

**APLICACIÓN WEB PARA LA PREDICCIÓN DEL TRATAMIENTO DE
PACIENTES CON APNEA OBSTRUCTIVA DEL SUEÑO USANDO MACHINE
LEARNING**

**LUIS GARCIA SANDOVAL
MANUEL LEMUS PERTUZ**

**UNIVERSIDAD AUTÓNOMA DEL CARIBE
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA MECATRÓNICA
BARRANQUILLA - COLOMBIA
2021**

**APLICACIÓN WEB PARA LA PREDICCIÓN DEL TRATAMIENTO DE
PACIENTES CON APNEA OBSTRUCTIVA DEL SUEÑO USANDO MACHINE
LEARNING**

**LUIS GARCIA SANDOVAL
MANUEL LEMUS PERTUZ**

**Trabajo de grado presentado para optar al título de
Ingeniero Mecatrónico**

**ASESORES DISCIPLINARES:
ING. JAIR VILLANUEVA, MSc.
ING. CARLOS DÍAZ, PhD.**

**UNIVERSIDAD AUTÓNOMA DEL CARIBE
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA MECATRÓNICA
BARRANQUILLA - COLOMBIA**

2021

Nota de aceptación

Firma del jurado 1

Firma del jurado 2

DEDICATORIA

Este proyecto está dedicado, primeramente, a Dios, por su infinita bondad y por guiarme en cada uno de mis pasos.

A mis padres y toda mi familia por su comprensión y estímulo constante, además de su apoyo incondicional a lo largo de mis estudios.

Y a todas las personas que, de una u otra forma, ayudaron a la realización de este trabajo.

– **Manuel Esteban Lemus Pertuz.**

Dedico en primer lugar a mis amigos por apoyarme, darme apoyo, fuerzas y creer en mí.

A mis profesores que me guiaron en cada una de mis asignaturas de mi carrera universitaria.

Y a mi familia por estar conmigo durante este proceso.

– **Luis Guillermo García Sandoval.**

TABLA DE CONTENIDO

LISTA DE FIGURAS	7
LISTA DE TABLAS.....	8
GLOSARIO.....	9
RESUMEN	10
ABSTRACT	11
INTRODUCCIÓN	12
1. PLANTEAMIENTO DEL PROBLEMA.....	13
1.1. ANTECEDENTES	13
1.1. FORMULACIÓN DEL PROBLEMA	13
1.2. JUSTIFICACIÓN Y ALCANCE	14
2. OBJETIVOS.....	16
2.1. OBJETIVO GENERAL.....	16
2.2. OBJETIVOS ESPECÍFICOS.....	16
3. MARCO DE REFERENCIA.....	17
3.1. ESTADO DEL ARTE.....	17
3.2. MARCO TEÓRICO	29
3.2.1 Apnea del sueño.....	29
3.2.2 Tipos de apnea del sueño	29
3.2.3 Apnea central del sueño	30
3.2.4 Apnea obstructiva del sueño	30
3.2.5 Diagnostico.....	31
3.2.6 Estrategias para el Tratamiento de OSA.....	33
3.2.7 Machine Learning	38
3.2.8 Feature selection	38
3.2.9 Algoritmos de aprendizaje	42
3.2.10 Aprendizaje supervisado	42
3.2.11 Aprendizaje no supervisado	43
3.2.12 Aplicación web.....	45
4. PROCEDIMIENTO METODOLÓGICO	47
4.1. METODLOGÍA.....	47
4.1.1. Conjunto de datos.....	47
4.1.2. Preprocesamiento de la base de datos	48
4.1.3. Feature Selection.....	53
4.1.4. Split de los datos	54

4.1.5.	Implementación de modelos supervisados y no supervisados	54
4.1.6.	Validación del modelo.....	55
4.1.7.	Métricas	55
4.1.8.	Aplicación web.....	56
4.2.	TIPO DE ESTUDIO.....	57
4.3.	CRONOGRAMA	58
5.	PRESUPUESTO.....	60
5.1.	PRESUPUESTO GENERAL.....	60
5.2.	PERSONAL CIENTÍFICO Y DE APOYO.....	61
5.3.	CONSULTORIA ESPECIALIZADA.....	61
5.4.	MATERIALES, INSUMOS Y EQUIPOS.....	61
5.5.	DIFUSIÓN DE RESULTADOS	62
6.	PRESENTACIÓN Y ANÁLISIS DE RESULTADOS.....	63
6.1.	DISEÑO DE LA APLICACIÓN WEB.....	63
6.2.	RECOLECCIÓN DE DATOS	68
6.2.1.	Muestra.....	68
6.3.	ANÁLISIS DE RESULTADOS	68
6.3.1.	Preprocesamiento.....	69
6.3.2.	Feature Selection.....	69
6.3.3.	Modelos supervisados	70
6.3.4.	Características indelebles.....	74
6.3.5.	Modelo no supervisado.....	75
6.3.6.	Pruebas realizadas en la aplicación web	80
6.2.	MANUAL DE USUARIO.....	¡Error! Marcador no definido.
	CONCLUSIONES Y RECOMENDACIONES	84
	BIBLIOGRAFÍA	86
	ANEXOS	¡Error! Marcador no definido.

LISTA DE FIGURAS

Figura 1. Apnea Obstructiva del Sueño (Mayo Clinic, 2019).	31
Figura 2. Polisomnografía (Clínica Aystesis, 2017).	32
Figura 3. CPAP (Clínica de cirugía maxilofacial, 2019).	34
Figura 4. Dispositivo de avance Mandibular (Mayo Clinic, 2019).	35
Figura 5. Dispositivo de retención de Lengua (Denlattix, 2019).	35
Figura 6. Feature Selection (Towards Data Science, 2021).	38
Figura 7. Método Forward Stepwise (thaddeussegura, 2020).	39
Figura 8. Método Backward Stepwise (thaddeussegura, 2020).	40
Figura 9. Método LASSO (Friedman, 2019).	41
Figura 10. Método Random Forest (shishirkant, 2021).	42
Figura 11. Representación de aprendizaje supervisado (ICHI.PRO, 2020).	43
Figura 12. Representación de aprendizaje no supervisado (ICHI.PRO, 2020).	44
Figura 13. Agrupación (PyPRr, 2010).	44
Figura 14. Metodología propuesta.	47
Figura 15. Porcentaje de datos nulos en los predictores del dataset PSG.	50
Figura 16. Porcentaje de datos nulos en los predictores del dataset HRP.	51
Figura 17. Dataframes con outcome desbalanceados.	52
Figura 18. Dataframes con outcome balanceados.	52
Figura 19. Resumen del preprocesamiento de la base de datos.	53
Figura 20. Curva ROC (Eric Melillanca, 2018).	56
Figura 21. Diseño de la aplicación web.	57
Figura 22. Visual Studio Code - Backend.	63
Figura 23. Archivos y carpetas Backend.	64
Figura 24. Script del modelo de predicción implementado en visual studio code.	65
Figura 25. Script de la predicción para el usuario evaluado.	66
Figura 27. Archivos y carpetas del frontend.	67
Figura 28. Paginas principales de la aplicación web.	68
Figura 29. Curva ROC comparando los modelos (Forware - Backward) (PSG).	72
Figura 30. Curva ROC comparando los modelos (LASSO – Random Forest) (PSG).	73
Figura 31. Curva ROC comparando los modelos (Forware - Backward) (HRP).	73
Figura 32. Curva ROC comparando los modelos (LASSO – Random Forest) (HRP).	74
Figura 33. Número de clústeres – Método de la rodilla.	76
Figura 34. Número de clústeres – Método de la silueta.	76
Figura 35. Número de clústeres – 30 métodos.	77
Figura 36. K-means.	79
Figura 37. Dominio de TREAT-OSA.	80
Figura 38. Página de inicio de sesión.	81
Figura 39. Sesiones de la aplicación web.	82
Figura 40. Paciente apto para CPAP.	83
Figura 41. Paciente no apto para CPAP.	83

LISTA DE TABLAS

Tabla 1. Estado del arte.....	17
Tabla 2. Cronograma objetivo específico 1.....	58
Tabla 3. Cronograma objetivo específico 2.....	58
Tabla 4. Cronograma objetivo específico 3.....	59
Tabla 5. Presupuesto general.....	60
Tabla 6. Costo personal científico.....	61
Tabla 7. Costo personal de apoyo.....	61
Tabla 8. Costo consultoría especializada.....	61
Tabla 9. Costo equipos usados.....	62
Tabla 10. Costo material de difusión.....	62
Tabla 11. Tamaño de los datasets (Supervisado).....	69
Tabla 12. Tamaño de los datasets (No supervisado).....	69
Tabla 13. Características importantes (PSG).....	70
Tabla 14. Características importantes (HRP).....	70
Tabla 15. Modelos Supervisados dataframe PSG.....	71
Tabla 16. Modelos Supervisados dataframe HRP.....	71
Tabla 17. Predictores finales.....	75
Tabla 18. Validación de modelos en predictores finales.....	75
Tabla 19. Variables importantes - Clustering (PSG).....	77
Tabla 20. Variables importantes - Clustering (HRP).....	78
Tabla 21. Promedio de las variables importantes (PSG).....	79
Tabla 22. Promedio de las variables importantes (HRP).....	80

GLOSARIO

APNEA: interrupción completa del flujo nasobucal.

APRENDIZAJE AUTOMÁTICO: se refiere al proceso por el cual los PC desarrollan el reconocimiento de patrones o la capacidad de aprender continuamente y realizar predicciones basadas en datos, realizando ajustes sin haber sido programados específicamente para ello.

CPAP: siglas inglesas para denominar un soporte ventilatorio mecánico que consiste en la aplicación de presión positiva en la vía aérea mediante máscara nasal. Es un tratamiento de elección en el síndrome de apneas obstructivas del sueño.

PREDICTORES: dicho de algo o alguien que predice o ayuda a predecir.

RESUMEN

Actualmente, con las nuevas implementaciones del uso de bases de datos y el aprendizaje automático en la medicina, todo tiene tendencia a mejorar, optimizar niveles de calidad para el paciente, mayor seguridad en los procesos y proporcionar mejores procesamientos de datos, lo que se ve reflejado en la toma de decisiones. El objetivo de este estudio es diseñar una aplicación web capaz de predecir la adherencia a la CPAP de pacientes que sufren de apnea obstructiva del sueño (AOS) usando técnicas del aprendizaje automático. Para conseguir esto, se realiza el preprocesamiento de la base de datos, la cual está conformada por las variables clínicas de 430 pacientes, consiguiendo así un conjunto de datos limpio y libre de ruido. Se hace uso de métodos de feature selection (Forward, Backward, LASSO, Random Forest), para así obtener las características más relevantes a las cuales se le implementan los modelos predictivos. Finalmente se diseña una aplicación web para a predicción, teniendo en cuenta las validaciones de los modelos, de esta manera se obtiene una herramienta digital de fácil acceso y con una alta precisión como apoyo en la toma de decisiones clínicas.

Palabras claves: Aplicación web, aprendizaje automático, modelos predictivos predicción, predictores.

ABSTRACT

Currently, with the new implementations of databases and machine learning in medicine, everything is tending to improve, optimizing quality levels for the patient, greater safety in the processes and providing better data processing, which is reflected in decision making. This study aims to design a web application capable of predicting CPAP adherence in patients suffering from obstructive sleep apnea (OSA) using machine learning techniques. To this end, the database is preprocessed, which consists of the clinical variables of 430 subjects, thus obtaining a clean and noise-free dataset. Feature selection methods such as forward, backward, LASSO and Random Forest are used in this study. These methods were applied to select the most relevant predictors to implement the predictive models. Finally, a web application was designed for prediction, considering the validations of the models, thus obtaining an easily accessible and highly accurate digital tool to support clinical decision-making.

Keyword: Machine learning, obstructive sleep apnea, prediction, predictive models, web application.

INTRODUCCIÓN

La apnea obstructiva del sueño (AOS) es un síndrome que se caracteriza por el estrechamiento cíclico o el cierre completo de las vías respiratorias superiores durante el sueño, lo que causa una disminución del flujo de aire y con frecuencia termina en la disminución del nivel de saturación de oxígeno [1][2].

El síndrome de apnea obstructiva del sueño (SAOS) es un trastorno muy común con una influencia aproximada del 5 al 14 por ciento en adultos con un rango de edad de 30 a 70 años [3]. En las últimas décadas su prevalencia ha aumentado sustancialmente, principalmente debido a la obesidad y el envejecimiento, dos factores principales de riesgo de AOS [4][5]. Estudios epidemiológicos recientes reportaron aumento de AOS entre la población general, así como una prevalencia mayor en grupos con un riesgo particularmente alto de consecuencias adversas, como aquellos que padecen de hipertensión, enfermedades cardiovasculares, diabetes o sujetos evaluados para la cirugía bariátrica [6].

1. PLANTEAMIENTO DEL PROBLEMA

1.1. ANTECEDENTES

Desde hace unos años la apnea obstructiva del sueño (AOS) al ser el trastorno respiratorio más común referente con el sueño ha despertado un interés entre los investigadores [7][8][9]. Este mismo trastorno últimamente se ha visto relacionado o en ocasiones utilizado como sinónimo de la hipertensión, la obesidad, la dislipidemia, entre otros [10][11][12].

Los métodos usados en el campo de la salud que son capaces de detectar este trastorno son tanto especiales como insuficientes en comparación a la cifra de pacientes en aumento parecientes de apnea obstructiva del sueño (AOS) [13]. Por consecuencia las bases de datos de pacientes con viable apnea obstructiva del sueño (AOS) son importantes para lograr un nuevo método más eficiente tomando de referencia las diversas variables y aspectos de cada paciente [14].

1.1. FORMULACIÓN DEL PROBLEMA

El diagnóstico del SAOS se realiza con la polisomnografía (PSG). La PSG es el método estándar de oro, que consiste en una prueba del sueño que se realiza en un laboratorio de sueño durante una noche y se registran las ondas cerebrales, los niveles de oxígeno en la sangre, frecuencia cardiaca y respiratoria, entre otras [15].

Aunque la PSG es un procedimiento estándar para la detección de trastornos de sueño, presenta varias limitaciones, entre las cuales se mencionan incomodidad del paciente, larga lista de espera, necesidad de personal médico especialista para la interpretación del estudio. Por esta razón, surge la necesidad de nuevos métodos alternativos para la detección y sobre todo el tratamiento de personas con AOS [16].

La presión positiva continua en las vías respiratorias (CPAP), es el tratamiento más utilizado en pacientes con un nivel moderado a severo de AOS. Es un dispositivo que bombea aire a presión en las vías respiratorias, evitando así el estrechamiento de las mismas [17]. Algunos pacientes a los cuales se les aplica este tratamiento, con el tiempo resultan no ser aptos, perdiendo así tiempo, dinero y la dedicación empleada en la CPAP. Es aquí donde técnicas y algoritmos predictivos son necesarios para facilitar estos análisis y resultados.

El aprendizaje automático es una rama de la inteligencia artificial y hace referencia al conjunto de técnicas computacionales con capacidad de aprender y ayudar en la solución de tareas complejas. Haciendo referencia a la medicina, el aprendizaje automático tiene como objetivo proporcionar mejor procesamiento y resultados a las herramientas informáticas existentes [18].

Cabe resaltar que, los métodos en donde se aplican algoritmos de inteligencia artificial pueden resolver todos aquellos problemas que presentan el proceso estándar de oro PSG. De esta manera, se propone la realización de un algoritmo y esquema completo para abordar ciertos problemas en la detección, pero mucho más importante en la selección del tratamiento que vaya acorde con el caso específico de cada paciente, utilizando aprendizaje automático [16].

1.2. JUSTIFICACIÓN Y ALCANCE

Este estudio contribuye a mejorar la toma de decisiones relacionado con el tratamiento de la OSA dado por el médico usando aprendizaje automático, tomando la aplicabilidad clínica para diferentes métodos de aprendizaje supervisado y no supervisado basados en información del paciente.

