechas de desarrollo de cada una:

Descripción	Fecha de inicio	Fecha de fin	Hito
<b>Definición del plan de trabajo. PEC1.</b>	<b>23/sep/2021</b>	<b>04/oct/2021</b>	
Recopilación y análisis exploratorio de los datos.	23/sep/2021	26/sep/2021	
Definición de los modelos que serán aplicados.	26/sep/2021	29/sep/2021	
Desarrollo del plan de trabajo. PEC1.	29/sep/2021	04/oct/2021	
Entrega y retroalimentación del plan de trabajo.	04/oct/2021	11/oct/2021	<b>Entrega de PEC1</b>
<b>Desarrollo del trabajo. Fase 1. PEC2.</b>	<b>05/oct/2021</b>	<b>08/nov/2021</b>	

Exploración de las técnicas y herramientas necesarias.	05/oct/2021	08/oct/2021	
Entender el formato de entrada de los datos que requiere cada modelo.	05/oct/2021	08/oct/2021	
Preprocesamiento de datos.	08/oct/2021	13/oct/2021	
Construcción y entrenamiento de los 4 modelos.	14/oct/2021	25/oct/2021	
Validar el rendimiento de los 4 modelos.	14/oct/2021	25/oct/2021	
Documentación de todo el trabajo realizado para la memoria.	26/oct/2021	29/oct/2021	
Mejorar el rendimiento de los modelos construidos.	30/oct/2021	06/nov/2021	
Validar el rendimiento de los modelos mejorados.	30/oct/2021	06/nov/2021	
Optimizar esquema y presentación la PEC 2.	07/nov/2021	08/nov/2021	
Entrega y retroalimentación de la fase 1.	08/nov/2021	15/nov/2021	<b>Entrega de PEC2</b>
<b>Desarrollo del trabajo. Fase 2. PEC3.</b>	<b>09/nov/2021</b>	<b>09/dic/2021</b>	
Análisis de resultados.	09/nov/2021	12/nov/2021	
Selección del modelo que alcance la precisión deseada.	09/nov/2021	12/nov/2021	
Aprendizaje de desarrollo de aplicaciones empleando Shiny.	13/nov/2021	20/nov/2021	
Definición del esquema de la aplicación.	16/nov/2021	02/dic/2021	
Desarrollo de la aplicación web con Shiny.	21/nov/2021	02/dic/2021	
Subir aplicación al servidor de Shiny.	01/dic/2021	04/dic/2021	
Probar funcionamiento de la aplicación.	01/dic/2021	04/dic/2021	
Documentación de todo el trabajo realizado para la memoria.	05/dic/2021	09/dic/2021	

Optimizar esquema y presentación la PEC 3.	05/dic/2021	09/dic/2021	<b>Entrega de PEC3</b>
<b>Cierre de la memoria. PEC4.</b>	<b>10/dic/2021</b>	<b>24/dic/2021</b>	<b>Entrega de PEC4 Memoria</b>
Optimizar esquema y presentación de la memoria.	10/dic/2021	24/dic/2021	
Optimizar esquema y presentación del código fuente.	10/dic/2021	24/dic/2021	
<b>Elaboración de la presentación. PEC5a.</b>	<b>27/dic/2021</b>	<b>03/ene/2022</b>	<b>Entrega de PEC5a Presentación</b>
<b>Defensa pública. PEC5b.</b>	<b>13/ene/2022</b>	<b>21/ene/2022</b>	<b>Entrega de PEC5b Defensa</b>

Tabla 1.- Planificación con hitos y temporización.

Diagrama de Gantt:

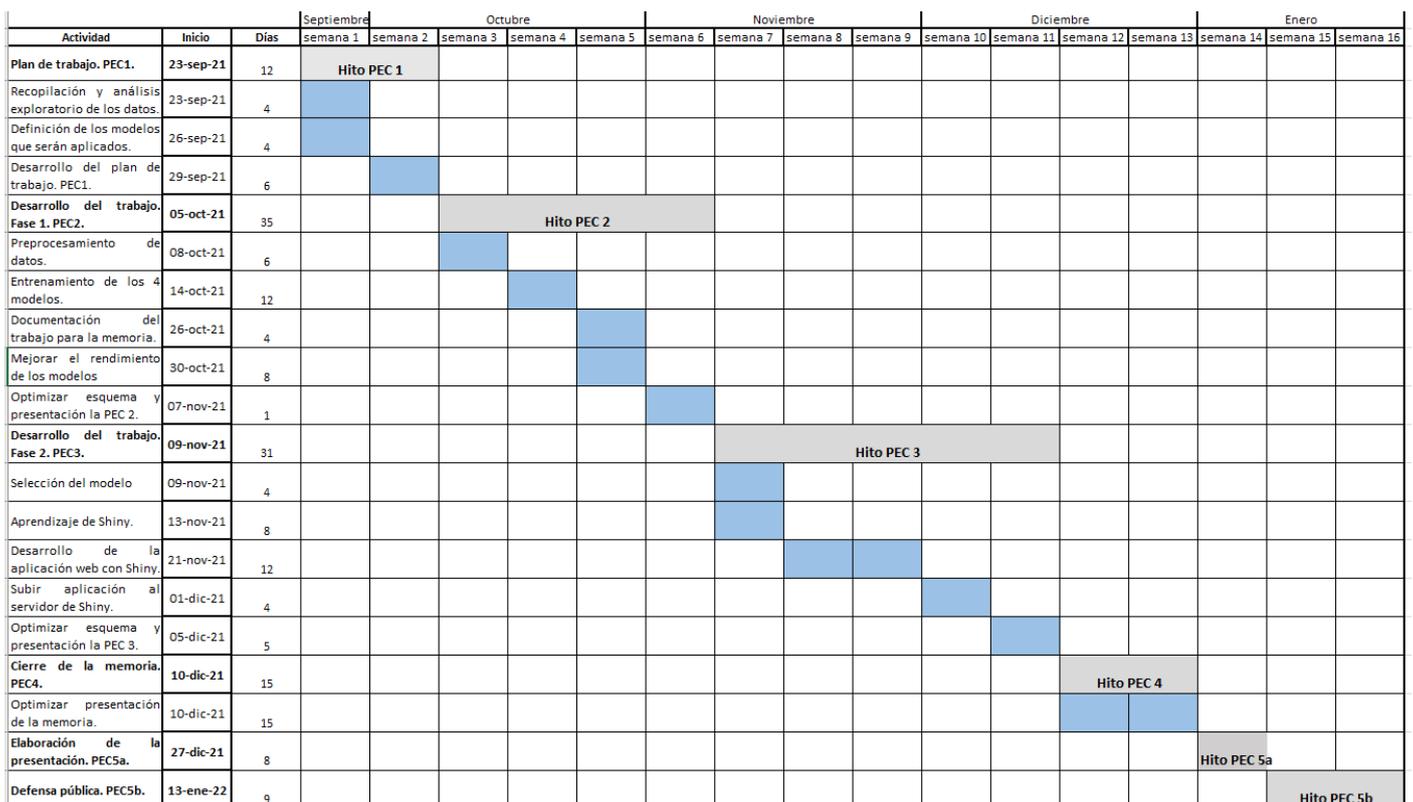


Figura 3.- Diagrama de Gantt planificado

## 2.5 Breve resumen de contribuciones y productos obtenidos

Como producto final de este trabajo se obtuvo la aplicación web interactiva funcionando según lo planificado. El enlace para acceder a la aplicación es:

[https://damaris24.shinyapps.io/Predictor\\_de\\_salud\\_lumbar/](https://damaris24.shinyapps.io/Predictor_de_salud_lumbar/)

Todo el código desarrollado tanto para el entrenamiento de los modelos como para el desarrollo de la aplicación web en Shiny se encuentra disponible en el repositorio de GitHub, se puede acceder a estos mediante el siguiente enlace:

[https://github.com/DamarisA91/Predictor\\_de\\_la\\_salud\\_lumbar](https://github.com/DamarisA91/Predictor_de_la_salud_lumbar)

Adicionalmente, se obtuvo la presente memoria que detalla el procedimiento completo hasta llegar a los resultados obtenidos, así como una presentación virtual que demuestra el funcionamiento de la aplicación.

## 2.6 Breve descripción de los otros capítulos de la memoria

Una vez planteado el contexto del presente trabajo, en los siguientes capítulos se profundiza el desarrollo de la aplicación, así como las técnicas aplicadas y los resultados obtenidos. La sección “Aprendizaje automático para el diagnóstico de enfermedades” describe las aplicaciones del aprendizaje automático en el diagnóstico médico, el diagnóstico tradicional de las patologías aquí tratadas y una descripción de los datos que ayudan a desarrollar un modelo predictivo para este caso. La metodología, detalla las técnicas de aprendizaje automático aplicadas, el preprocesamiento de los datos, los pasos para la validación de los modelos y el desarrollo del producto final, la aplicación en Shiny. El capítulo de resultados detalla las métricas obtenidas sobre la precisión, sensibilidad, especificidad y valor kappa de todos los modelos entrenados y el análisis respectivo para seleccionar el mejor modelo sobre el cual se desarrolló la aplicación. Finalmente, en discusión y conclusiones, se comparan los resultados planteados inicialmente con el producto final, analizando los métodos y técnicas empleadas para lograr el resultado obtenido.

# 3. Aprendizaje automático para el diagnóstico de enfermedades

Un diagnóstico acertado es la parte más importante del tratamiento de cualquier enfermedad [10]. Algunos signos y síntomas no son específicos para ciertas enfermedades, lo que representa un desafío para los médicos al momento de identificar una patología. En la actualidad, se está investigando y aplicando técnicas de aprendizaje automático para acelerar el diagnóstico de varias enfermedades, incluyendo el cáncer. El aprendizaje automático es un enfoque de la inteligencia artificial, cuyo objetivo es descubrir patrones en cantidades masivas de datos que incluyen imágenes, palabras o valores numéricos [11]. Se

utilizan datos de entrenamiento etiquetados, los cuales son analizados mediante algoritmos que “aprenden” el comportamiento de estos datos y establecen modelos, a partir los cuales se puede realizar predicciones sobre nuevos datos [12]. Las técnicas de aprendizaje automático han ganado popularidad debido a que ayudan a la detección temprana de enfermedades de manera rápida y precisa [10]. Varios algoritmos de clasificación como K-vecinos más cercanos, árboles de decisión, máquina de vectores de soporte (SVM) y random forest se han implementado para el desarrollo de modelos que permitan tomar decisiones exitosas en la industria de la salud [10]. La siguiente tabla obtenida del artículo “Aplicaciones de modelos predictivos de aprendizaje automático en el diagnóstico de enfermedades crónicas” [13], muestra algunos ejemplos de aplicaciones de algoritmos de aprendizaje automático en diferentes datos médicos para la predicción de enfermedades:

CD Diagnosis	Study Type	Input Features	Outcomes	Models
Hepatic fibrosis	Cross-sectional	Age, sex and RTE images	Accuracy, Sensitivity, and Specificity	NB, RF, KNN, SVM, and NN
Chronic hepatitis B stages	Case study	Gene expressions	Precision and AU-ROC	RF, KNN, SVM
COPD exacerbation events	Retrospective	COPD symptoms	TP, FP, ROC	BN
Aggravating event identification of COPD	Longitudinal	EDGE digital health system	AU-ROC	LR
Exacerbations of COPD patients	Case-control	Equi-ripple bandpass (BP)	Sensitivity, specificity, accuracy, PPV, NPV	PCA coupled SVM
Diabetes classification	Case study	Age and clinical data	Sensitivity, specificity, accuracy, AU-ROC	LR, ANN, NB, KNN, and RF
Glomerulus filtration rate estimation	Retrospective cohort study (RCT)	Age, sex, and serum creatinine 99mTc-DTPA imaging	Accuracy	ANN, SVM
Asthma exacerbations events	Case-control	Telemonitoring data	Sensitivity, specificity, accuracy	NB, adaptive Bayesian network, and SVM
Stage of lung cancer	Prospective cohort study	Cyrano’s 320 sensor device, age	Accuracy, sensitivity, and specificity	SVM
Pulmonary function tests	RCT	Blood analysis, lung images	Accuracy	DT
Dementia prediction	Case-control	MRI	Accuracy, precision, and specificity	SVM
Identification of ischemic stroke lesions	Cross-sectional	MRI	Accuracy	NB
Course of depression	Case study	A shortened version of the IDS (QIDS)	Accuracy	LR
Late-life dementia assessment	Prospective cohort	MRI/CT, Blood Tests	ROC, AUC and, MCA	SVM
Degenerative movement disorders	Cross-sectional	Pathological	Not defined	Hierarchical clustering analyses
Checking CT imaging effectiveness	Case study	CT images, Age, and sex	Accuracy, AU-ROC	NN
Discriminatory peptide identification of heart failures	Experimental	Age, sex, and renal function	Sensitivity, specificity	SVM

Tabla 2.- Ejemplos de aplicaciones de aprendizaje automático en datos médicos. [12]

En el campo de la medicina se genera una gran cantidad de información que puede ser aprovechada utilizando aprendizaje automático, para facilitar la identificación de enfermedades y minimizar las fallas de diagnóstico.

### 3.1 Diagnóstico de hernia de disco y espondilolistesis

Las patologías que afectan a la columna vertebral son muy frecuentes en la población y pueden desencadenar efectos negativos en la calidad de vida de los pacientes [26]. Debido a esto es muy importante fortalecer el procedimiento para lograr un diagnóstico acertado, que permita al especialista aplicar un tratamiento adecuado. Este trabajo se enfocó en la detección de dos de las patologías que afectan a la columna lumbar: la hernia de disco y la espondilolistesis. Ambas enfermedades se diagnostican por lo general mediante imágenes de resonancia magnética de la columna vertebral [18]. Sin embargo, en algunos casos el diagnóstico desde el punto de vista imagenológico resulta difícil y controversial, ya que existen diferentes criterios de clasificación para estas patologías [14].

### 3.2 Medidas biomecánicas de la columna

Con el objetivo de fortalecer el diagnóstico de hernia de disco, espondilolistesis y de la condición normal de la columna lumbar, es posible utilizar información recopilada a partir de imágenes de resonancia magnética de la columna [15]. Los parámetros radiológicos que se pueden emplear son: incidencia pélvica (PI), inclinación pélvica (PT), pendiente sacra (SS), lordosis lumbar (LA), radio pélvico (PR) y el grado de espondilolistesis. Estas medidas describen la orientación y el grado de inclinación tanto de la columna como de la pelvis.

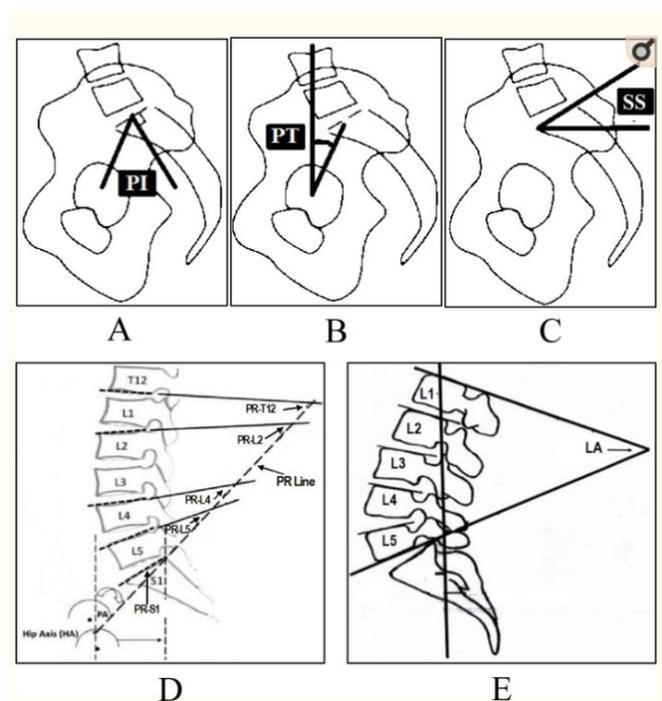


Figura 4.- Medidas biomecánicas para el diagnóstico de hernia y espondilolistesis. [15]

### 3.3 Algoritmos de aprendizaje automático

Un modelo predictivo recibe pautas claras sobre lo que necesita aprender y cómo hacerlo, debido a esto se conoce como aprendizaje supervisado al procedimiento de entrenamiento de este tipo de modelos [9]. Una de las tareas más frecuentes de este tipo de aprendizaje automático es la clasificación, donde los modelos predicen la categoría a la que pertenece una observación. Dentro de los algoritmos de clasificación se encuentran: K-vecinos más cercanos, árboles de decisión, máquina de vectores de soporte (SVM) y bosque aleatorio, algoritmos que se utilizaron en este trabajo.

- K-vecinos más cercanos, este algoritmo define un ejemplo no etiquetado de acuerdo a sus características, le asigna la misma clase de otros ejemplos etiquetados con características similares. Aplicando la teoría de que “es probable que las cosas que son iguales tengan propiedades similares” [9]. Selecciona los ejemplos etiquetados más cercanos o con menor distancia, de acuerdo a la función que se utilice. Este algoritmo se ha utilizado en estudios para la detección de fibrosis hepática, hepatitis b, clasificación de diabetes, entre otros [12].

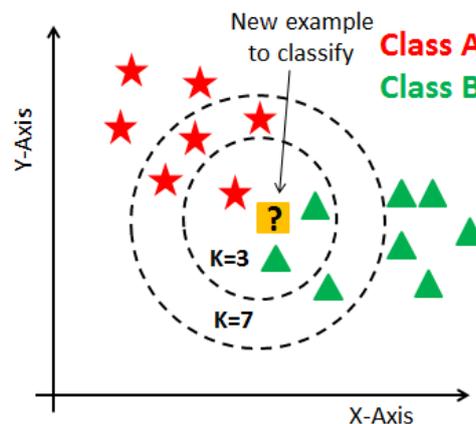


Figura 5.- Funcionamiento del algoritmo KNN.  
[29]

- Árboles de decisión / Decision trees, estos algoritmos son clasificadores muy poderosos que trabajan de manera similar a la estructura de un árbol para modelar las relaciones entre las variables y las clases. Empieza como un tronco ancho que se divide en ramas que se hacen cada vez más estrechas. Cuando es un modelo clasificador, las ramas son decisiones que el modelo va tomando para llegar a una predicción [9]. Este algoritmo se ha utilizado en estudios para detectar funciones pulmonares [12].

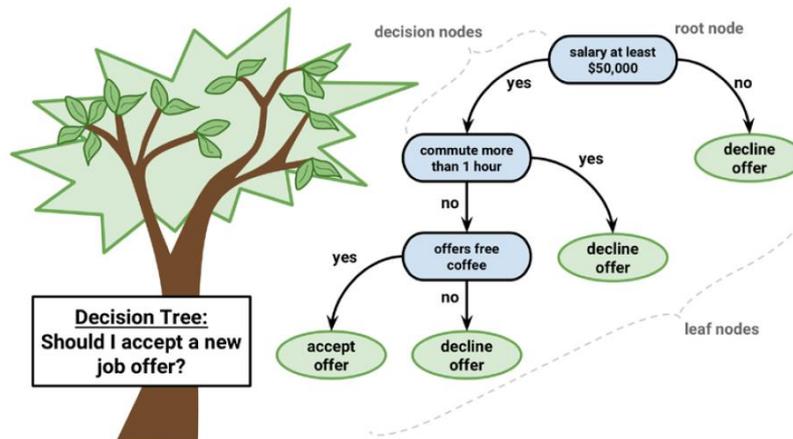


Figura 6.- Funcionamiento del algoritmo árbol de decisión.  
[9]

- Máquina de vectores de soporte, es un algoritmo de aprendizaje supervisado utilizado para la clasificación de grupos. Este algoritmo busca hiperplanos que dividan los grupos, tratando de seleccionar la recta que tiene mayor distancia hacia los puntos de cada clase. Esta distancia se conoce como margen y le da flexibilidad al modelo para poder utilizarlo sobre nuevos datos, evitando así el sobre ajuste del modelo a los datos de entrenamiento [30]. Este algoritmo se ha utilizado en estudios para el diagnóstico de cáncer de pulmón, estados de demencia, fallos cardíacos, entre otros [12].