Se propone este proyecto para mejorar la elección del tratamiento oportuno y personalizado al paciente, identificando las variables clínicas claves para que el

medico pueda dictar sobre el paciente un tratamiento, evitando que el paciente gaste tiempo y dinero haciéndose valoraciones como la polisomnografía, de esta forma se evaden largas listas de espera, tratamientos costosos y malas tomas de elecciones [3][19][20].

El lenguaje de programación para el manejo, manipulación, modelado y validación es R (v4.1.2, 2021, Boston), utilizando como entorno de desarrollo integrado (IDE) a RStudio (v1.3.1073, 2020, Boston), que son herramientas open Source las cuales están disponibles en <https://cran.r-project.org/bin/windows/base/> y <https://www.rstudio.com/products/rstudio/download/>. Para el diseño de la aplicación web los lenguajes de programación a utilizar son Python, JavaScript, TypeScript, CSS y SCSS [21].

Este proyecto ayuda a mejorar la respuesta del tratamiento de los pacientes dándoles así una opción más rápida para poder decidir si iniciar o no el tratamiento [22][23]. En este orden de ideas, el interés de los modelos predictivos en la medicina tiende a crecer y mejorar, además cabe resaltar que la realización de una aplicación web del modelo predictivo beneficia no solo a la comunidad clínica, sino también a todo aquel que practique el uso de machine learning [24][25].

El aprendizaje automático da a las computadoras la habilidad de aprender sobre algo para lo que no han sido explícitamente programadas. La aplicación del machine learning en la medicina revoluciona la forma de hacer ciencia, permite que grupos de investigaciones, grandes y pequeños, puedan tener acceso a las mismas oportunidades, que ya no necesitan sofisticados laboratorios para realizar los análisis o experimentos, debido a la facilidad con que los datos son obtenidos, generados, procesados y almacenados, mediante diferentes técnicas que pueden ser desarrolladas y ejecutadas en pequeñas computadoras de escritorio o grandes servidores [26][27].

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Desarrollar una aplicación web para la predicción del tratamiento de pacientes con apnea obstructiva de sueño (OSA) usando técnicas de aprendizaje automático supervisado y no supervisado como soporte en la toma de decisiones clínicas.

2.2. OBJETIVOS ESPECÍFICOS

- Desarrollar el preprocesamiento de un conjunto de datos que contienen las variables clínicas a utilizar para la implementación de los algoritmos de aprendizaje automático.
- Implementar modelos predictivos basados en aprendizaje automático supervisado y no supervisado para la detección del tratamiento de pacientes con AOS.
- Diseñar una aplicación web para la implementación del sistema de predicción para la toma de decisiones médicas.

3. MARCO DE REFERENCIA

3.1. ESTADO DEL ARTE

En la Tabla 1 se presenta el resumen de los estudios más significativos y relevantes encontrados, acorde al tema principal del proyecto de grado. Posterior a esto, se realiza una descripción más detallada a cada uno en particular.

Tabla 1. Estado del arte.

Autores y año	Título	Muestra	Métodos	Resultados	Conclusiones
Farahnaz Hajipour, et al, 2020. [2]	Regularized logistic regression for obstructive sleep apnea screening during wakefulness using daytime tracheal breathing sounds and anthropometric information.	199.	Regresión logística regularizada.	Predicción el nivel de gravedad de la AOS de 199 individuos. La precisión fue del 81,1% con una sensibilidad del 84,4% y especificidad del 77%.	El análisis de TBS se puede utilizar como una predicción rápida y confiable de la presencia y gravedad de la AOS durante la vigilia sin un estudio del sueño.
Corrado Mencar, et al, 2019. [3]	Application of machine learning to predict obstructive sleep apnea syndrome severity.	313.	Seleccionamos 19 variables que se utilizaron para un procesamiento adicional y para eventualmente entrenar 7 tipos de modelos de clasificación y 5 tipos de modelos de regresión para evaluar la capacidad de predicción.	La mejor precisión es del modelo support vector machine.se obtuvo una precisión del 44,7%, con la misma sensibilidad promedio. Los resultados de la regresión muestran un error cuadrático medio mínimo alcanzado de 22,17%.	Los resultados obtenidos con los datos disponibles sugieren el uso de métodos de aprendizaje automático como herramientas para proporcionar a los pacientes un nivel de prioridad para la prueba de polisomnografía, pero aún no pueden usarse para el diagnóstico automatizado.

<p>Daniel Álvarez, et al, 2020. [6]</p>	<p>A machine learning-based test for adult sleep apnoea screening at home using oximetry and airflow.</p>	<p>303 pacientes elegibles con sospecha de padecer AOS.</p>	<p>Máquinas vectoriales de soporte de regresión.</p>	<p>El modelo support vector machine para el rendimiento de doble canal obtuvo una precisión del 81,3%, para la oximetría individual del 75% y para el flujo de aire una precisión del 61,5%.</p>	<p>El estudio genera evidencia significativa de la superioridad de un enfoque de doble canal en el marco de la monitorización abreviada desatendida para la detección de AOS. Es más probable que las grabaciones de flujo de aire se vean afectadas por problemas de pérdida de señal permanente que la oximetría en entornos desatendidos.</p>
<p>Robert Stretch, et al, 2019. [4]</p>	<p>Predicting Nondiagnostic Home Sleep Apnea Tests Using Machine Learning.</p>	<p>Pacientes que deben ser derivados directamente a polisomnografía (PSG) por su alta probabilidad de HSAT no diagnóstica.</p>	<p>HSAT realizada como la prueba inicial para la sospecha de AOS, índice de eventos respiratorios (REI) ≥ 5 eventos / h. Las pruebas con REI <5 eventos / h o insuficiencia técnica se consideraron no diagnósticas.</p>	<p>Utilizando el modelo Random Forest se obtuvo una precisión del 81% con una sensibilidad del 46% y una especificidad del 95%.</p>	<p>En comparación con los modelos de regresión logística estándar, los modelos de aprendizaje automático mejoran la predicción de pacientes que requieren PSG en el laboratorio.</p>
<p>Berk Ustun, et al, 2016. [28]</p>	<p>Clinical Prediction Models for Sleep Apnea: The Importance of Medical History over Symptoms.</p>	<p>1.922 pacientes evaluados.</p>	<p>Se analiza la polisomnografía, se usa SLIM y 7 métodos de clasificación de última generación para producir modelos predictivos para la detección de la AOS utilizando características.</p>	<p>El método de aprendizaje automático SLIM creado, a partir de otros métodos como la regresión logística, árboles de decisión, máquina de vectores, entre otros, obtuvo una sensibilidad del 64,2% y especificidad 77%.</p>	<p>SLIM produce una herramienta clínica procesable que se puede aplicar a los datos que están habitualmente disponibles en los registros médicos electrónicos modernos, lo que puede facilitar la detección de AOS automatizada, en lugar de manual.</p>

<p>Gonzalo C, et al, 2019. [29]</p>	<p>Evaluation of Machine-Learning Approaches to Estimate Sleep Apnea Severity from At-Home Oximetry Recordings.</p>	<p>320 sujetos.</p>	<p>La señal de saturación de oxígeno en sangre (SpO2) transporta información útil sobre SAHS y se puede adquirir fácilmente mediante oximetría nocturna.</p>	<p>Los modelos entrenados (análisis discriminante lineal, regresión logística 1 contra todos, perceptrón multicapa bayesiano, y AdaBoost) obtuvieron aciertos del 92,9 %, 87,4 % y 78,7 % en tareas de clasificación binaria utilizando umbrales de gravedad crecientes.</p>	<p>El aprendizaje se puede utilizar junto con la información de SpO2 adquirida en el hogar de los pacientes para ayudar en la simplificación del diagnóstico de SAHS.</p>
<p>Huang W, et al, 2018. [30]</p>	<p>Prediction of Obstructive Sleep Apnea Using Machine Learning Technique.</p>	<p>5.301 pacientes.</p>	<p>Support Vector Machine (SVM). Se volvieron a tomar mediante la técnica de sobremuestreo de minorías sintéticas.</p>	<p>Modelo para predecir OSA basado en SVM dio un área bajo la curva de 0.818, precisión del 77.5%, sensibilidad del 86.5% y especificidad del 57.4%.</p>	<p>El modelo de 9 características tuvo un buen desempeño para identificar pacientes con AOS. Podría aplicarse para priorizar a los pacientes para la polisomnografía que puede traer un tratamiento más temprano.</p>
<p>Muhammed Kursad Ucar, et al, 2018. [1]</p>	<p>Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques.</p>		<p>Adquisición de señales. Preprocesamiento de señales. Extracción de características de la fotoplethysmografía y señales de variabilidad de la frecuencia cardíaca. Selección de funciones usando F-score. Etapa de clasificación. Criterios de rendimiento.</p>	<p>Implementando algoritmos de clasificación (k-vecinos más cercanos y support vector machine) se obtuvo una precisión 73,36%, sensibilidad 81% y especificidad del 77%.</p>	<p>Las señales PPG y HRV pueden utilizarse para el proceso de estadificación del sueño. La mejora de los sistemas en los que se utilizará esta señal mejorará los métodos de diagnóstico.</p>

<p>Satoru Tsuiki, et al, 2021. [31]</p>	<p>Machine learning for image-based detection of patients with obstructive sleep apnea: an exploratory study.</p>	<p>1389 radiografías cefalométricas laterales obtenidas de individuos diagnosticados con AOS grave o sin AOS.</p>	<p>Se desarrolló una red neuronal convolucional profunda y se probó con la muestra. Se prepararon tres datos en una sola imagen: la imagen original sin ninguna modificación, una imagen que contiene un perfil facial y una imagen que contiene parte de la región occipital</p>	<p>Utilizando una red neuronal convolucional profunda se obtuvo una sensibilidad y especificidad del 87% y 82% para a imagen completa, 88% y 75% para la región principal, 71% y 63% para la cabeza y 54% y 80%.</p>	<p>Una red neuronal convolucional profunda, identificó a las personas con AOS grave con alta precisión. La investigación futura sobre este concepto utilizando IA e imágenes puede fomentarse aún más cuando se discuta la clasificación de la AOS.</p>
--	---	---	---	--	---

A continuación se describen los trabajos relacionados en la Tabla 1, para un análisis más completo y detallado:

- Regularized logistic regression for obstructive sleep apnea screening during wakefulness using daytime tracheal breathing sounds and anthropometric information:** Este artículo de revista, muestra el desarrollo y aplicación de un modelo de regresión logística regularizada con una penalización LASSO para la predicción de la gravedad de la AOS utilizando información antropométrica y espectros de potencia de los sonidos respiratorios (TBS).

Las características seleccionadas por este enfoque de modelado de regresión, son congruentes con las seleccionadas por otros métodos en los estudios anteriores; son fisiológicamente significativas y están altamente correlacionadas con el índice de apne/hipoapnea (IAH). Los datos de este estudio se recogieron de 199 participantes con sospecha de AOS, antes de la evaluación nocturna con PSG en el Laboratorio de Trastornos del Sueño del Centro de Salud Misericordia (Winnipeg, MB, Canadá).

Después de la toma de datos, los participantes procedieron a la preparación de la PSG durante la noche y a la evaluación del sueño. Se recogió su información antropométrica y calculó su IAH (a partir de la evaluación PSG) para ser comparados con la predicción del IAH de este método propuesto. La precisión para todo el conjunto de datos fue del 81.1%, con una sensibilidad y especificidad del 84.4% y el 77.0%, respectivamente.

Las contribuciones más importantes de este trabajo son:

- (1) El método presentado en este trabajo no necesita la identificación de la fase respiratoria, ya que las mejores características no dependen de ello.
- (2) El método es simple, rápido y computacionalmente más eficaz que el SVM y otros métodos anteriores para el cribado de la AOS durante la vigilia; por lo tanto, es adecuado para aplicaciones en línea.
- (3) el método puede utilizarse para el cribado de individuos con AOS con una gravedad tan baja como 15 con una alta sensibilidad y una especificidad razonable.

La principal limitación de este estudio es la falta de un tamaño de muestra grande para tener una mejor manera de manejar la predicción de la gravedad de la AOS de los individuos con $5 < \text{AHI} < 15$.

Hasta donde se sabe, esta fue la primera investigación en la que se utilizan estas técnicas para ayudar al diagnóstico de la AOS a partir de señales de TBS registradas durante el día, validando el enfoque en un mayor número de participantes en comparación con el de estudios anteriores [2].

- **Application of machine learning to predict obstructive sleep apnea syndrome severity:** El artículo de revista muestra la eficacia y la aplicabilidad clínica de la implementación y el estudio de diferentes métodos de aprendizaje automático basados en información demográfica y datos de

cuestionarios, para predecir la gravedad del síndrome de apnea obstructiva del sueño.

Se recogieron datos sobre las características demográficas, los valores de espirometría, el intercambio de gases (PaO₂, PaCO₂) y los síntomas (escala de somnolencia de Epworth, ronquidos, etc.) de 313 pacientes con diagnóstico previo de síndrome de apnea obstructiva del sueño. Tras el análisis de componentes principales, se seleccionaron 19 variables que se utilizaron para el preprocesamiento y para entrenar finalmente siete tipos de modelos de clasificación y cinco tipos de modelos de regresión.

Los resultados comparativos mostraron la superioridad de los modelos de máquina de vectores de apoyo y “Random Forest” para la clasificación, mientras que la máquina de vectores de apoyo y la regresión lineal fueron más adecuadas para predecir el índice de apnea. La mejor precisión es del 44.7%, con la misma sensibilidad promedio y los resultados de la regresión muestran un error cuadrático medio mínimo alcanzado de 22.17%.

Los resultados obtenidos de este estudio con los datos disponibles sugieren el uso de métodos de aprendizaje automático como herramientas para proporcionar a los pacientes un nivel de prioridad para la prueba de polisomnografía, pero aún no pueden usarse para el diagnóstico automatizado y también han confirmado que la problemática de predecir el índice de apnea-hipopnea es muy difícil cuando se utilizan solo los datos recopilados antes de la prueba de polisomnografía [3].

- **Predicting Nondiagnostic Home Sleep Apnea Tests Using Machine Learning:** La investigación científica busca optimizar el diagnóstico de la AOS, mediante el uso de modelos predictivos para identificar a los pacientes que deberían ser remitidos directamente a la polisomnografía (PSG).

Los modelos se entrenaron con el 80% de los datos disponibles y se validaron con el 20% restante. El rendimiento se evaluó mediante el área parcial bajo la curva de precisión y recuperación (pAUPRC). Las técnicas de aprendizaje automático dieron sistemáticamente un pAUPRC más alto que la regresión logística estándar, que tuvo un pAUPRC de 0,574. El modelo de bosque aleatorio superó a todos los demás modelos (pAUPRC de 0,862).

La calibración preferida de este modelo arrojó lo siguiente: sensibilidad 0.46, especificidad 0.95, valor predictivo positivo 0.81, valor predictivo negativo 0.80. En comparación con los modelos de regresión logística estándar, los modelos de aprendizaje automático mejoran la predicción de los pacientes que requieren PSG en el laboratorio.

Estos modelos de modelos podrían implementarse en una herramienta de apoyo a la decisión clínica para ayudar a los médicos a seleccionar la prueba óptima para diagnosticar la AOS [4].

- **A machine learning-based test for adult sleep apnea screening at home using oximetry and airflow:** El estudio utiliza las variables de saturación de oxígeno en sangre (SpO₂) y el flujo de aire para realizar un prediagnóstico de OSA usando técnicas de aprendizaje automático.

Un total de 303 pacientes mayores de 18 años con sospecha de padecer AOS participaron. Se evaluó el posible aumento del rendimiento de las pruebas de detección de AOS simplificadas cuando se utilizan conjuntamente los registros de SpO₂ y de flujo de aire.

Se utilizaron métodos de procesamiento de señales y de aprendizaje automático para conocer la complementariedad de estos registros en un entorno no atendido, también, un exhaustivo procedimiento de selección

automatizada de características condujo a un subconjunto óptimo de características compuesto por variables de oximetría y flujo de aire casi en la misma proporción.

Se concluyó que el análisis conjunto de los registros simultáneos de SpO2 y el flujo de aire mediante técnicas de aprendizaje automático proporcionan estimaciones precisas del IAH, lo que sugiere su uso como prueba de cribado de rutina extensiva para la AOS en el hogar [6].