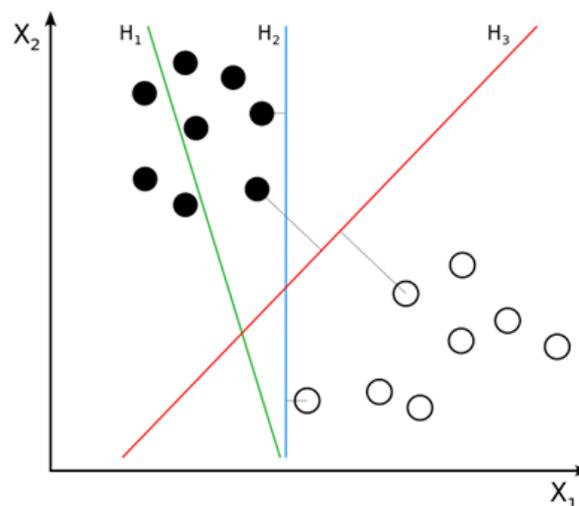


Figura 7.- Funcionamiento del algoritmo máquina de vectores de soporte.  
[30]

- Bosque aleatorio / Random forest, este algoritmo implementa el aprendizaje conjunto, que es un proceso que combina múltiples clasificadores para resolver un problema complejo y mejorar el rendimiento del modelo. En este caso, el algoritmo de bosque aleatorio contiene varios árboles de decisión en varios subconjuntos de una base de datos determinada y obtiene el promedio para mejorar la precisión

predictiva del modelo. Mientras más árboles de decisión, se previene el sobre ajuste y se tendrá una mayor precisión [31].

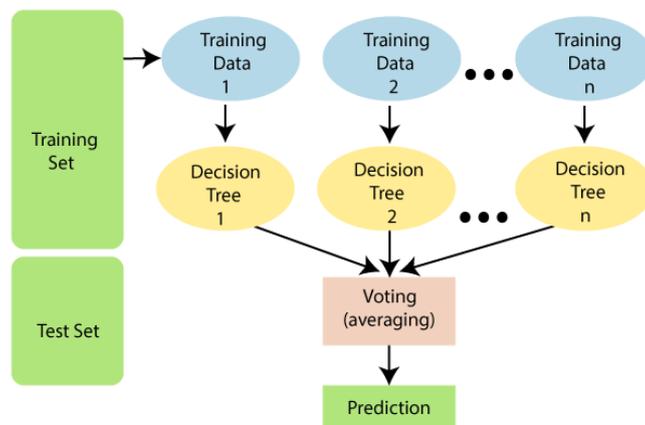


Figura 8.- Funcionamiento del algoritmo bosque aleatorio. [31]

## 4. Metodología

### 4.1 Implementación de algoritmos de Aprendizaje Automático en RStudio

Teniendo en cuenta las características de la base de datos de referencia, “column3Cweka.csv”, se aplicó cuatro algoritmos de aprendizaje automático: k-NN (vecino más cercano), árbol de decisión, SVM (máquina de vectores de soporte) y bosque aleatorio o random forest. Para el entrenamiento de los modelos se siguió los siguientes pasos:

#### 4.1.1 Exploración y preprocesamiento de datos:

Se cargó la base de datos “column3Cweka.csv” y se realizó un análisis exploratorio de sus características. Contiene 310 pacientes de los cuales; 60 pacientes tienen hernia de disco, 150 pacientes tienen espondilolistesis y 100 están sanos, por lo que están etiquetados con un diagnóstico “normal”. Se presentan 7 variables: incidencia pélvica (IP), inclinación pélvica (PT), pendiente sacra (SS), lordosis lumbar (LA), radio pélvico (PR), grado de espondilolistesis y diagnóstico. Las 6 primeras variables describen las características biomecánicas, son numéricas y al explorar el resumen estadístico de los datos se observó que la variable “grado de espondilolistesis” presenta una gran diferencia entre la media y la mediana, lo que sugiere la presencia de outliers o valores atípicos, esto se pudo evidenciar con un boxplot de todas las variables.

	Incidencia pélvica (PI)	Inclinación pélvica (PT)	Pendiente sacra (SS)	Lordosis lumbar (LA)	Radio pélvico (PR)	Grado de espondilolistesis
<b>Media</b>	60,50	17,54	51,93	42,95	117,92	29,30
<b>Mediana</b>	58,69	16,35	49,56	42,40	118,27	11,76
<b>Mínimo</b>	26,15	-6,55	14,00	13,37	70,08	-11,05
<b>Máximo</b>	129,83	49,43	125,74	121,43	163,07	418,54

Tabla 3.- Resumen de las métricas de las características biomecánicas.

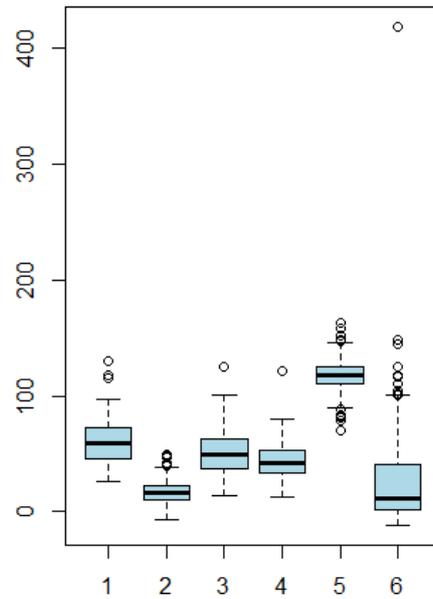


Figura 9.- Boxplot de las medidas biomecánicas.

Se identificó la clase, en este caso es el diagnóstico y tiene tres niveles: “Hernia”, “Espondilolistesis” y “Normal”. Para facilitar el entrenamiento de los modelos se cambió esta variable a factor y se analizó la cantidad de observaciones de cada diagnóstico:

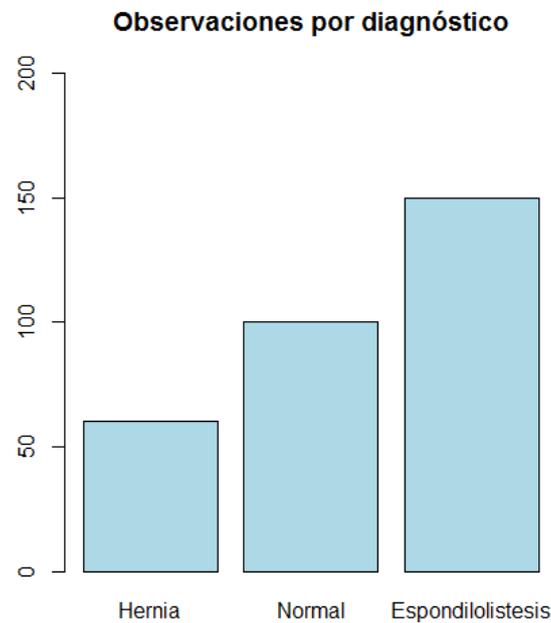


Figura 10.- Número de observaciones por diagnóstico.

Posteriormente se dividió la base de datos en dos partes, 70% para entrenamiento y 30% para evaluación. Se balancearon los datos de entrenamiento para evitar que las clases con mayor número de observaciones tengan mayor influencia en el entrenamiento del modelo.

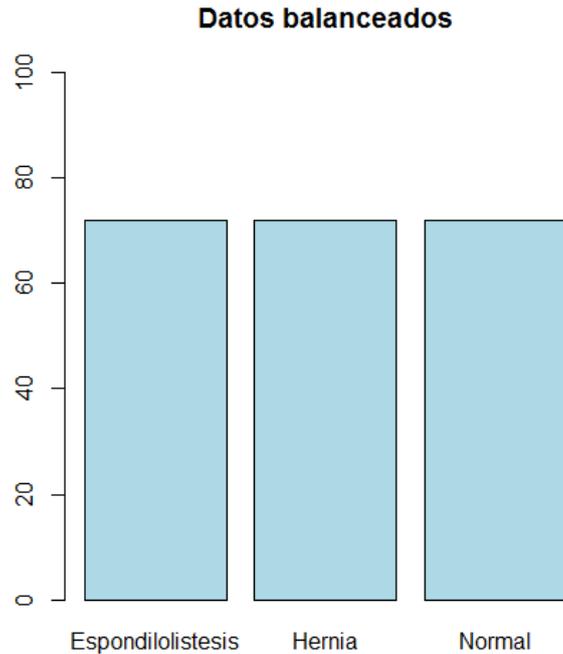


Figura 11.- Datos de entrenamiento balanceados.

Al analizar los datos se observó que la escala en la que se mide cada característica biomecánica es muy diferente, por lo que se requiere escalar los datos. Se utilizó la técnica Z-score; se guardó las medias y desviaciones estándar de cada parámetro de los datos de entrenamiento para escalar los datos de entrenamiento y de evaluación. Estos parámetros se utilizaron también para escalar los nuevos datos ingresados en la aplicación.

Por otro lado, se guardó la clase de los datos de evaluación en el vector "test\_lab", para obtener el rendimiento de los modelos predictivos.

#### 4.1.2 Construcción y mejora de los modelos

Teniendo en cuenta estas características de la base de datos utilizada, se seleccionó cuatro técnicas de aprendizaje automático para evaluar en estos datos: k-NN (vecino más cercano), árbol de decisión, SVM (máquina de vectores de soporte) y random forest, ya que estos algoritmos funcionan bien con variables numéricas y en su mayoría no son muy influenciados por datos atípicos [9].

Para el algoritmo de kNN se utilizó la función *knn* del paquete *class* con  $k=13$  ya que es aproximadamente la raíz de  $n=215$  [9] y distancia euclídea. Para optimizar el funcionamiento del modelo se modificó el valor de  $k$  hasta obtener un mejor rendimiento, adicionalmente se desarrolló un algoritmo que permita modificar la distancia utilizada por el modelo para identificar a los vecinos más cercanos. Los parámetros con los que se obtuvieron los mejores resultados fueron  $k=13$  con la distancia coseno.

El algoritmo de árbol de decisión fue implementado con la función *c5.0* del paquete *c50*, con los parámetros que utiliza el modelo por defecto, *trials=1*. Para mejorar el rendimiento del modelo se modificó el número de *trials=27*.

Para implementar la técnica de máquina de vectores de soporte (SVM), se utilizó el algoritmo *ksvm* del paquete *kernlab*, en un inicio con el kernel *rbfdot* y después para mejorar el modelo se evaluó todas las opciones de kernel disponibles. El mejor resultado se obtuvo con el kernel *vanilladot*.

Finalmente, para la construcción del modelo random forest se utilizó el algoritmo *randomForest* del paquete *randomForest*, inicialmente con los parámetros que aplica por defecto y para mejorarlo se modificó el parámetro *n tree=100*.

#### 4.1.3 Evaluar los modelos desarrollados

Con el objetivo de analizar el rendimiento del modelo frente a nuevos y futuros datos ingresados, es necesario evaluarlo. Para esto, se utilizó la base de datos “test” o de evaluación, ya que son datos cuyo diagnóstico se conoce y nos ayuda a comprobar la precisión de los diagnósticos devueltos por los modelos. Para esto se utilizó la función *predict* del paquete *stats*, ingresando como parámetros el nombre del modelo y los datos de evaluación. Luego de esto, los diagnósticos predichos son comparados con los diagnósticos reales. Se desarrolló una matriz de confusión para analizar varias métricas relacionadas al modelo, con la ayuda de la función *confusionMatrix* del paquete *caret*.

#### 4.2 Desarrollo de la aplicación web en Shiny

Para desarrollar la aplicación web se implementó el modelo SVM con el kernel *vanilladot*, la selección de este modelo se detalla en el capítulo de resultados. Se guardó el modelo seleccionado y los datos de entrenamiento balanceados utilizando la función *save* en el archivo “data.RData”, este fichero se guardó en la carpeta que contiene los elementos de la aplicación.

Para el desarrollo de la aplicación web, la sección *ui* contiene la distribución y diseño de todo el contenido de la interfaz; una tabla con los campos donde se ingresarán las nuevas medidas biomecánicas y el botón de ejecución con la palabra “Diagnóstico”, un campo que describe brevemente las patologías tratadas en este trabajo, seguido por una sección en la que se visualizará el diagnóstico devuelto por la aplicación. Por otro lado, la sección *server* contiene instrucciones para el correcto funcionamiento de la aplicación. Primero, se desarrolló instrucciones para validar la información ingresada por el usuario, para esto se tomó en cuenta los valores mínimos y máximos de la base de datos de referencia. Se implementó la función *sv\_between(valor mínimo, valor máximo)* para todos los parámetros, junto con el algoritmo *sv\_required(message=“valor requerido”)* para impedir que la aplicación corra sin alguno de los valores solicitados. Adicionalmente, se configuró el mensaje de alerta: “Por favor ingrese medidas biomecánicas válidas” en caso de que se presione el botón de acción

sin ingresar correctamente los valores solicitados. Segundo, la sección *server* contiene instrucciones para trabajar con los datos ingresados; al presionar el botón de acción “Diagnóstico”, se escalan los datos ingresados utilizando las medias y desviaciones estándar de los datos de entrenamiento, luego estos datos se aplican al modelo predictivo para generar como respuesta el diagnóstico de la salud lumbar. Finalmente se implementó un token en el scrip para enlazar la aplicación desarrollada con el servidor de Shiny.