- **Clinical Prediction Models for Sleep Apnea: The Importance of Medical History over Symptoms:** Esta investigación científica utiliza un nuevo método de aprendizaje automático conocido como SLIM (Supersparse Linear Integer Models), para probar la hipótesis de que una herramienta de cribado de diagnóstico basada en la información médica disponible de forma rutinaria sería superior a otra, basada únicamente en los síntomas relacionados con el sueño comunicados por el paciente.

Se analizó la PSG y la información clínica autodeclarada de 1922 pacientes examinados en el laboratorio clínico del sueño. Se utilizó SLIM y 7 métodos de clasificación de última generación para producir modelos predictivos para el cribado de la AOS.

Para el diagnóstico de la AOS, se observó que el rendimiento del modelo que utilizaba sólo las características de la historia clínica, era superior al rendimiento del modelo que utilizaba sólo los síntomas y similar al rendimiento del modelo que utilizaba todas las características. El rendimiento fue similar al de otras herramientas ampliamente utilizadas: sensibilidad del 64.2% y especificidad del 77%.

La precisión de SLIM fue similar a la de los modelos de clasificación más avanzados aplicados a este conjunto de datos, pero con la ventaja de la total transparencia, lo que permite una predicción práctica utilizando respuestas sí/no a un pequeño número de consultas clínicas.

SLIM produce una herramienta clínica procesable que puede aplicarse a los datos que están disponibles de forma rutinaria en las historias clínicas electrónicas modernas, lo que puede facilitar automatizado, en lugar de manual, el cribado de la AOS [28].

- **Evaluation of Machine-Learning Approaches to Estimate Sleep Apnea Severity from At-Home Oximetry Recordings:** Este estudio muestra la hipótesis sobre el proceso de diagnóstico del trastorno del sueño potencialmente grave (SAHS) que puede simplificarse mediante el uso de un enfoque de aprendizaje automático y la información contenida en la señal de SpO2 en el hogar.

Participaron 320 sujetos adultos remitidos al Hospital Universitario Río Hortega de Valladolid (España) por sospecha de SAHS. Todos ellos fueron diagnosticados mediante una PSG de noche (E-series, Compumedics).

Los participantes fueron divididos en dos conjuntos: un conjunto de entrenamiento compuesto por el primer 60% consecutivos (n_{tr}=193, 19 sin SAHS, 31 leves, 35 moderados, 108 graves) y un conjunto de prueba compuesto por el 40% restante 40% (n_{test}=127, 10 sin SAHS, 24 leves, 21 moderados, 72 graves).

Se demuestra y explica las tendencias relacionadas con la gravedad del SAHS en las características estadísticas, espectrales, no lineales y clínicas extraídas de los registros de SpO2 obtenidos en el domicilio de los pacientes.

El modelo AB-LDA de aprendizaje automático, validado rigurosamente, alcanzó la mayor capacidad de diagnóstico.

Estos resultados sugieren que el enfoque de aprendizaje automático puede utilizarse junto con la información de SpO2 adquirida en el domicilio de los pacientes para ayudar a simplificar el diagnóstico del SAHS [29].

- **Prediction of Obstructive Sleep Apnea Using Machine Learning Technique:** El estudio científico propuso un nuevo modelo para predecir la AOS con las características mínimas. Se recogieron las características clínicas y los parámetros polisomnográficos de 5301 pacientes (edad media de 47.5 ± 14.4 años, hombres 76,5%).

Entre los 5301 pacientes, la media del IAH fue de $29.7 \pm 26,1/h$ y el 70.1% tenía AOS. Se utilizaron 35 características que incluían comorbilidades, información antropométrica y síntomas para investigar la relevancia de la AOS.

El modelo de nueve características proporcionó un AUROC, sensibilidad, especificidad, precisión LR (+) y LR (-) de 0.818, 86.5%, 57.4%, 77.5%, 2.03 y 0.24, respectivamente. El modelo de 9 características tuvo un buen rendimiento para identificar a los pacientes con AOS. Este podría aplicarse para priorizar a los pacientes para polisomnografía, lo que puede permitir un tratamiento más temprano [30].

- **Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques:** Este estudio busca identificar las etapas del sueño de la manera más eficiente y práctica, evaluando un nuevo método de estadificación del sueño basado en el algoritmo de clasificación k-nearest neighbors (kNN) y máquinas de vectores de apoyo (SVM).

Se extrajeron 46 características de la señal PPG y 40 características de la señal HRV. Las características extraídas de PPG y HRV se clasifican con él con el método de aprendizaje automático, utilizando tanto el método de extracción de características de cifras brutas como el de Fscore.

La base de datos utilizada en el estudio se formó en el Laboratorio de Laboratorio de Enfermedades Torácicas del Hospital Público Sakarya Hendek, formada a partir de datos de 33 canales obtenidos, grabando a 10 personas a lo largo de la noche con dispositivo PSG de la marca SOMNOscreen Plus.

La tasa de precisión obtenida para el sueño-vigilia es de aproximadamente el 70-80%. Se concluyó que las señales de PPG y HRV pueden utilizarse para el proceso de del sueño. Es una ventaja que la señal PPG pueda medirse más práctica que las otras señales de estadificación del sueño [1].

- **Machine learning for image-based detection of patients with obstructive sleep apnea: an exploratory study:** En este estudio se prueba la hipótesis que el aprendizaje automático, una aplicación de la inteligencia artificial (IA), podría utilizarse para detectar a los pacientes con AOS grave basándose en imágenes bidimensionales.

Se desarrolló una red neuronal convolucional profunda (n = 1258; 90%) y se probó (n = 131; 10%) utilizando datos de 1389 (100%) radiografías cefalométricas laterales obtenidas de individuos diagnosticados con AOS grave (n = 867; índice de apnea-hipopnea > 30 eventos/h de sueño) o sin AOS (n = 522; índice de apnea-hipopnea < 5 eventos/h de sueño) en un único centro de trastornos del sueño.

La sensibilidad/especificidad fue de 0.87/0.82 para la imagen completa, 0.88/0.75 para la región principal, 0.71/0.63 para la cabeza solamente y 0.54/0.80 para el análisis manual. El área bajo la curva de características operativas del receptor fue la más alta para la región principal: 0.92, para la imagen completa: 0.89, para la cabeza solamente: 0.70, y para el análisis cefalométrico manual: 0.75.

Una red neuronal convolucional profunda identificó a los individuos con AOS grave con gran precisión. Se puede fomentar la investigación futura de este concepto mediante el uso de IA e imágenes cuando se hable del triaje de la AOS [31].

3.2. MARCO TEÓRICO

En esta sección se describen los principales conceptos, teorías y términos necesarios para el correcto desarrollo y entendimiento de este estudio.

3.2.1 Apnea del sueño

La apnea del sueño es un trastorno del sueño grave en la cual la respiración se detiene y empieza repetidas veces [32]. La apnea es la ausencia de flujo de aire inspiratorio durante al menos 10 segundos y se presenta en aproximadamente el 10% de los adultos de la población general, pero en determinadas enfermedades cardiovasculares, y en condiciones particulares caracterizadas por retención de sodio y agua [33][34].

El diagnóstico de la apnea del sueño regularmente se basa en la demostración de al menos 10 a 15 apneas e hipopneas por hora de sueño y cuando adicionalmente presenta 1 o más síntomas de ronquidos, sueño inquieto, dolores de cabeza matutinos y excesiva somnolencia diurna, constituye un síndrome de apnea del sueño [35].

3.2.2 Tipos de apnea del sueño

Una apnea es la ausencia de flujo de aire inspiratorio durante al menos 10 segundos. Una hipopnea es una disminución menor del flujo de aire, que puede durar 10 segundos o más, y se da parte a una disminución de la saturación de oxihemoglobina arterial o una excitación electroencefalográfica.

Las apneas y las hipopneas se clasifican como obstructivas o centrales, pero independientemente, son el resultado de una ausencia o reducción de la producción neural del tronco encefálico hacia los músculos de las vías respiratorias superiores y/o músculos de la bomba inspiratoria torácica inferior (diafragma y músculos intercostales) [34].

El patrón de salida neural determina el fenotipo: La apnea obstructiva del sueño ocurre cuando se presenta la oclusión completa de la vía aérea superior (flujo de aire ausente, lengua cayendo hacia atrás) frente a la actividad continua de los músculos de la bomba torácica inspiratoria. Por el contrario, la apnea central del sueño ocurre cuando hay una reducción transitoria por parte del marcapasos pontomedular en la generación del ritmo respiratorio [34][36].

3.2.3 Apnea central del sueño

La apnea central del sueño es un trastorno en el que disminuye el esfuerzo respiratorio o presenta una intermitencia cíclica durante el sueño. La apnea central es causada por una falla temporal de los centros bulbares inspiratorios y espiratorios, que constituyen el generador central del ritmo respiratorio. Un evento de apnea central ocurre cuando hay una interrupción del flujo del aire del 90% durante 10 segundos o más [37][38].

La apnea central es un trastorno con baja influencia en la sociedad, menor al 1% en población general y al 10% en pacientes diagnosticados con trastornos de sueño, además, la presencia de esta en fallas cardíacas es del 25 al 40% y en enfermedades cerebrovasculares del 20% [34].

3.2.4 Apnea obstructiva del sueño

Los trastornos respiratorios del sueño engloban un grupo heterogéneo de trastornos relacionados con el sueño que se caracterizan por pausas anormales en la respiración durante el sueño. La AOS es el tipo más común y constituye más del 85% de todos los casos de trastornos respiratorios del sueño; CSA es mucho menos común [32].

La apnea obstructiva del sueño (AOS) es un trastorno muy común que se caracteriza por parálisis repetitivas del flujo del aire en las vías respiratorias faríngeas durante el sueño (ver Figura 1). La posterior reducción del flujo de aire a

menudo conduce a trastornos agudos en el intercambio de gases y despertares frecuentes en el sueño [39].

Las consecuencias para la salud de la persona que padece de AOS son numerosas, si no se trata, provoca somnolencia diurna excesiva, disfunción cognitiva, deterioro del rendimiento laboral y disminución de la calidad de vida, todo esto relacionado con la salud de las personas que lo padecen. La evidencia experimental también sugiere que la AOS puede contribuir al desarrollo de hipertensión sistémica, enfermedades cardiovasculares y anomalías en el metabolismo de la glucosa [40].

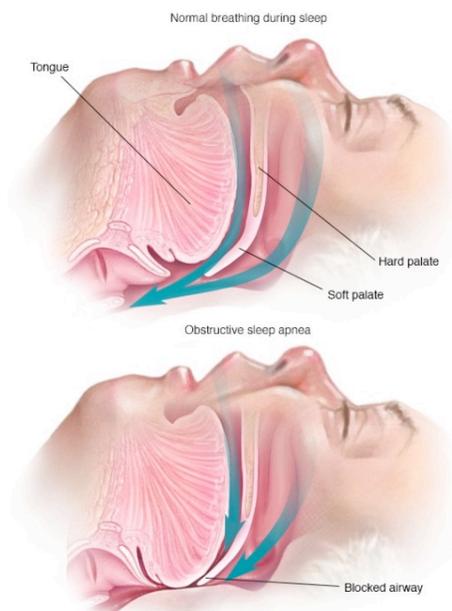


Figura 1. Apnea Obstructiva del Sueño (Mayo Clinic, 2019).

Estudios epidemiológicos recientes informaron una mayor incidencia y crecimiento de AOS entre la población general, así como una prevalencia sustancialmente mayor en grupos con un riesgo particularmente alto de consecuencias adversas anteriormente mencionadas [6]. La AOS es un desafío médico y económico que afecta del 4% al 37% de la población adulta según los criterios diagnósticos utilizados y la población estudiada [4].

3.2.5 Diagnostico

Los factores de riesgo como lo son: jadeo observado durante el sueño, dolor de cabeza matutino, somnolencia diurna excesiva, ronquidos fuertes y una

circunferencia del cuello de más de 16 pulgadas; además de numerosos signos y síntomas que pueden sugerir AOS, la polisomnografía (PSG) ha sido y sigue siendo el método estándar para el diagnóstico [41].

La PSG es el estándar de oro para el diagnóstico de trastornos respiratorios relacionados con el sueño. Es un proceso sistemático que se utiliza para recopilar parámetros fisiológicos durante el sueño, es un procedimiento que utiliza electroencefalograma, electrooculograma, electromiograma, electrocardiograma y oximetría de pulso, así como también el flujo de aire y el esfuerzo respiratorio, para evaluar las causas latentes de los trastornos del sueño (ver Figura 2) [42][43].

La PSG también se puede utilizar para evaluar otros trastornos del sueño, que incluyen convulsiones nocturnas, narcolepsia, trastorno del movimiento periódico de las extremidades y trastorno del comportamiento del sueño con movimientos oculares rápidos.

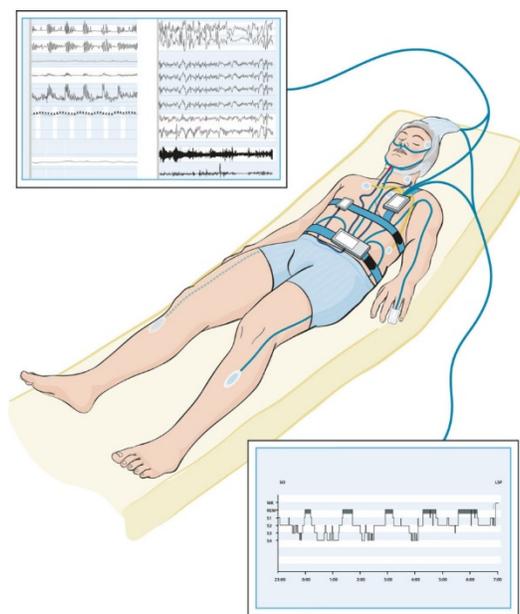


Figura 2. Polisomnografía (Clínica Aystesis, 2017).

Aunque la PSG es el método estándar, esta presenta ciertas problemáticas y limitaciones para su implementación, debido a que el los pacientes gastas mucho tiempo y dinero haciéndose valoraciones, además de las largas listas de espera, sin mencionar la incomodidad que presentan las personas a la hora de la prueba. Por

tal razón, surge la necesidad de desarrollar nuevos métodos alternativos para la detección de personas con AOS [16].

La poligrafía respiratoria (HRP) es un método similar a la PSG realizada en el domicilio, es método usado para el diagnóstico de AOS en paciente con un alto riesgo de padecerla y además, cumple con ciertos requerimientos técnicos de calidad. Una de las principales desventajas, es que no permite evaluar las variables neurofisiológicas, sin embargo, se pueden inferir datos como la cantidad de sueño con la ayuda de los canales de ronquido, posición, actimetría, entre otros [44].

3.2.6 Estrategias para el Tratamiento de OSA

En cuanto a las medidas a tomar para la AOS, en todos los casos se deben de seguir unas recomendaciones generales de prevención, claramente dependiendo de la gravedad y demás factores, todo esto son decisiones tomadas por los médicos. Entre los más comunes tratamientos se encuentran:

Los tratamientos para la AOS varían según los síntomas, la gravedad, la edad, el peso corporal y la historia clínica de cada paciente. Para los pacientes con AOS leve, se recomiendan buenos hábitos de sueño y medidas dietéticas que incluyen evitar la posición supina durante el sueño, el ejercicio, la pérdida de peso, la abstinencia de fumar y el consumo de alcohol. Sin embargo, estas estrategias normalmente son insuficientes para las formas graves de AOS [45]. En pacientes con AOS grave, la presión positiva continua en las vías respiratorias (CPAP) se considera la opción de tratamiento de primera línea [46].

Presión positiva en las vías aéreas.

Las terapias de presión positiva vienen dadas en presión positiva continua en las vías respiratorias (CPAP), presión positiva de dos niveles en las vías respiratorias y presión positiva auto titulada en las vías respiratorias [41].

La CPAP es el tratamiento estándar para la apnea obstructiva del sueño. Es un aparato que usa un tubo y una mascarilla o un dispositivo para la nariz para proporcionar una presión constante y estable en las vías respiratorias obstruidas (ver Figura 3). Esta terapia ha demostrado ser un tratamiento eficaz para mejorar los síntomas de apnea obstructiva del sueño, como somnolencia diurna. Funciona mediante ferulización neumática de las vías respiratorias superiores [47].

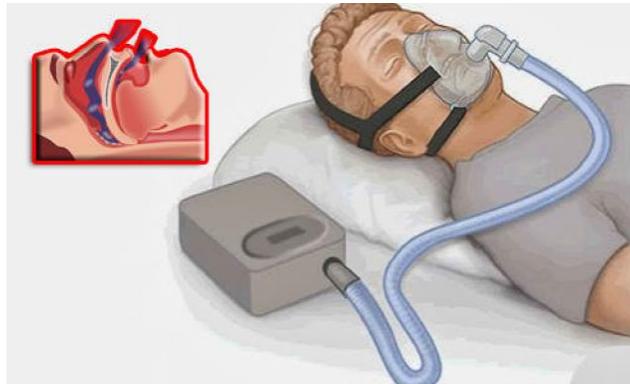


Figura 3. CPAP (Clínica de cirugía maxilofacial, 2019).

Una encuesta de las interfaces de CPAP demostró que los pacientes prefieren las mascarillas nasales, pero se requieren más estudios para evaluar los beneficios comparativos de cada sistema de administración. Las tasas de adherencia a la CPAP oscilan entre el 17% y el 85%, con una mejor adherencia en los pacientes que reciben educación temprana, continua y capacitaciones sobre el uso de CPAP.

Se ha demostrado que no solo mejora la calidad de vida y los índices de sueño en pacientes con AOS, sino que también disminuye la presión arterial, la arritmia, accidente cerebrovascular, mejora la fracción de eyección del ventrículo izquierdo en pacientes con insuficiencia cardíaca y mejora las tasas de eventos cardiovasculares fatales y no fatales [41].

Dispositivos orales.

Los aparatos orales son una alternativa razonable si los pacientes presentan dificultades con la CPAP. Las dos terapias principales con aparatos orales son los dispositivos de avance mandibular (MAD), que mueve la mandíbula del paciente hacia adelante para mantener la vía aérea abierta como se muestra en la Figura 4,

y los dispositivos de retención de la lengua, que inmovilizan la lengua en el lugar que le corresponde para mantener la vía aérea abierta como se muestra en la Figura 5 [41][47].

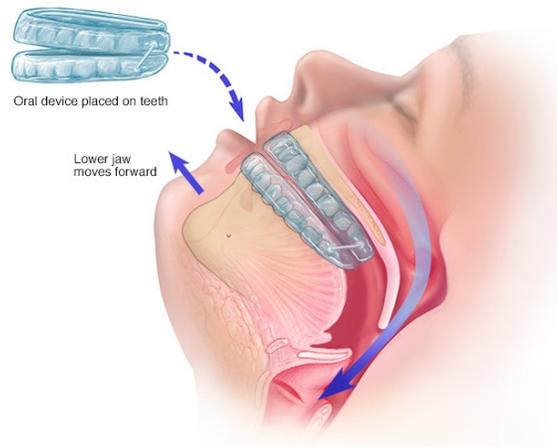


Figura 4. Dispositivo de avance Mandibular (Mayo Clinic, 2019).



Figura 5. Dispositivo de retención de Lengua (Denlattix, 2019).

Es muy importante aclarar que la evidencia a corto plazo que arroja la implementación de la MAD es convincente, pero al mismo tiempo es poco claro si estos beneficios obtenidos por los aparatos orales se mantienen. Hasta hace poco, ha prevalecido una falta de información a más largo plazo, pero los investigadores han comenzado a abordar esta brecha de conocimiento [48].

Cambio de estilo de vida.

La obesidad es uno de los principales factores de riesgo, debido a que produce depósitos de grasa alrededor del cuello, que contribuyen al colapso faríngeo. Aunque se ha demostrado que una disminución de peso disminuye las presiones

críticas de cierre de las vías respiratorias, existen hallazgos inconsistentes sobre la asociación entre la reducción de peso y la mejora general del sueño junto con la respiración.

Las estrategias para evitar la posición supina incluyen colocar pelotas de tenis en un calcetín o bolsillo y sujetarlas en la parte de atrás de una camisa; usar chalecos con parachoques posteriores; y el uso de alarmas posicionales, instrucciones verbales y almohadas. Cabe resaltar que respecto al cumplimiento deficiente a largo plazo, la terapia posicional no se recomienda de manera rutinaria sobre la terapia de presión positiva en las vías respiratorias estándar [49][50].

Cirugía.

Como una alternativa correctiva de casos severos de AOS están las intervenciones quirúrgicas. La cirugía consiste en una alteración general de la vía aérea respiratoria, donde el especialista responsable (cirujano) debe realizar una modificación correcta y satisfactoria [51].

Se ha propuesto numerosas cirugías para corregir la obstrucción anatómica en pacientes con AOS: procedimientos nasales, orales, hipofaríngeos, laríngeos y procedimientos globales de las vías respiratorias, pero la cirugía bariátrica frecuentemente se promociona al paciente como la opción más efectiva para tratar la obesidad y sus comorbilidades, incluida la AOS [36].

Según las revisiones sistemáticas, no hay evidencia y estudios suficientes que respalden la cirugía en general o cualquier tipo de cirugía en particular como tratamiento para las personas que padecen de AOS. En el contexto clínico, a menudo se han empleado modelos estadísticos como los modelos lineales multivariados o de regresión logística para identificar la relación entre diferentes factores clínicos (por ejemplo, datos demográficos, comorbilidades y síntomas) y la gravedad de la AOS [50].

Las principales ventajas de estos métodos son su simplicidad e interoperabilidad de los resultados obtenidos. Sin embargo, tales modelos tienen limitaciones en situaciones en las que la relación entre las características clínicas y el resultado no es lineal o cuando las características tienen una alta correlación entre sí [51].

Actualmente, los algoritmos de aprendizaje automático (ML) basados en el modelado predictivo se utilizan para abordar múltiples limitaciones de los métodos estadísticos tradicionales. ML se refiere a la aplicación y desarrollo de algoritmos para analizar datos, y aprender de ellos para realizar una tarea en particular (es decir, clasificación o predicción) o tomar una decisión automáticamente [52].

Son dos tipos principales de métodos que se utilizan para aplicar ML:

El aprendizaje supervisado se utiliza para desarrollar modelos de entrenamiento para predecir valores futuros de categorías de datos o "etiquetados" (variable dependiente) a partir de un conjunto dado de predictores (variables independientes), generando así una función que mapea las entradas a las salidas deseadas. Los métodos supervisados más comunes incluyen regresión logística (LR), bosque aleatorio (RF), máquina de vectores de soporte (SVM), red neuronal artificial (ANN), bayes (NB) y árboles de decisión (DT) [52].

El aprendizaje no supervisado es un enfoque de aprendizaje automático adecuado para un conjunto de datos sin etiqueta previa o anotación disponible. Su objetivo es aprender la relación entre las variables y descubrir patrones ocultos en el conjunto de datos analizado y después de agruparlo en categorías, incluir métodos de agrupamiento y reducción de dimensionalidad. Incluye métodos de agrupamiento (jerárquico o K-means, análisis de componentes principales, mapas autoorganizados y aprendizaje profundo [52].

3.2.7 Machine Learning

Actualmente se han desarrollado diferentes alternativas que permiten tanto el diagnóstico como la predicción del tratamiento. Una de estas alternativas es el uso de algoritmos que recopilan diferentes datos de variables para que con estos tengan la capacidad de predecir el tratamiento más óptimo que se deba realizar con el paciente, esto se le conoce como machine learning (aprendizaje automático).

¿Qué es?

El aprendizaje automático es una tecnología que pretende desarrollar algoritmos informáticos capaces de emular la inteligencia humana, aprendiendo del entorno e incorporando ideas de la neurociencia, la estadística, la informática, la teoría de la información, con aplicaciones famosas en la perspectiva de una computadora, la robótica, el entretenimiento, la ecología, la biología y la medicina. El "big data", y las técnicas de aprendizaje automático ocupan el asiento del conductor para guiar la comprensión de estos datos en la medicina [53].

3.2.8 Feature selection

Feature selection es el proceso de selección de variables relevantes, o de un subconjunto de variables candidatas. La selección de variables es un proceso que aísla y selecciona las variables más consistentes, no redundantes y relevantes para utilizarlas en la construcción de modelos (ver Figura 6) [54].

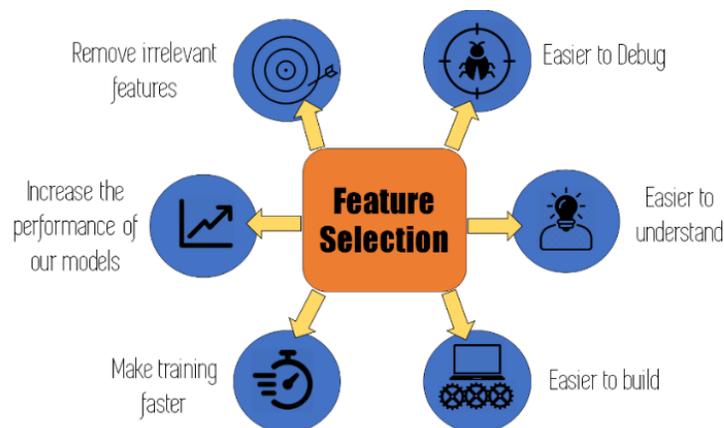


Figura 6. Feature Selection (Towards Data Science, 2021).

El objetivo principal de la selección de variables es mejorar el rendimiento de un modelo predictivo y reducir el coste computacional del modelado. Reducir metódicamente el tamaño de los conjuntos de datos es importante, ya que el tamaño y la variedad de los conjuntos de datos siguen creciendo [54].

Forward Stepwise.

Comienza con un modelo que no contiene variables (llamado Modelo Nulo), Luego comienza a añadir las variables más significativas secuencialmente en el modelo una tras otra como se muestra en la Figura 7 [55].

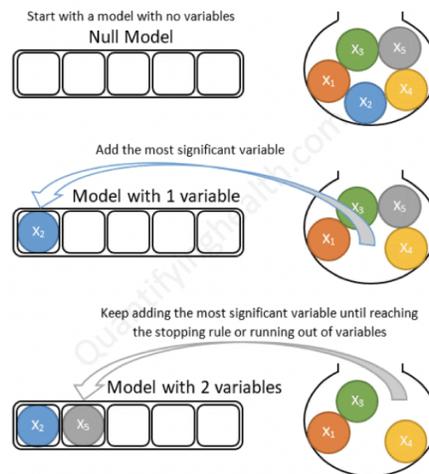


Figura 7. Método Forward Stepwise (thaddeussegura, 2020).

Este examina todas las variables predictoras seleccionadas y de estas variables la primera en ser añadida es la que posee mayor grado de correlación con la variable dependiente. Esa variable se añade al modelo, esto se repite con la variable que más predice en la medida dependiente [55].

Este pequeño procedimiento continúa hasta que la adición de predictores ya no añade nada al modelo de predicción o hasta que se incluyen en el modelo todas las variables consideradas, el procedimiento culmina cuando ya no quedan variables que cumplan el criterio de entrada [56].

Backward Stepwise.

Comienza con un modelo que contiene todas las variables consideradas (llamado Modelo Completo), estas variables se introdujeron en la ecuación y luego se comienza a eliminar las variables menos significativas una tras otra como se muestra en la Figura 8, hasta que se alcanza una regla de parada preestablecida o hasta que no queda ninguna variable en el modelo [57].

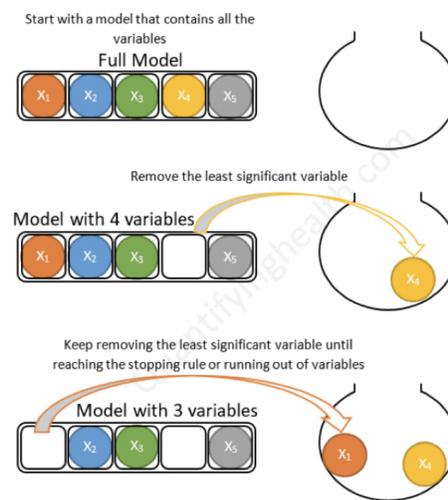


Figura 8. Método Backward Stepwise (thaddeussegura, 2020).

A continuación, las variables que no predicen (significativamente) nada en la medida dependiente se eliminan del modelo una por una. En cada fase o etapa del método se va eliminando secuencialmente la variable menos influyente, es decir la que posee menor grado de correlación [56][57].

LASSO (Least Absolute Shrinkage and Selection Operator).

Es un método que realiza una regularización y selección de características para mejorar la precisión y facilitar la forma en que se interpreta un modelo estadístico. LASSO combina un modelo de regresión con un procedimiento de contracción de

algunos parámetros hacia cero y selección de variables, imponiendo una restricción o una penalización sobre los coeficientes de regresión (ver Figura 9) [56].

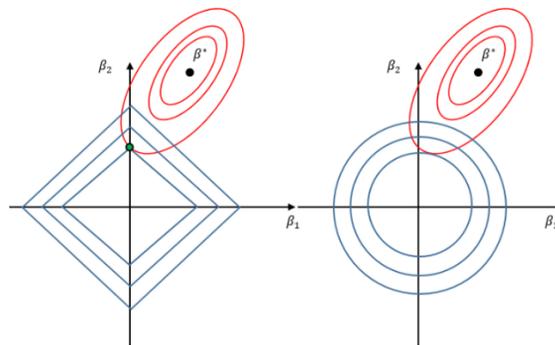


Figura 9. Método LASSO (Friedman, 2019).

En particular, la función objetivo del modelo LASSO es encontrar el número mínimo de variables independientes, lo que difiere de otros enfoques de regresión tradicionales [56].

Random Forest.

Es un método estadístico utilizado en la predicción supervisada, donde los datos se utilizan para predecir un valor de respuesta. Es una combinación de árboles de decisión en la que cada árbol depende de los valores asociados a un vector aleatorio, el cual es evaluado de forma independiente y con la misma distribución a la del resto [58].

Se han producido mejoras significativas en la precisión a la hora de clasificar y hacer crecer un conjunto de árboles, dejando que voten por la clase más popular. Para hacer crecer estos conjuntos, a menudo se generan vectores aleatorios que rigen el crecimiento de cada árbol del conjunto (ver Figura 10) [59].

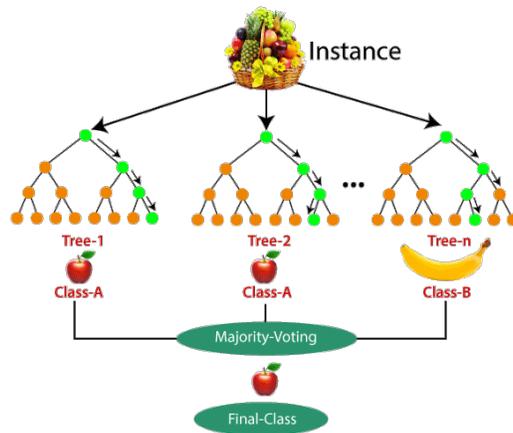


Figura 10. Método Random Forest (shishirkant, 2021).

3.2.9 Algoritmos de aprendizaje

El ideal del aprendizaje automático es emular la forma en que los seres humanos aprenden a procesar las señales sensoriales para lograr un objetivo. Cada objeto tiene características que son únicas, pero aun así somos capaces de distinguir una de otra. En lugar de codificar una máquina con muchísimas representaciones exactas de estas características, se puede programar para que aprenda a distinguirlas mediante la experiencia repetida con los objetos [60].

Cabe mencionar los tres tipos de algoritmos de aprendizaje automático que existen: supervisado, no supervisado y por refuerzo.

3.2.10 Aprendizaje supervisado

Técnicas donde los datos de entrada se emparejan con su etiqueta de clasificación conocida. Permiten a la máquina tratar con las similitudes y diferencias cuando los objetos a clasificar tienen muchas propiedades variables dentro de sus propias clases, pero siguen teniendo cualidades fundamentales que los identifican (ver Figura 11). Lo más importante es que la máquina que tenga éxito sea capaz de reconocer un perro o un gato que nunca haya visto antes [60].