## 5. Resultados

### 5.1 Métricas de los modelos desarrollados

Para evaluar el rendimiento de los modelos se requiere analizar los tipos de predicciones correctas e incorrectas de los modelos desarrollados. Se tomó en cuenta la precisión, valor kappa, la sensibilidad y especificidad para predecir cada diagnóstico. La precisión se refiere a la fracción de predicciones correctas, se planteó como objetivo lograr el 75%. El coeficiente kappa “ajusta la precisión teniendo en cuenta la posibilidad de una predicción correcta solo por casualidad” [9], sus valores oscilan entre 0 y 1 indicando el nivel de concordancia entre las predicciones de los modelos y los valores reales. La sensibilidad se refiere a la proporción de observaciones positivas que se clasifican correctamente, también se lo conoce como tasa de verdaderos positivos. Por otro lado, la especificidad es la proporción de observaciones negativas que se clasifican correctamente, es decir, la tasa de verdaderos negativos [9]. Para obtener todas estas métricas se generó una matriz de confusión con la función *confusionMatrix* del paquete *caret*. A continuación se muestra un resumen de los resultados obtenidos:

Modelo	Precisión	Kappa	Sensibilidad			Especificidad		
			Hernia	Normal	Espondilo.	Hernia	Normal	Espondilo.
KNN distancia=euclídea	0,79	0,67	0,81	0,47	0,89	0,87	0,89	0,94
KNN mejorado distancia=euclídea	0,80	0,68	0,81	0,68	0,89	0,89	0,89	0,94
KNN dist=manhattan	0,78	0,65	0,81	0,59	0,91	0,89	0,89	0,94
KNN mejorado dist=manhattan	0,80	0,69	0,88	0,65	0,89	0,89	0,89	0,94
KNN dist=sim cos	0,82	0,72	0,94	0,62	0,93	0,87	0,93	0,94
KNN mejorado dist=sim cos	0,81	0,70	0,94	0,59	0,93	0,86	0,93	0,94
Árbol de decisión	0,82	0,72	0,88	0,68	0,91	0,86	0,90	1,00
Árbol de decisión mejorado	0,82	0,72	0,69	0,71	0,96	0,87	0,88	1,00
SVM Kernel="rbfdot"	0,82	0,72	0,75	0,71	0,93	0,89	0,90	0,96
SVM mejorado	0,85	0,77	0,75	0,76	0,96	0,89	0,92	1,00

Kernel= "vanilladot"								
Rain forest	0,82	0,72	0,69	0,74	0,93	0,89	0,87	1,00
Rain Forest mejorado	0,83	0,73	0,75	0,74	0,93	0,89	0,89	1,00

Tabla 4.- Resumen de las métricas obtenidas.

Analizando los resultados obtenidos se determinó que el mejor modelo para desarrollar la aplicación es el modelo desarrollado con el algoritmo de máquina de vectores de soporte, SVM por sus siglas en inglés, con el kernel *vanilladot*. Se seleccionó este modelo debido a que todos los estadísticos analizados superan el 0.75, valor establecido como objetivo mínimo de precisión del modelo.

## 5.2 Aplicación en Shiny

Como se mencionó anteriormente, el producto final de este trabajo es una aplicación web que puede predecir con un 85% de precisión, tres diagnósticos de la salud lumbar; la condición "normal" y dos patologías "hernia" y "espondilolistesis". La aplicación contiene una tabla donde el usuario ingresa las seis medidas biomecánicas solicitadas y si los valores ingresados son correctos, al dar click en el botón "Diagnóstico" la aplicación devuelve el diagnóstico generado por el modelo entrenado con el algoritmo de máquina de vectores de soporte (SVM).

La aplicación web desarrollada se encuentra activa en el servidor de Shiny, el enlace para acceder es:

[https://damaris24.shinyapps.io/Predictor\\_de\\_salud\\_lumbar/](https://damaris24.shinyapps.io/Predictor_de_salud_lumbar/)

La siguiente imagen muestra el diseño de la interfaz cuando se ingresan valores incorrectos, se activa el mensaje de alerta para que el usuario pueda corregir los datos.



Figura 12.- Imagen de la interfaz con valores erróneos.

Una vez que se ingresen los valores de forma correcta y se presione el botón "Diagnóstico", la aplicación devuelve el diagnóstico predicho por el modelo SVM. En este caso, se muestra un ejemplo donde las medidas ingresadas corresponden al diagnóstico de Hernia con un 85% de precisión.

**PREDICTOR DE LA SALUD LUMBAR**

Incidencia pélvica  
24

Inclinación pélvica  
27

Ángulo de lordosis lumbar  
17

Pendiente sacra  
10

Radio pélvico  
100

Grado espondilolistesis  
34

**Diagnóstico**

La región lumbar de la columna vertebral está compuesta por vértebras más grandes y gruesas, debido a que en esta zona se deposita la mayor parte del peso del cuerpo, lo que hace que sea más propensa a la aparición de dolores y afecciones. El dolor en la región lumbar es una de las principales causas de consulta médica y se presenta con frecuencia en el 80-90% de la población adulta. La hernia de disco y la espondilolistesis son patologías degenerativas que afectan generalmente a la columna lumbar, siendo algunas de las posibles causas del dolor esta zona.

Esta aplicación ha sido desarrollada con el objetivo de optimizar el diagnóstico tanto de hernia de disco como de espondilolistesis, interpretando rápidamente las características biomecánicas de los pacientes y orientando a los médicos para aplicar un tratamiento oportuno. Se desarrolló a partir de la base de datos "column\_3C\_weka" disponible en UCI Machine Learning Repository, entrenando un modelo predictivo con el algoritmo de aprendizaje automático SVM (máquina de soporte de vectores) y alcanza una precisión del 85%.

**Diagnóstico:**  
**Hernia**

Normal      Hernia de disco      Espondilolistesis

La hernia de disco consiste en la salida del material gelatinoso contenido en el disco intervertebral hacia el canal raquídeo. Mientras que la espondilolistesis es el desplazamiento de una vértebra, por lo general hacia adelante, de la vértebra que se encuentra debajo. Ambas patologías generan

Figura 13.- Imagen de la interfaz con el diagnóstico predicho.

## 6. Discusión

El modelo seleccionado para desarrollar la aplicación fue el que implementa el algoritmo máquina de vectores de soporte (SVM). La precisión alcanzada sugiere que en un 85% de los casos, la aplicación devuelve un diagnóstico acertado. Según la literatura, esta medida de precisión corresponde a un modelo muy bueno ya que se encuentra entre 80%-90% de precisión [9]. Adicionalmente el valor kappa fortalece la validez del modelo, demostrando un grado de concordancia de 0.77 considerado como "bueno" [9]. La sensibilidad del modelo SVM para clasificar pacientes con hernia discal, sanos y con espondilolistesis, es de 0.75, 0.76 y 0.96 respectivamente. A pesar de que otros modelos tienen mayor sensibilidad para diagnosticar hernia discal como el modelo kNN con la distancia similitud coseno con un valor de 0.94, la sensibilidad de este modelo frente a pacientes sanos es de 0.62, generando mayor incertidumbre en este tipo de diagnóstico. Debido a esto se seleccionó el modelo que en todas sus métricas

supera el nivel de precisión mínima establecida como objetivo del trabajo que fue 0.75. En cuanto a la especificidad o tasa de verdaderos negativos, todos los modelos alcanzaron valores entre 0.87 y 1. Para esta métrica igualmente el modelo SVM tiene los valores más altos; 0.89, 0.92 y 1 para diagnóstico de hernia, normal y espondilolistesis respectivamente.

Una de las tareas planteadas para el desarrollo de la aplicación web fue validar su correcto funcionamiento una vez subida al servidor de Shiny. En este paso se evaluó todas las respuestas de la aplicación frente a varios datos ingresados, lo que ayudó a perfeccionar su funcionamiento estableciendo reglas para la validación de los datos ingresados. Todo esto con el fin de garantizar el beneficio que implica la aplicación de este tipo de técnicas en el campo de la salud, respetando la importancia de la calidad de los datos ya que de esto depende el tratamiento oportuno de los pacientes.

## 7. Conclusiones

### 7.1 Conclusiones

- Se logró cumplir los objetivos del trabajo; se desarrolló una aplicación web interactiva que permite predecir dos patologías de la salud lumbar, con una precisión del 85%, a partir de características biomecánicas de la columna.
- Tomando en cuenta que esta aplicación fue desarrollada como una herramienta para facilitar la toma de decisiones en cuanto al diagnóstico médico de hernia de disco y espondilolistesis, se puede decir que esta necesidad ha sido cubierta con el nivel de precisión alcanzado.
- Aplicar técnicas de aprendizaje automático en el campo de la salud permite convertir datos clínicos desde mediciones hasta imágenes en conclusiones importantes para la toma de decisiones sobre el diagnóstico de enfermedades.
- El paquete de Shiny permite desarrollar aplicaciones web de manera sencilla e intuitiva para quienes estamos relacionados con el manejo de R. Permite diseñar y dar estilo a una interfaz para el manejo de algoritmos como en este caso, el modelo predictivo, todo desde el entorno de R.

### 7.2 Líneas de futuro

Para los siguientes pasos de este trabajo se puede considerar, en medida de lo posible, con la ayuda de un especialista ampliar el uso de estas medidas biomecánicas y vincularlas a otro tipo de enfermedades que afectan a la salud lumbar, así como acceder a más datos sobre este tipo de enfermedades que puedan enriquecer el funcionamiento de la aplicación desarrollada.

### 7.3 Seguimiento de la planificación

La planificación establecida al inicio del trabajo se cumplió en su mayoría. Surgieron tareas no consideradas como el desarrollo de un algoritmo desde cero para aplicar la técnica knn con otro tipo de distancias como la de manhattan y similitud coseno ya que la versión utilizada de R no permitió trabajar con paquetes antiguos donde se puede modificar este parámetro. Sin embargo, se pudo considerar el tiempo invertido en esta tarea como parte de la tarea de mejoramiento y evaluación de los modelos mejorados.

Otra tarea con la que no se contaba, es la validación de datos ingresados en la aplicación. Al momento de tener desarrollada la base de la interfaz, se vio la necesidad de optimizar su funcionamiento. Por otro lado, establecer el estilo de la interfaz y analizar las opciones de diseño y herramientas interactivas de la aplicación es otra tarea no considerada que tomó bastante tiempo. Sin embargo, ninguna de las anteriores tareas adicionales comprometió el apropiado desarrollo del trabajo ni el tiempo para cumplirlo.

## 8 Glosario

**Espondilolistesis.** - Enfermedad que afecta generalmente a la columna lumbar, en la que una vértebra se desplaza por lo general hacia adelante, de la vértebra que se encuentra debajo [5].

**Hernia de disco.** - Enfermedad que afecta a la columna, consiste en la salida del material gelatinoso contenido en el disco intervertebral hacia el canal raquídeo [4].

**Características biomecánicas.** - Las características biomecánicas derivadas de la forma y orientación de la columna lumbar y de la pelvis, describen la relación de la columna con respecto al piso y con respecto a la pelvis [18].

**Aprendizaje automático.** - Es un enfoque de la Inteligencia Artificial que crea sistemas que aprendan patrones complejos en cantidades gigantes de datos [19].

**Shiny.** - Es un paquete de R que facilita construir aplicaciones web interactivas directamente desde R [20].

## 9 Bibliografía

1. Sierra, Isaac; Rincón, Laura, Dávila Claudia (Mar. 2018). Anatomía de la columna vertebral en radiografía convencional. Revista médica Sanitas. Recuperado el 17 de septiembre de 2021 de <https://revistas.unisanitas.edu.co/index.php/RMS/article/view/312>
2. Ortiz, Jessie (Jun. 2016). Anatomía de la columna vertebral. Revista Mexicana de Anestesiología. Recuperado el 17 de septiembre de <https://www.medigraphic.com/pdfs/rma/cma-2016/cmas161bh.pdf>
3. Vargas, Karen (Sep. 2012). Lumbalgias. Costa Rica: *SciELO*. Recuperado el 18 de septiembre de 2021 de [https://www.scielo.sa.cr/scielo.php?script=sci\\_arttext&pid=S1409-00152012000200011](https://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1409-00152012000200011)
4. García de Sola, Rafael (Abr. 2014). Hernia discal lumbar: Patología degenerativa de columna vertebral. Unidad de Neurocirugía RGS. Recuperado el 17 de septiembre de 2021 de [https://neurorgs.net/docencia/pregraduados/hernia-discal-lumbar-patologia-degenerativa-de-columna-vertebral/#Hernia\\_discal\\_lumbar-Concepto](https://neurorgs.net/docencia/pregraduados/hernia-discal-lumbar-patologia-degenerativa-de-columna-vertebral/#Hernia_discal_lumbar-Concepto)
5. Garza, Félix; Cuervas, Lagarda; Bernáldez, Ladewing (2014). Espondilolistesis lumbar degenerativa de un solo segmento tratada con espaciador interespinoso. Ciudad de México: *SciELO*. Recuperado el 18 de septiembre de 2021 de [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S2306-41022014000200004](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2306-41022014000200004)
6. Vega, Carlos (Dic. 2019). Parámetros espinopélvicos como factor predictivo de lumbalgia en pacientes con coxartrosis degenerativa en la Unidad Médica de Alta Especialidad Hospital de Traumatología y Ortopedia de Puebla durante enero 2017 a diciembre 2019. Puebla: *Benemérita Universidad Autónoma de Puebla*. Recuperado el 18 de septiembre de 2021 de <https://repositorioinstitucional.buap.mx/bitstream/handle/20.500.12371/1/1264/20200813171315-5518-T.pdf?sequence=1>
7. Gómez, Carlos (2019). Desarrollo de Soluciones software mediante aprendizaje automático en el ámbito de la salud: situación tecnológica y perspectivas. Universidad Politécnica de Cartagena. Recuperado el 02 de octubre del 2021 de <https://repositorio.upct.es/handle/10317/8110>
8. Da Mota, Henrique (Ago. 2011). Vertebral Column Data Set. Brazil: *UCI Machine Learning Repository*. Recuperado el 16 de septiembre de 2021 de: <http://archive.ics.uci.edu/ml/datasets/vertebral+column>