Entre los modelos supervisados se encuentran:

- Regresión lineal y logística.
- Árboles de decisión.
- Redes neuronales.
- K-NN (k -nearest neighbor).
- Support vector machine.

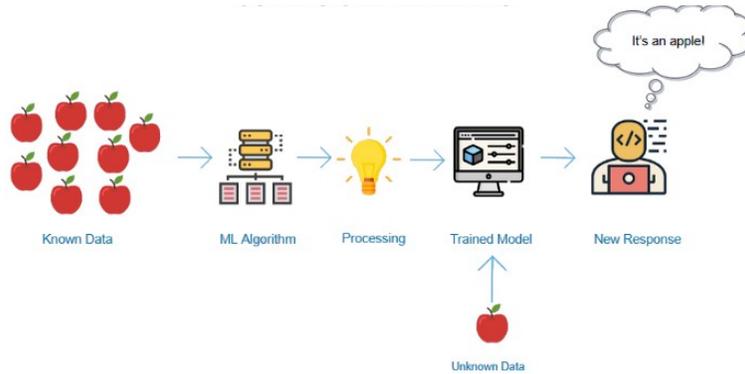


Figura 11. Representación de aprendizaje supervisado (ICHI.PRO, 2020).

3.2.11 Aprendizaje no supervisado

Son algoritmos que basan su proceso de entrenamiento en un análisis de datos sin etiquetas o un valor de salida. Este tipo de algoritmos van dirigidos a las tareas de agrupamiento, también llamadas clustering o segmentación. Esto es no supervisado, en el sentido de que el entrenamiento no asocia una determinada configuración de propiedades variables de entrada con un determinado resultado (ver la Figura 12) [60].

Entre los modelos no supervisados se encuentran:

- Agrupamiento jerárquico.
- K-meas.

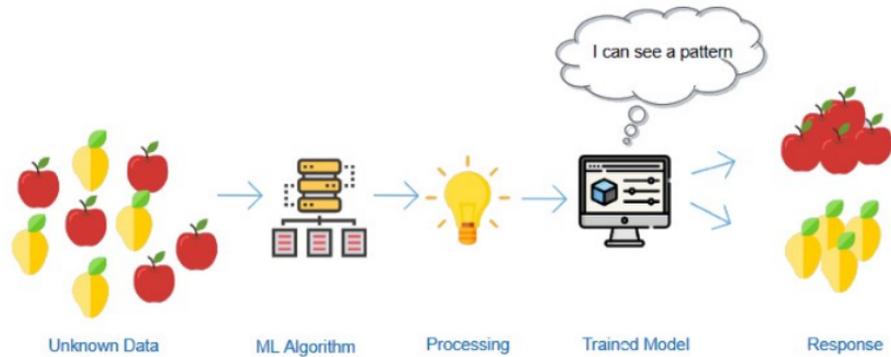


Figura 12. Representación de aprendizaje no supervisado (ICHI.PRO, 2020).

Clustering.

La agrupación o clustering es un tipo de técnica de aprendizaje no supervisado que, cuando se inicia, crea grupos automáticamente. Los elementos que poseen características similares se colocan en el mismo clúster. Este algoritmo crea grupos distintos como se ilustra en la Figura 13. La media de los valores en un cluster particular es el centro de dicho clúster [61].

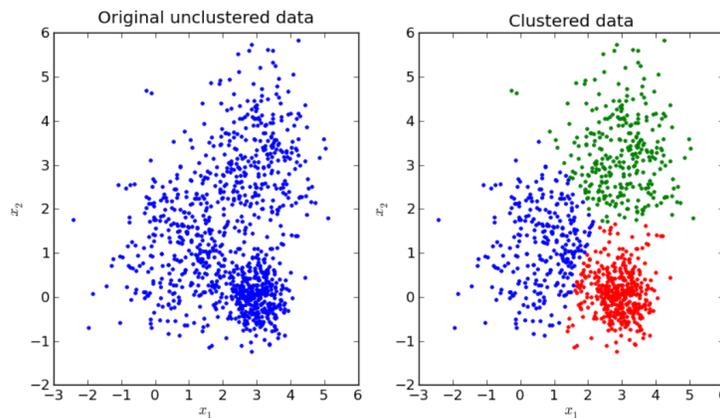


Figura 13. Agrupación (PyPRr, 2010).

El objetivo del algoritmo de clustering aglomerativo es agrupar un conjunto de n objetos basándose en una matriz de similitud $n \times n$. Estos algoritmos de clustering han crecido en popularidad debido a su capacidad para descubrir simultáneamente varias capas de estructura de clustering, y visualizar estas capas mediante

diagramas de árbol, es decir, dendrograma que es una agrupación jerárquica binaria [62].

En general los algoritmos de machine learning supervisados y no supervisados tienen la capacidad de ser usados para predecir tanto el tratamiento como el diagnóstico del paciente.

Actualmente existe una problemática en donde se manifiesta que estos algoritmos no se implementan lo suficiente en el área clínica por su complejidad para comprenderlos por el personal médico. Para ello existen ciertas alternativas como lo es usar aplicaciones web que le permitan al médico y/o personal clínico acceder a todo este tipo de información.

3.2.12 Aplicación web

Las aplicaciones web utilizadas en el área clínica y en los centros médicos son herramientas ideales para que usuarios y cuidadores de la salud puedan acceder a un servidor web que permita el acceso a ciertas funcionalidades específicas, a través de un navegador u software [63].

Para el uso generalizado de los métodos estadísticos de análisis de datos en la investigación médica, un software de código abierto eficaz es el entorno R, Python, JavaScript, entre otros. [64].

En resumen se busca desarrollar que el algoritmo que caracterice un tipo de tratamiento hacia el paciente tomando en cuenta sus variables clínicas, las cuales fueron adquiridas a través del médico mediante la consulta médica y este algoritmo será incluido en una aplicación web para que el personal clínico acceda a la plataforma y pueda hacer su análisis.

Una vez desarrollado el modelo del algoritmo se tiene como iniciativa que todo el algoritmo que se es utilizado sea embebido en una aplicación web que permita una

visualización dinámica y atractiva para quienes la utilicen, en este caso para el área clínica.

El personal médico será capaz de usar la aplicación web introduciéndoles las variables de los pacientes para proceder a obtener una respuesta predictiva por parte del algoritmo, el cual se encargara de arrojar el resultado del diagnóstico del paciente sobre la AOS, tanto como la predicción del tratamiento.

4. PROCEDIMIENTO METODOLÓGICO

4.1. METODOLOGÍA

Para el cumplimiento de los objetivos se aplican modelos predictivos supervisados y no supervisados para predecir el tratamiento con CPAP para pacientes con OSA. La Figura 14 muestra el diagrama de flujo detallado del estudio.

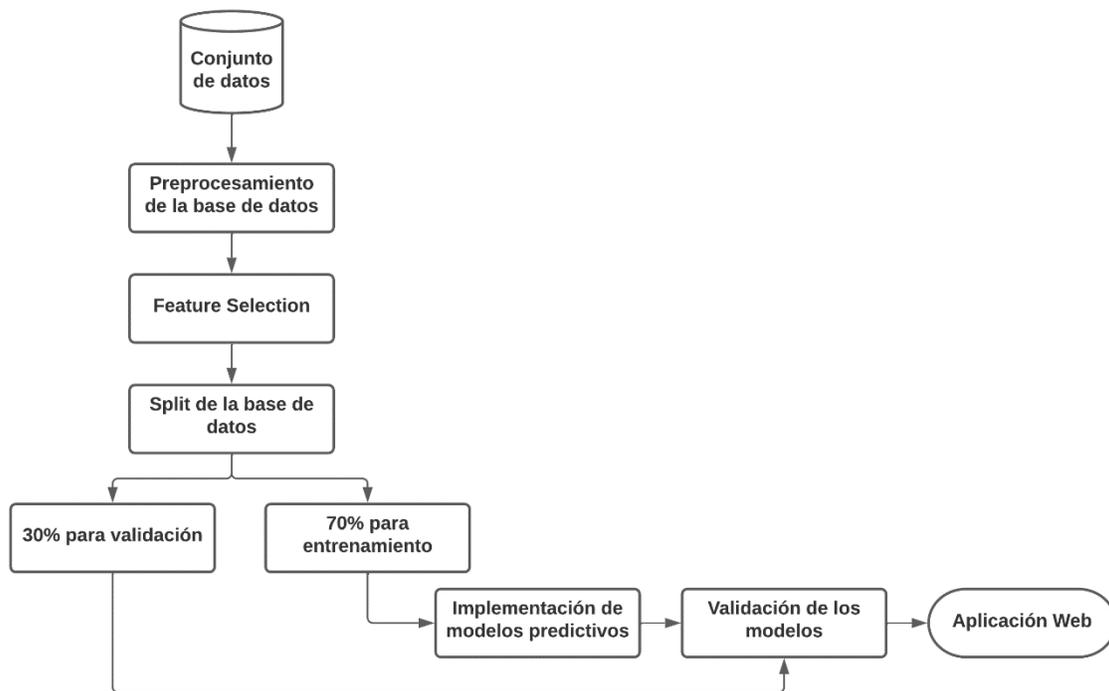


Figura 14. Metodología propuesta.

4.1.1. Conjunto de datos

Para el modelo predictivo propuesto se utiliza una base de datos la cual cuenta con 430 pacientes y X variables clínicas diagnosticado con apnea obstructiva del sueño (AOS). Los datos de las señales fisiológicas proceden del diagnóstico usando la polisomnografía (PSG) y poligrafía respiratoria (HRP). Estos datos se trabajan de forma independiente, debido a que cada estudio presenta ciertos datos asociados y que no se encuentran en el otro.

Los datos antropométricos y las variables clínicas, los datos polisomnográficos, los valores de las pruebas analíticas, el tipo de tratamiento y las comorbilidades se obtuvieron de un conjunto de datos de un estudio [65]. El conjunto de datos inicial incluyó datos de 430 pacientes diagnosticados con AOS a lo largo de tres visitas de seguimiento de CPAP.

4.1.2. Preprocesamiento de la base de datos

La base de datos con la que se trabaja cuenta con 430 observaciones y 335 predictores (variables). La mayoría de estos registros presentan valores nulos, atípicos y datos que no se pretenden usar. Estos valores causan ruido al modelo, no logrando conseguir su máximo desempeño.

El objetivo principal del preprocesamiento es la selección de un subconjunto adecuado de características y observaciones que se utilizarán para diseñar los modelos de ML. Con este fin, las observaciones que tienen datos faltantes relacionados con el uso del dispositivo CPAP se eliminaron del conjunto de datos inicial.

Posteriormente, se realizó un proceso de eliminación e imputación para las observaciones con datos faltantes. Los datos con menos del 20% de los valores faltantes se imputaron utilizando la media para las variables cuantitativas. Los predictores para en donde el número de datos nulos es mayor o igual al 20% se les eliminó del conjunto de datos.

Para llevar a cabo todo el preprocesamiento, se inicia importando la base de datos a RStudio. Este es el entorno de desarrollo integrado basado en el lenguaje R, con el que se trabajará el procesamiento de datos e implementación de modelos. Seguido de esto se convierten los datos a dataframe (estructura de datos), debido

a que este tipo de organización es más utilizada para realizar estudios estadísticos en R.

Teniendo en cuenta la variable de salida (outcome), es decir, aquella que se desea predecir, se resalta que esta debe tener un valor de cualquier tipo, si no es así, la observación correspondiente debe ser removida, debido a que no aporta nada a nuestro análisis. Para la base de datos utilizada el outcome corresponde a la variable “n_horas_dia_A”, la cual representa si el paciente es o no apto para el tratamiento de la CPAP.

Continuando con la limpieza, se eliminan las variables que no hacen parte de los tiempos correspondientes en que se tomaron los primeros datos (_B, _C, _D). También se divide la base de datos entre los pacientes sometidos a PSG y HRP, esto con el motivo de analizar ciertas variables pertenecientes únicamente a cada grupo.

La limpieza se inicia graficando y evaluando el porcentaje de datos nulos que contiene cada uno de los predictores, para posteriormente eliminar aquellos con un porcentaje mayor al 20%. En la Figura 15 y Figura 16 se muestra los porcentajes pertenecientes a las variables de cada grupo de datos (PSG y HRP).

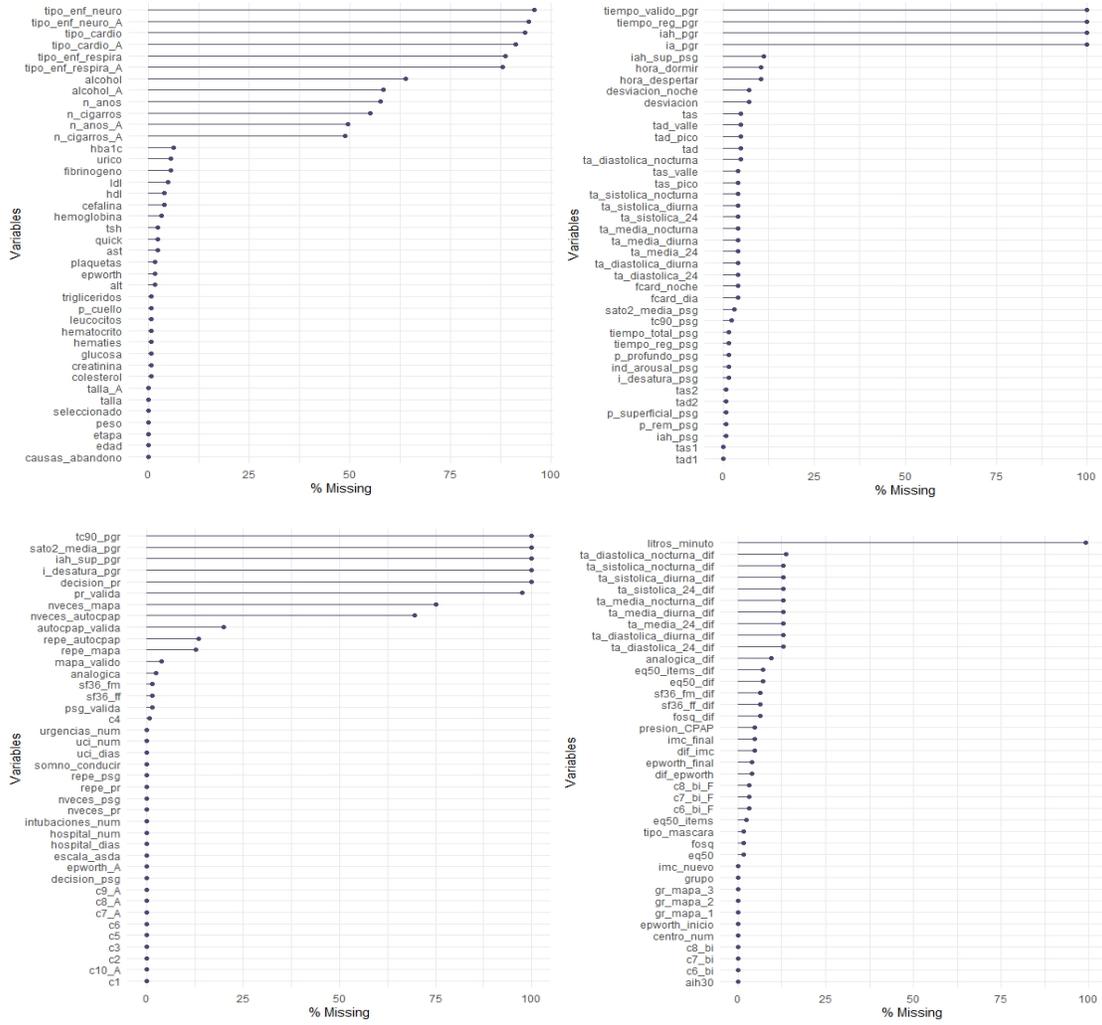


Figura 15. Porcentaje de datos nulos en los predictores del dataset PSG.