9. Lantz, Brett (Jul. 2015). Machine Learning with R. Birmingham: *PACKT Publishing*.
10. Mahmood, Ibrahim (Mar. 2021). The role of Machine Learning Algorithms for Diagnosing Diseases. *JASTT*. Recuperado el 16 de diciembre de 2021 de <https://jastt.org/index.php/jasttpath/article/view/79>.
11. Smiti, Abir (2020). When machine learning meets medical world: Current status and future challenges. *ELSEVIER*. Recuperado el 18 de diciembre de 2021 de <https://www.sciencedirect.com/science/article/abs/pii/S157401372030126X?via%3Dihub>
12. Shaheen, Mohammed (2021). Adoption of Machine Learning for medical diagnosis. *ScienceOpen Preprints*. Recuperado el 19 de diciembre de 2021 de <https://www.scienceopen.com/hosted-document?doi=10.14293/S2199-1006.1.SOR-.PPHMKA6.v1>
13. Battineni, Gopi; Gamo, Getu (2020). Applications of Machine Learning Predictive Models in Chronic Disease Diagnosis. *MDPI*. Recuperado el 19 de diciembre de 2021 de <https://www.mdpi.com/2075-4426/10/2/21>
14. Aroche, Yoandra; Pons, Laura (Mar. 2015). Patogenia, cuadro clínico y diagnóstico imagenológico por resonancia magnética de las hernias discales. *SCIELO*. Recuperado el 20 de diciembre de 2021 de [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30192015000300012](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192015000300012)
15. Reshi, Aijaz; Ashraf, Imran (Jul. 2021). Diagnosis of vertebral column pathologies using concatenated resampling with machine learning algorithms. *NCBI*. Recuperado el 18 de diciembre de 2021 de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8323723/>
16. Barrios, Juan Ignacio (2019). La matriz de confusión y sus métricas. *Health Big Data*. Recuperado el 20 de diciembre de 2021 de <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>
17. Abraira, V (May, 2001). El índice kappa. Unidad de Bioestadística Clínica del Hospital Ramón y Cajal. Recuperado el 18 de diciembre de 2021 de [file:///C:/Users/damar\\_000/Downloads/S113835930173955X%20\(1\).pdf](file:///C:/Users/damar_000/Downloads/S113835930173955X%20(1).pdf)
18. Guidoy, Alfredo; Gagliardi, Martín; Sícoli, Alfredo; González Nicolás (2018). Parámetros sagitales espino-pélvicos en una población asintomática Argentina. *Surgical Neurology International*. Recuperado el 19 de septiembre de 2021 de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5799942/>

19. Gonzales, Andrés (s/f). ¿Qué es Machine Learning?. *Cleverdata*. Recuperado el 20 de diciembre de 2021 de <https://cleverdata.io/que-es-machine-learning-big-data/>
20. Shiny from RStudio. Description. Recuperado el 20 de diciembre de 2021 de <https://shiny.rstudio.com/>
21. Balthazard, P (Nov. 2015). Fundamentos de biomecánica. ELSEVIER. Recuperado el 21 de septiembre de 2021 de <https://www.sciencedirect.com/science/article/abs/pii/S1293296515741423>
22. Berthonnaud, Eric; Dimmet, Joannés; Roussouly, Pierre; Lebel, Hubert (2005). Analysis of the sagittal balance of the spine and pelvis using shape and orientation parameters. PubMed. Recuperado el 21 de septiembre de 2021 de <https://pubmed.ncbi.nlm.nih.gov/15687851/>
23. Tejeda, Martín (2015). Balance sagital en columna vertebral. Su significado e importancia. México: Medigraphic. Recuperado el 20 de septiembre de 2021 de <https://www.medigraphic.com/pdfs/orthotips/ot-2015/ot153b.pdf>
24. Labelle, Hubert (2005). The importance of spino-pelvic balance in L5-s1 developmental spondylolisthesis: a review of pertinent radiologic measurements. PubMed. Recuperado el 20 de septiembre de 2021 de <https://pubmed.ncbi.nlm.nih.gov/15767882/>
25. Villae, Raphael; Levassor, Nicolas; Rillardon, Ludovic; Templier, Alexandre (2005). Radiographic analysis of the sagittal alignment and balance of the spine in asymptomatic subjects. PubMed. Recuperado el 20 de septiembre de 2021 de <https://pubmed.ncbi.nlm.nih.gov/15687145/>
26. ICAC (Mar. 2021). Columna vertebral: enfermedades principales. *Instituto Cirugía Avanzada de Columna*. Recuperado el 20 de diciembre de 2021 de <https://columna.com/columna-vertebral-enfermedades-principales/>.
27. Shiny from RStudio. Recuperado el 21 de diciembre de 2021 de <https://shiny.rstudio.com/>
28. Solymos, Peter. The Anatomy of a Shiny Application. *Hosting*. Recuperado el 21 de diciembre de 2021 de <https://hosting.analythium.io/the-anatomy-of-a-shiny-application/>
29. ICHI.PRO. K-Vecinos más cercanos. Recuperado el 22 de diciembre de 2021 de <https://ichi.pro/es/k-vecinos-mas-cercanos-knn-157240814281359>

30. Ollé, Jordi. ¿Qué es una máquina de vectores de soporte (Support Vector Machine) y cómo interpretarla? Referencias e intuición práctica. *Conceptos claros*. Recuperado el 21 de diciembre de 2021 de <https://conceptosclaros.com/que-es-maquina-vectores-soporte/>
31. Random Forest Algorithm. JavaTpoint. Recuperado el 22 de diciembre de 2021 de <https://www.javatpoint.com/machine-learning-random-forest-algorithm>