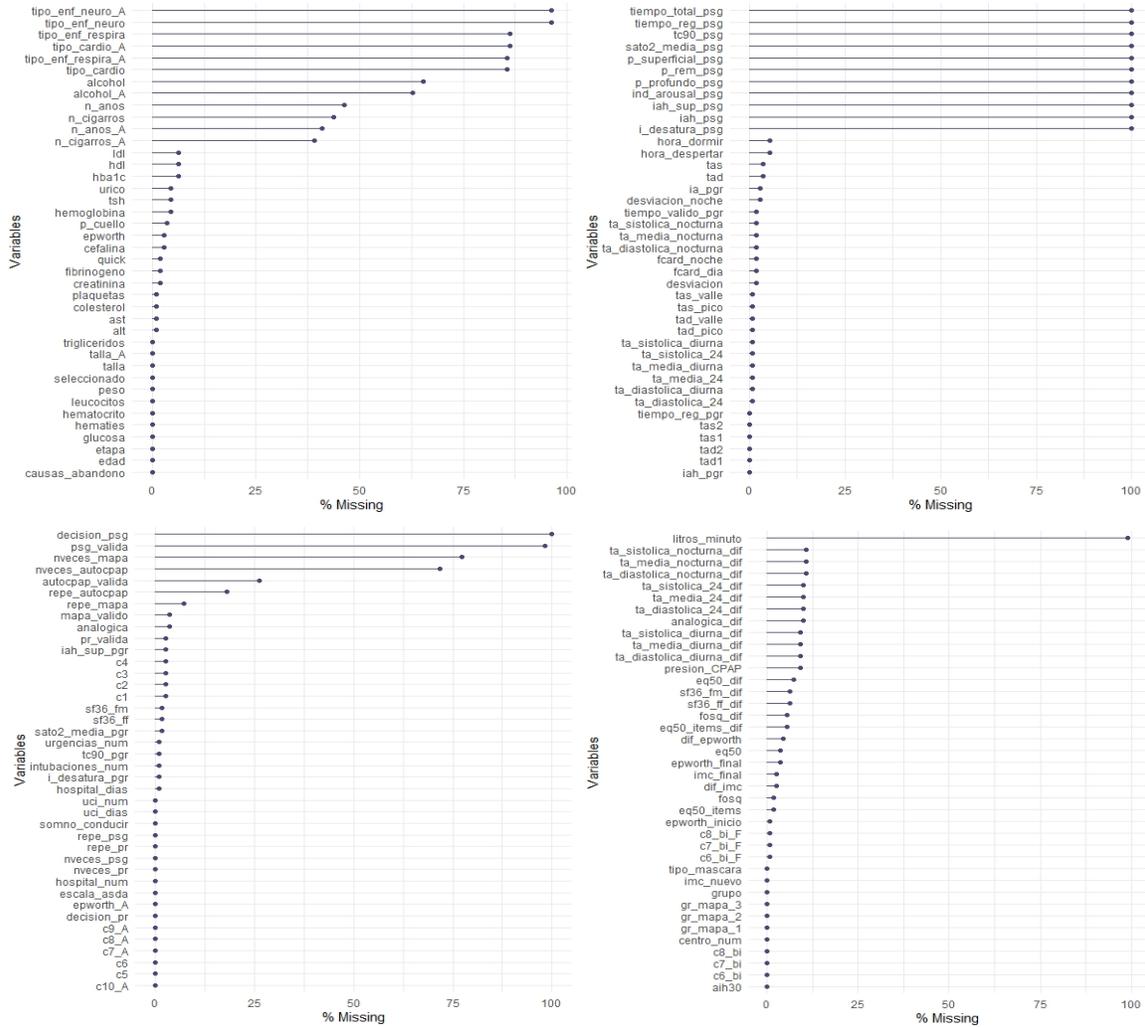


Figura 16. Porcentaje de datos nulos en los predictores del dataset HRP.

Un grupo de variables presentan datos nulos pero estos no son menores al 20%, para este caso en particular, se le imputó la media aritmética a la variable que le corresponde esa medida, todo esto para eliminar los datos nulos presentes en los datasets. Como resultado se obtienen dos dataframes parcialmente limpios, con un outcome desbalanceado, esto quiere decir que la variable a predecir sabiendo que es dicotómica presenta una cantidad muy superior de un valor en comparación con el otro.

Para el caso del dataframe PSG, el outcome presenta 30 pacientes aptos para CPAP y 95 no aptos para. En el caso de los datos de HRP, el outcome presenta 32 pacientes aptos y 78 no aptos para CPAP tal como se muestra en la Figura 17. Finalmente se procede a balancear los datasets como se muestra en la Figura 18 aplicando un proceso de imputación de observaciones.

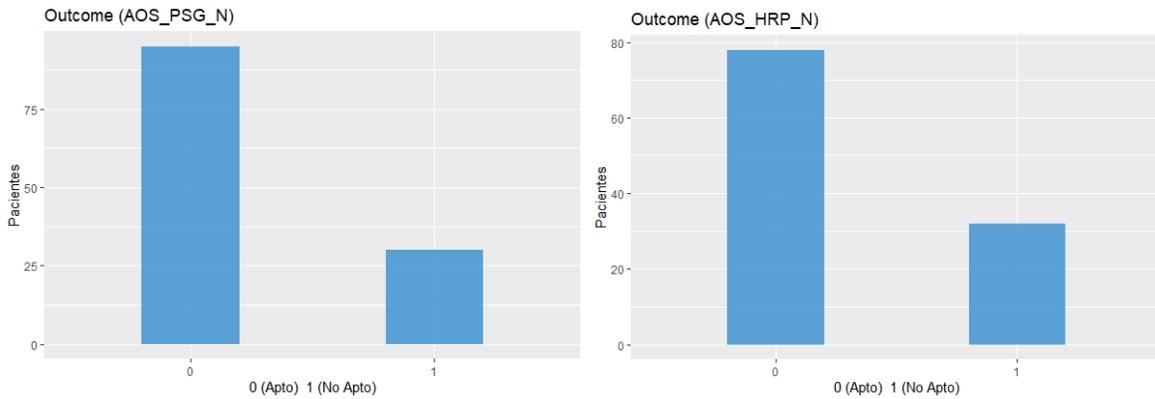


Figura 17. Dataframes con outcome desbalanceados.

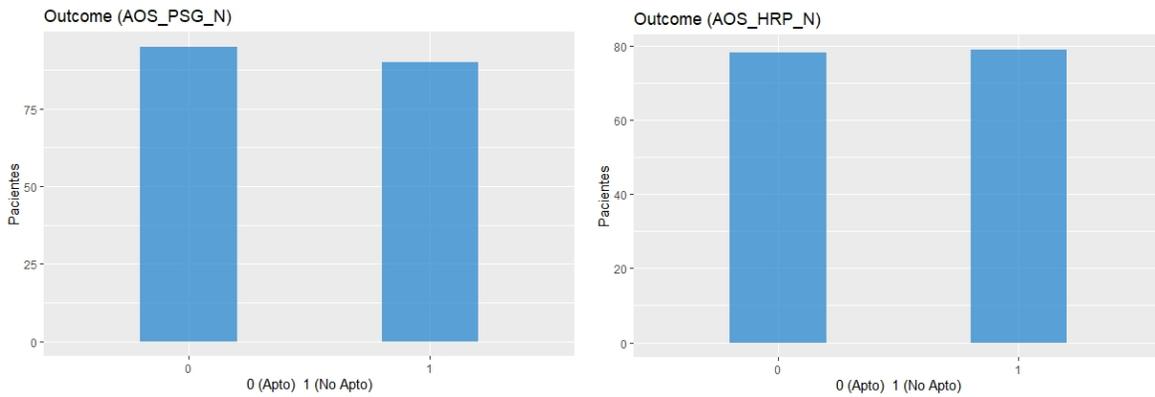


Figura 18. Dataframes con outcome balanceados.

La Figura 19 muestra un diagrama de flujo usando el método prisma, con el fin de describir la etapas de del pre-procesamiento de la base de datos.

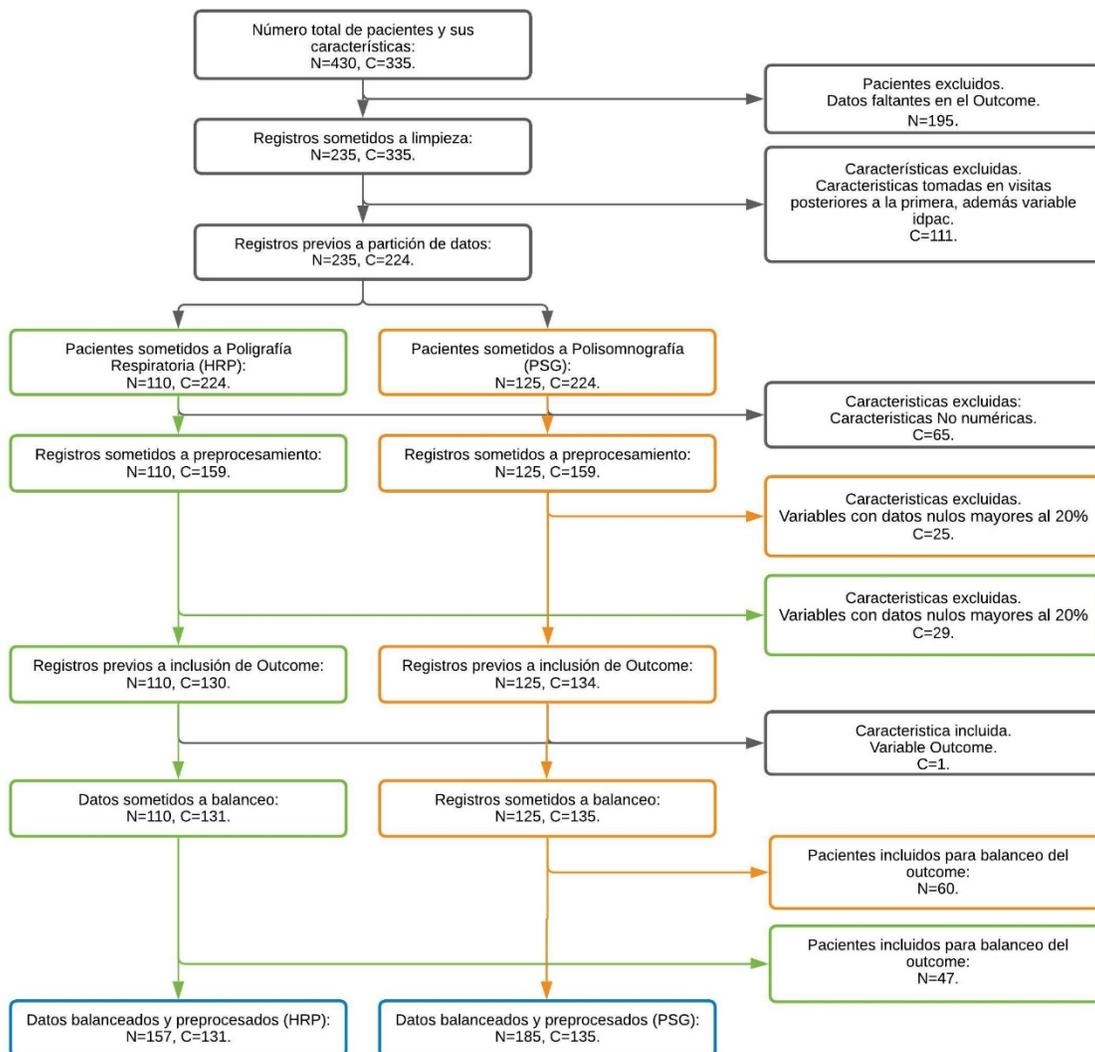


Figura 19. Resumen del preprocesamiento de la base de datos.

4.1.3. Feature Selection

Los conjuntos de datos resultantes del preprocesamiento aun presentan variables predictoras. El conjunto de datos con la prueba de PSG contiene 185 sujetos y 135 variables y el conjunto de datos con la prueba HRP contiene 157 sujetos y 131 variables, por tal razón se deben implementar métodos de selección de característica, de esta manera se obtiene un modelo eficiente y viable. Estas características son las más relevantes e importantes para llegar al resultado deseado.

En este orden de ideas se implementaron 4 métodos de feature selection, para comparar y evaluar los resultados de cada uno. Los métodos implementados son:

- Forward Stepwise Selection.
- Backward Stepwise Selection.
- LASSO.
- Random Forest.

4.1.4. Split de los datos

Una vez obtenidos los datasets correspondientes al proceso de selección de características, se realiza la división aleatoria de las observaciones de la siguiente manera: el 70% de los datos se utilizarán para el entrenamiento de los modelos a evaluar y el 30% de los datos se utilizan para la validación de dichos modelos.

4.1.5. Implementación de modelos supervisados y no supervisados

En primera instancia se implementan y evalúan modelos supervisados, esto con el fin de analizar y comparar los diferentes resultados y alcances que te arrojan los modelos de machine learning para este proyecto en particular.

Los **modelos supervisados** implementados son:

- Regresión Logística.
- K-Nearest-Neighbor.
- Redes Neuronales.
- Support Vector Machine.

El **modelo no supervisado** implementado es:

- Clustering K-meas.

4.1.6. Validación del modelo

Para los modelos supervisados utilizados se realiza una evaluación objetiva y cuantitativa de los grupos de datos utilizados. Se construyen matrices de confusión que realizan una evaluación de la predicción realizada versus los valores reales, tomando como referencia métricas como lo es la exactitud y curvas ROCs.

En el caso de los modelos no supervisados, se evalúa el número de clústeres ideal para el conjunto de datos en específico, así mismo, se obtienen las variables más representativas en la selección de los clústeres y las medias de las variables. De esta manera se conoce a detalle la agrupación y se logra entender las medidas que los hacen semejantes y diferentes a otros clústeres.

4.1.7. Métricas

Para las métricas de los modelos de lenguaje supervisados y no supervisado se utilizan la exactitud, sensibilidad, especificidad y la curva ROC [66].

Exactitud: Indica el número de elementos clasificados correctamente en comparación con el número total de artículos. Hay que tener en cuantas ciertas limitaciones: no funciona bien con las clases desequilibradas que pueden tener muchos elementos de la misma clase e incluir algunas otras clases.

Sensibilidad: Relaciona la capacidad de la predicción para detectar la enfermedad en sujeto enfermo, es decir los resultados positivos y que en lo real pertenecen a los casos positivos.

Especificidad: Relaciona la capacidad de la predicción para detectar la no enfermedad en sujeto que se encuentra enfermo, es decir los resultados negativos y que en lo real pertenecen a los casos negativos.

Curva ROC: Gráfico que nos muestra el rendimiento de un modelo de clasificación en todos los umbrales de clasificación. Esta curva es necesariamente creciente, propiedad que refleja el compromiso existente entre sensibilidad y especificidad como se muestra en la Figura 20.

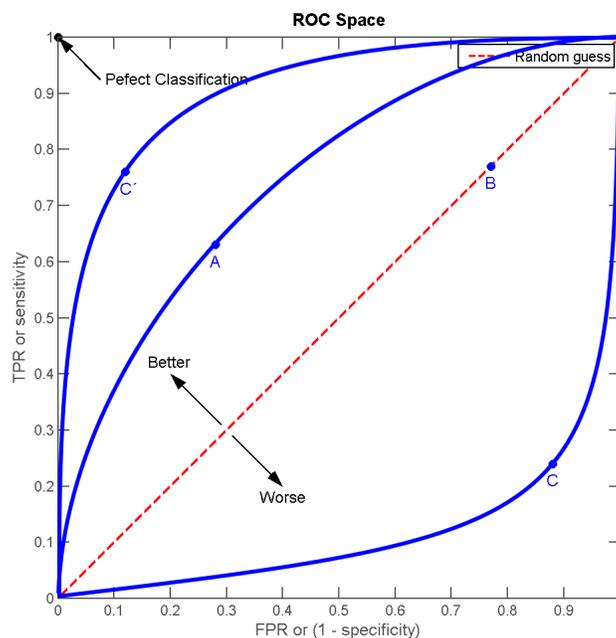


Figura 20. Curva ROC (Eric Melillanca, 2018).

4.1.8. Aplicación web

Como producto final de este estudio se desarrolla una aplicación web, en la cual se incluya el modelo que arroje mejores resultados con las características más relevantes obtenidas en el feature selection, de esta manera el personal clínico puede acceder a la plataforma y pueda hacer su análisis, además de recibir por parte de la aplicación web una respuesta acerca de qué tan viable es el tratamiento de la CPAP para el paciente.

En la **¡Error! No se encuentra el origen de la referencia.** se muestra un esquema general de la construcción de la aplicación web, la cual está conformada por el dominio, frontend, backend y hosting.

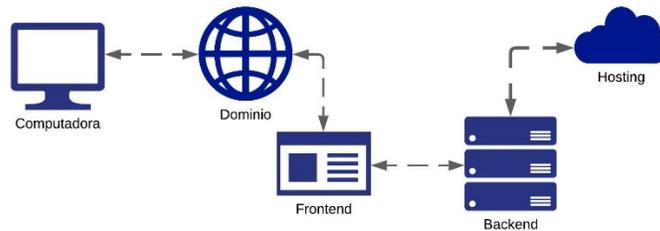


Figura 21. Diseño de la aplicación web.

Dominio: Es la dirección que permite acceder a un activo web, desde el navegador.

Frontend: es la interfaz gráfica con la que interactúa el usuario es decir, todo lo que se alcanza a percibir cuando en un sitio web.

Backend: es donde se encuentra toda la lógica de la aplicación y un cliente puede comunicarse con el servidor.

Hosting: El hosting es el sitio en la nube donde va estar almacenado el frontend y backend. Se hace uso de Vercel y Amazon Web Services.

4.2. TIPO DE ESTUDIO

La investigación para este proyecto es del tipo descriptiva con enfoque cuantitativo. Se recopiló la información de manera cuantificable con diferentes variables para ser utilizadas en el análisis de la base de datos. La población objetivo que se tomó en cuenta en el estudio son todas aquellas personas diagnosticadas con apnea obstructiva del sueño.

4.3. CRONOGRAMA

Para llevar a cabo el proyecto se determinó y elaboró una serie de entregables, actividades y tareas que conforman cada objetivo específico. Esto para tener una mejor visualización de los tiempos necesarios que se requieren en cada fase, además de tener definido el plan de trabajo.

Tabla 2. Cronograma objetivo específico 1.

 Universidad Autónoma del Caribe Proyecto de Grado - Ingeniería Mecatrónica CRONOGRAMA DE ACTIVIDADES 				
Componentes	Descripción	Fecha Inicio	Fecha Final	Duración (Días)
Planeación del Proyecto	Aplicación web para la predicción del tratamiento de pacientes con apnea obstructiva del sueño usando machine learning	24/5/2021	23/12/2021	213
OBJETIVO 1	Desarrollar el preprocesamiento de un conjunto de datos que contienen las variables clínicas a utilizar para la implementación de los algoritmos de machine learning	24/5/2021	17/9/2021	116
Entregable # 1	Resumen detallado de la base de datos original	24/5/2021	12/6/2021	19
Actividad 1	Selección de la base de datos	24/5/2021	29/5/2021	5
Tarea 1	Revisión de la base de datos	24/5/2021	28/5/2021	4
Tarea 2	importar base de datos en el software de programación	28/5/2021	29/5/2021	1
Actividad 2	Análisis exploratorio de las variables	31/5/2021	12/6/2021	12
Tarea 1	Realizar un estudio de las características de cada variable	31/5/2021	12/6/2021	12
Entregable # 2:	Informe de las características seleccionadas para implementación de los modelos	14/6/2021	17/9/2021	95
Actividad 3	Selección de características como base para prueba de modelos	14/6/2021	10/7/2021	26
Tarea 1	Clasificación de variables con matriz de coorelación	14/6/2021	30/6/2021	16
Tarea 2	Eliminación de variables altamente correlacionadas	1/7/2021	10/7/2021	9
Actividad 4	Reporte gráfico de la selección de características	12/7/2021	17/9/2021	67
Tarea 1	Identificar los outliers y datos nulos para corrección	12/7/2021	4/8/2021	23
Tarea 2	Aplicar técnicas de selección de características	5/8/2021	4/9/2021	30
Tarea 3	Realizar una tabla de los diferentes conjuntos de datos arrojados	6/9/2021	17/9/2021	11

Tabla 3. Cronograma objetivo específico 2.

OBJETIVO 2	Implementar modelos predictivos basados en aprendizaje automático supervisado y no supervisado para la detección del tratamiento de pacientes con AOS	20/9/2021	24/11/2021	65
Entregable # 3	Implementación de los modelos supervisados y no supervisados	20/9/2021	26/10/2021	36
Actividad 5	Implementación de modelos supervisados	20/9/2021	2/10/2021	12
Tarea 1	Split de los datos para entrenamiento y evaluación	20/9/2021	21/9/2021	1
Tarea 2	Implementar los modelos logístico, KNN, redes neuronales y SVM	21/9/2021	2/10/2021	11
Actividad 6	Implementación del modelo no supervisado	4/10/2021	26/10/2021	22
Tarea 1	determinar el número de clústeres ideal para el conjunto de datos	4/10/2021	6/10/2021	2
Tarea 2	Determinar variables importantes en la selección de clústeres	7/10/2021	12/10/2021	5
Tarea 3	implementar modelo K-means	13/10/2021	26/10/2021	13
Entregable # 4	Validación de los modelos supervisados y no supervisados	27/10/2021	24/11/2021	28
Actividad 7	Validación de los modelos supervisados	27/10/2021	10/11/2021	14
Tarea 1	Implementación de una matriz de confusión para cada modelo y conjunto de datos	27/10/2021	2/11/2021	6
Tarea 2	Implementación de la curva ROC	3/11/2021	9/11/2021	6
Tarea 3	Determinar el mejor modelo a implementar	9/11/2021	10/11/2021	1
Actividad 8	Validación del modelo no supervisado	11/11/2021	24/11/2021	13
Tarea 1	Resumen de la media de las variables para cada clúster	11/11/2021	17/11/2021	6
Tarea 2	Análisis de gráficos de agrupación y dendogramas	18/11/2021	24/11/2021	6

Tabla 4. Cronograma objetivo específico 3.

OBJETIVO 3	Diseñar una aplicación web para la implementación del sistema de predicción para la toma de decisiones médicas	25/11/2021	23/12/2021	28
Entregable # 5	Aplicación web funcional	25/11/2021	16/12/2021	21
Actividad 9	Implementación del modelo seleccionado en el backend de la aplicación web	25/11/2021	9/12/2021	14
Tarea1	Elaboración del código para el correcto funcionamiento del modelo	25/11/2021	4/12/2021	9
Tarea2	Validación del modelo predictivo en el backend	4/12/2021	9/12/2021	5
Actividad 10	Diseño del frontend de la aplicación web	9/12/2021	16/12/2021	7
Tarea1	Elaboración del código para el maquetado y visualización por parte del usuario	9/12/2021	15/12/2021	6
Tarea2	Pruebas para el correcto funcionamiento de la aplicación web	15/12/2021	16/12/2021	1
Entregable # 6	Manual de usuario	17/12/2021	23/12/2021	6
Actividad 11	Elaboración del documento con toda la información necesaria sobre la aplicación web	17/12/2021	23/12/2021	6
Tarea 1	Elaboración de un documento detallado acerca del uso de la aplicación	17/12/2021	23/12/2021	6

5. PRESUPUESTO

Para el cálculo y planificación de los gastos del proyecto, se construyen una serie de tablas donde se especifica y detalla los aspectos necesarios para su realización y puesta en marcha, a su vez del costo monetario que aspecto implica. De esta manera se genera un presupuesto general del proyecto.

5.1. PRESUPUESTO GENERAL

Tabla 5. Presupuesto general.

	FORMATO DE PRESUPUESTO PARA PRESENTACIÓN DE PROYECTOS DE INVESTIGACIÓN EN CONVOCATORIAS INTERNAS				GI-02-PR-03-F02
					Versión 1
					12/06/2019
PRESUPUESTO GENERAL DEL PROYECTO					
RUBROS	Fuentes de Financiamiento				Total
	Vicerrectoría de Investigaciones y transferencia	INVESTIGADORES	Otras fuentes Externas	Contrapartida UAC	
1. Personal Científico	\$ 0,00	\$ 0,00	\$ 0,00	\$ 3.733.280,00	\$ 3.733.280,0
2. Personal de Apoyo	\$ 0,00	\$ 0,00	\$ 0,00	\$ 713.920,00	\$ 713.920,0
3. Consultaría especializada y Servicios Técnicos externos	\$ 0,00	\$ 2.000.000,00	\$ 0,00	\$ 0,00	\$ 2.000.000,0
4. Materiales e Insumos	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,0
5. Trabajo de Campo	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,0
6. Equipos	\$ 0,00	\$ 600.000,00	\$ 0,00	\$ 0,00	\$ 600.000,0
7. Bibliografía	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,0
8. Material de difusión y Promoción de resultados	\$ 0,00	\$ 850.000,00	\$ 0,00	\$ 0,00	\$ 850.000,0
TOTAL, PRESUPUESTO DEL PROYECTO	\$ 0,00	\$ 3.450.000,00	\$ 4.447.200,00	\$ 0,00	\$ 7.897.200,0

5.2. PERSONAL CIENTÍFICO Y DE APOYO

Tabla 6. Costo personal científico.

1. PERSONAL CIENTÍFICO										
Nombres y Apellidos	Tipo de Contrato	Función dentro del Proyecto	Valor Hora (\$)	Dedica ción Horas/ semana	No. de Semanas	Fuentes de Financiamiento				SUB-TOTAL
						Vicerrectoría de Investigaciones y transferencia	INVESTIGADORES	Otras Fuentes Externas	Contrapartida UAC	
1. Jair Villanueva	Titular	Investigador Principal	\$46.666	3	16				\$2.239.968	\$2.239.968
2. Carlos Díaz	Asociado	Co-Investigador	\$46.666	2	16				\$1.493.312	\$1.493.312
SUB-TOTAL									\$3.733.280	\$3.733.280

Tabla 7. Costo personal de apoyo.

2. PERSONAL DE APOYO								
Nombres y Apellidos	Tipo de Vinculación	Función dentro del Proyecto	Valor Hora (\$)	Dedica ción Horas/ semana	No. de Semanas	Fuentes de Financiamiento		SUB-TOTAL
						Vicerrectoría de Investigaciones y transferencia	INVESTIGADORES	
1. Manuel Lemus	Practicante	Auxiliar de Investigación	\$2.231	10	16		\$356.960	\$356.960
2. Luis García	Practicante	Auxiliar de Investigación	\$2.231	10	16		\$356.960	\$356.960
SUB-TOTAL							\$713.920	\$713.920

5.3. CONSULTORIA ESPECIALIZADA

Tabla 8. Costo consultoría especializada.

3. CONSULTORIA ESPECIALIZADA Y SERVICIOS TECNICOS EXTERNOS				
Descripción	Justificación	Fuentes de Financiamiento		
		Vicerrectoría de Investigaciones y transferencia	INVESTIGADORES	SUB-TOTAL
1. Luis Márquez (Ing. Electrónico)	Asesoría en desarrollo web y programación		\$600.000	\$600.000
SUB-TOTAL		\$ 0	\$600.000	\$600.000

5.4. MATERIALES, INSUMOS Y EQUIPOS

Tabla 9. Costo equipos usados.

6. EQUIPOS						
Descripción	Justificación	Cantidad	Fuentes de Financiamiento			
			Vicerrectoría de Investigaciones y transferencia	INVESTIGADORES	Contrapartida UAC	SUB-TOTAL
1. Computador	Realizar procesamiento de datos, programación y creación de la aplicación web	1		\$2.000.000		\$ 2.000.000
<i>SUB-TOTAL</i>			\$ 0	\$ 2.000.000	\$ 0	\$ 2.000.000

5.5. DIFUSIÓN DE RESULTADOS

Tabla 10. Costo material de difusión.

8. MATERIAL DE DIFUSIÓN Y PROMOCIÓN DE RESULTADOS				
Descripción	Justificación	Fuentes de Financiamiento		
		Vicerrectoría de Investigaciones y transferencia	INVESTIGADORES	SUB-TOTAL
1. Artículo	Publicación de resultados en revista científica		\$850.000	
<i>SUB-TOTAL</i>		\$ 0	\$850.000	\$ 850.000

6. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

6.1. DISEÑO DE LA APLICACIÓN WEB

Backend.

La estructura del backend de la aplicación se realizó en el lenguaje de programación Python, debido a la flexibilidad que ofrece en la migración y/o implementación del modelo predictivo desde RStudio. Para la edición y escritura del código, se usa Visual Studio Code (v1.65.0, 2022, Estados Unidos), el cual es un editor de código fuente desarrollado para Windows, Linux y macOS, disponible en <https://code.visualstudio.com/>. En la Figura 22 se muestra el entorno básico de Visual Studio Code.

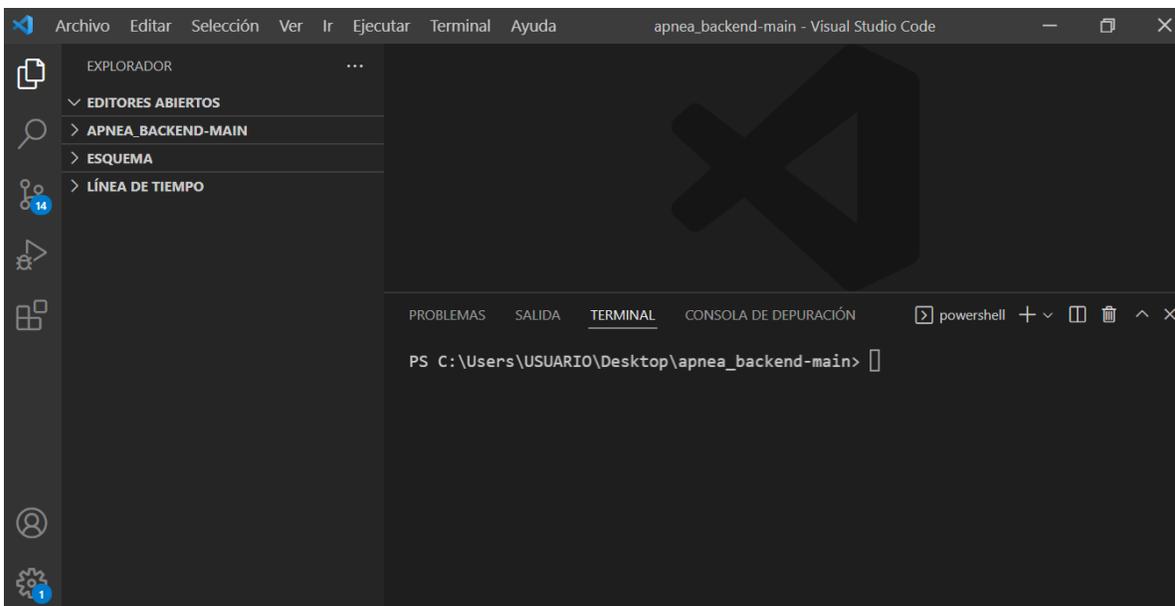


Figura 22. Visual Studio Code - Backend.

En el apartado izquierdo del editor de texto se encuentra la estructura de cada uno de los archivos y carpetas las cuales conforman el backend de la aplicación como se muestra en la Figura 23.

El entrenamiento del modelo predictivo se realiza en el lenguaje Python migrando tanto las librerías utilizadas en la implementación de los modelos, como el proceso mismo paso a paso (ver Figura 24). La predicción se realiza del mismo modo, pero se tiene en cuenta solo una observación a la vez, es decir, el paciente al cual se le desee realizar el estudio (ver Figura 25).

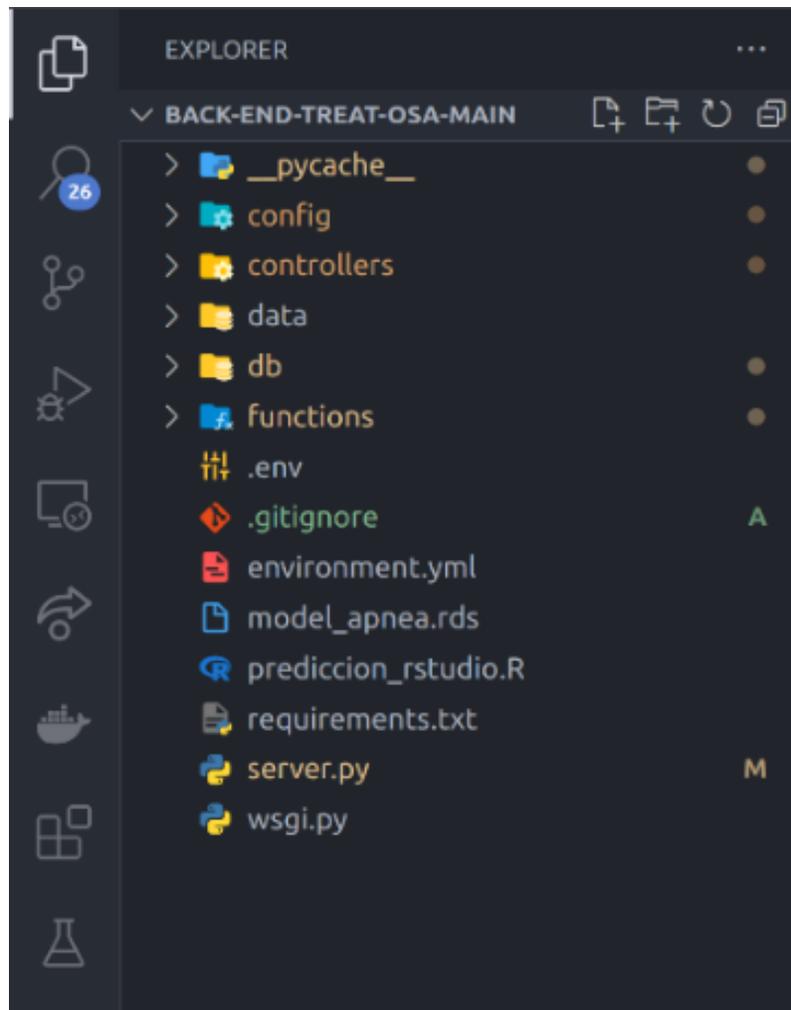


Figura 23. Archivos y carpetas backend

```
index.py 3
config > index.py > trainNgetModel
9 def trainNgetModel():
10     path = os.getcwd()
11     file = '/data/datos.xlsx'
12     filename = path + file
13     data_frame = redf.read_datos_excel_convert_to_data_frame(filename)
14     precision = 0
15     while precision <= 0.83:
16         train, test = gdt.get_test(data_frame)
17         x_test = test.drop(["outcome"], axis=1)
18         x_test_head = list(x_test.columns)
19         train_test = h2o.H2OFrame(train)
20         train_test["outcome"] = train_test["outcome"].asfactor()
21         model_predictive = H2ODeepLearningEstimator(
22             activation = 'Rectifier',
23             epochs = 100,
24             train_samples_per_iteration = -2,
25             hidden=[10,10],
26         )
27         model_predictive.train(
28             x=x_test_head,
29             y='outcome',
30             training_frame = train_test,
31         )
32         prob_test = model_predictive.predict(h2o.H2OFrame(x_test)).as_data_frame(True)
33         test_out = prob_test["predict"].tolist()
34         real_out = test["outcome"].tolist()
35         matrix = confusion_matrix(real_out, test_out)
36         precision = (matrix[0][0] + matrix[1][1]) / len(test)
37     model_path = h2o.save_model(model=model_predictive, path = os.getcwd()+"/data/models", force = True)
38     return model_path, precision
```

Figura 24. Script del modelo de predicción implementado en visual studio code.

```
predict.py 1,1 X
controllers > predict.py > predictController
6 def predictController(req):
7     db = init()
8     now = datetime.now()
9     autocpap_valida = 1
10    data_req = {
11        "talla": req["talla"],
12        "urico": req["urico"],
13        "ta_media_24": req["ta_media_24"],
14        "c10_A": req["c10_A"],
15        "hospital_num": req["hospital_num"],
16        "trigliceridos": req["trigliceridos"],
17        "ind_arousal_psg": req["ind_arousal_psg"],
18        "autocpap_valida": autocpap_valida,
19        "gr_mapa_2": req["gr_mapa_2"],
20    }
21    df = pd.DataFrame([data_req])
22    print('df', df)
23    result = predict(df)
24    print(db.collection("predictions"))
25    db.collection("predictions").document().set({
26        "prediction": result[0],
27        "userId": req["userId"],
28        "date": now,
29        "talla": req["talla"],
30        "urico": req["urico"],
31        "ta_media_24": req["ta_media_24"],
32        "c10_A": req["c10_A"],
33        "hospital_num": req["hospital_num"],
34        "trigliceridos": req["trigliceridos"],
35        "ind_arousal_psg": req["ind_arousal_psg"],
36        "autocpap_valida": autocpap_valida,
37        "gr_mapa_2": req["gr_mapa_2"],
38        "full_name_patient": req["name"],
39        "age": req["age"],
40        "weight": req["weight"],
41    })
42    return result[0]
```

Figura 25. Script de la predicción para el usuario evaluado.

Frontend.

La estructura está diseñada en los lenguajes de programación JavaScript, TypeScript, y CSS. En la Figura 26 se muestra la estructura establecida para el apartado del cliente de la aplicación, así como también, los archivos que la conforman.

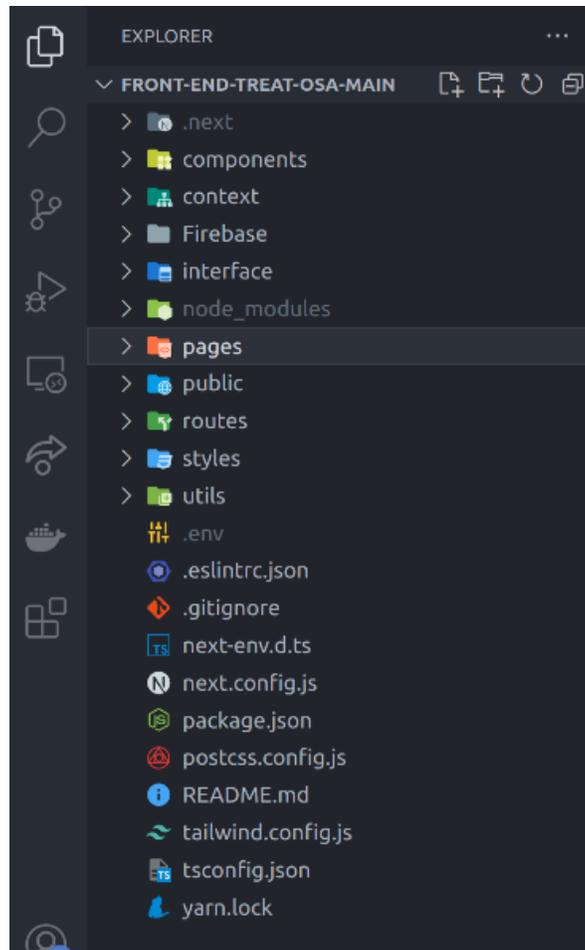


Figura 26. Archivos y carpetas del frontend.

La aplicación web cuenta con una serie de páginas básicas para una mejor experiencia de los usuarios (ver Figura 27). Las páginas previamente diseñadas son:

Index. Página de inicio en donde se encuentra información detallada del problema abordado y la aplicación.

- **Login.** Página donde se ingresan las credenciales del usuario registrado, esto con el propósito de imitar su uso para el área médica.
- **Prediction.** Página principal en donde se realiza el ingreso de los datos del paciente y predicción de la adherencia a la CPAP.
- **Profile.** Página para la edición del perfil del usuario registrado.

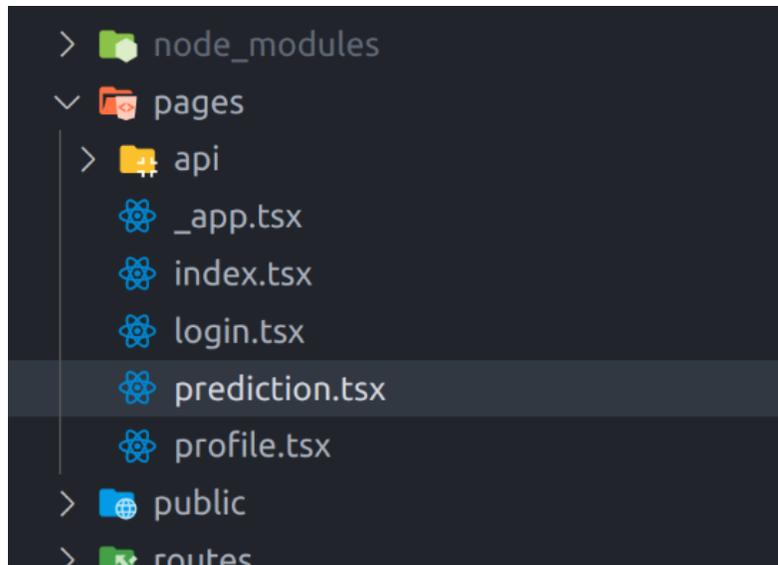


Figura 27. Páginas principales de la aplicación web.

6.2. RECOLECCIÓN DE DATOS

Se llevó a cabo una medición y registro de las características de los pacientes arrojadas por el estudio de sueño. El estudio de sueño realizó con los métodos de la polisomnografía o poligrafía respiratoria.

6.2.1. Muestra

El conjunto de datos con los que se trabajó en el proyecto fueron recopilados por una clínica en la población de España, y corresponden a 430 pacientes a los cuales se les realizó un estudio de sueño.

6.3. ANÁLISIS DE RESULTADOS

En el siguiente apartado se detallan los resultados para cada proceso realizado en la metodología. El código elaborado para este estudio se encuentra en un repositorio en GitHub, al cual se puede acceder con el enlace <https://github.com/manuel-lemus-pertuz/Processing-and-Predictive-Models.git>.

6.3.1. Preprocesamiento

Una vez aplicada la metodología para la limpieza de la base de datos se obtienen los dos principales datasets. Uno corresponde a los datos de los pacientes a quienes se les realizó la polisomnografía y el conjunto de datos correspondientes a los pacientes a los que se les realizó la poligrafía respiratoria.

En Tabla 11 se especifica el tamaño para cada dataset utilizados previo a la selección de características como base para la implementación de los modelos predictivos supervisados. Para el caso del modelo no supervisado, se omiten ciertas acciones realizadas en el procesamiento, debido a que este tipo de modelos los datos no cuentan con una variable outcome. En la Tabla 12 se especifica el tamaño para los datasets utilizados previo a la selección de características relevantes en el de los clústeres.

Tabla 11. Tamaño de los datasets (Supervisado).

DATASET	PREDICTORES	OBSERVACIONES
PSG	135	185
HRP	131	157

Tabla 12. Tamaño de los datasets (No supervisado).

DATASET	PREDICTORES	OBSERVACIONES
PSG	135	125
HRP	131	110

6.3.2. Feature Selection

En la Tabla 13 se puede analizar las variables que arrojaron cada uno de los métodos de selección de características implementados en el dataframe PSG y en la Tabla 14 las características correspondientes al dataframe HRP.

Tabla 13. Características importantes (PSG).

Métodos	Características principales relacionadas con el tratamiento de la apnea (PSG)
Forward Stepwise Selection	causas_abandono, talla, tiempo_reg_psg, ind_arousal_psg, c10_A, hospital_num, intubaciones_num. (N _{feature} =7)
Backward Stepwise Selection	talla, urico, trigliceridos, ta_media_24, autocpap_valida, gr_mapa_2, ta_sistolica_diurna_dif. (N _{feature} =7)
LASSO	ta_sistolica_diurna_dif, ta_media_24_dif, c7_bi, ta_media_diurna_dif, ta_diastolica_24_dif, c8_A, desviacion, ta_diastolica_diurna_dif, tsh, c9_A, c7_A, urico, hospital_dias, c6, c7_bi_F, gr_mapa_1, hematies, c10_A, eq50_items_dif, gr_mapa_2, c6_bi, etapa, creatinina, hospital_num, autocpap_valida. (N _{feature} =25)
Random Forest	p_rem_psg, hemoglobina, ta_diastolica_24_dif, alt, urico, ta_diastolica_nocturna_dif, ta_media_diurna_dif, ta_diastolica_diurna_dif, trigliceridos, ind_arousal_psg. (N _{feature} =10)

Tabla 14. Características importantes (HRP).

Métodos	Características principales relacionadas con el tratamiento de la apnea (HRP)
Forward Stepwise Selection	causas_abandono, quick, tad1, hora_despertar, c10_A, epworth_final, gr_mapa_1. (N _{feature} =7)
Backward Stepwise Selection	quick, tas2, tas_pico, hora_despertar, tiempo_valido_pgr, tc90_pgr, c5. (N _{feature} =7)
LASSO	desviacion, somno_conducir, c1, c6_bi_F, c2, nveces_pr, repe_pr, tipo_mascara, escala_asda, c7_A, c9_A, tsh, aih30, c6, c5, c10_A, creatinina, gr_mapa_3, gr_mapa_2, c6_bi, eq50_items_dif, repe_mapa, causas_abandono, c7_bi_F, gr_mapa_1. (N _{feature} =25)
Random Forest	ta_media_nocturna, colesterol, tas2, ta_media_24, ta_diastolica_24, sf36_fm, urico, ta_media_diurna, tad1, trigliceridos. (N _{feature} =10)

6.3.3. Modelos supervisados

Teniendo en cuenta las variables seleccionadas por los diferentes métodos de selección de características implementados en el conjunto de datos original, se tienen diferentes sub-conjuntos de datos con las variables seleccionadas. La selección de características se generó nuevos datasets, de esta manera se implementa cada uno de los modelos supervisados para los diferentes conjuntos de datos obtenidos, y así, se analiza y escogen los mejores resultados.

En la Tabla 15 y Tabla 16 muestra la evaluación del desempeño de los modelos de predicción implementado, es decir, la precisión de los modelos para cada conjunto de datos arrojados por los métodos de selección de características correspondientes al dataset PSG y HRP. El porcentaje para la precisión hace referencia a los sujetos que el modelo evaluó correctamente respecto a los sujetos totales.

Tabla 15. Modelos Supervisados dataframe PSG.

PSG		Modelos			
		Logistic Regression	KNN	Redes Neuronales	Support Vector Machine
Feature Selection	Forward Stepwise	70.91%	76.36%	81.82%	74.55%
	Backward Stepwise	72.73%	69.09%	76.36%	65.45%
	LASSO	69.09%	54.55%	76.36%	70.91%
	Random Forest	60%	61.81%	70.91%	72.73%

Tabla 16. Modelos Supervisados dataframe HRP.

HRP		Models			
		Logistic Regression	KNN	Neural Networks	Support Vector Machine
Feature Selection	Forward Stepwise	78.26%	73.71%	73.91%	73.91%
	Backward Stepwise	69.57%	69.57%	73.91%	80.43%
	LASSO	65.22%	52.17%	76.09%	71.74%
	Random Forest	58.7%	71.74%	67.39%	69.57%

Teniendo en cuenta los resultados anteriores, se puede concluir que el modelo de **redes neuronales** para el primer dataset (Forward, 7 variables) proveniente del dataframe **PSG**, fue el que mejor resultados obtuvo, con una precisión del 81,82%.

Se debe mencionar que para definir finalmente el mejor modelo, también se deben considerar aspectos como la sensibilidad y la especificidad, para esto se implementa la curva ROC en cada uno de los métodos de selección de características tanto para los datos del grupo PSG (Figura 28, Figura 29) como para el grupo HRP (Figura 30, Figura 31). La curva ROC muestra que tan exacto es un modelo en términos de sensibilidad y especificidad, en términos visuales, la curva que esté más cerca de la esquina superior izquierda será la mejor.

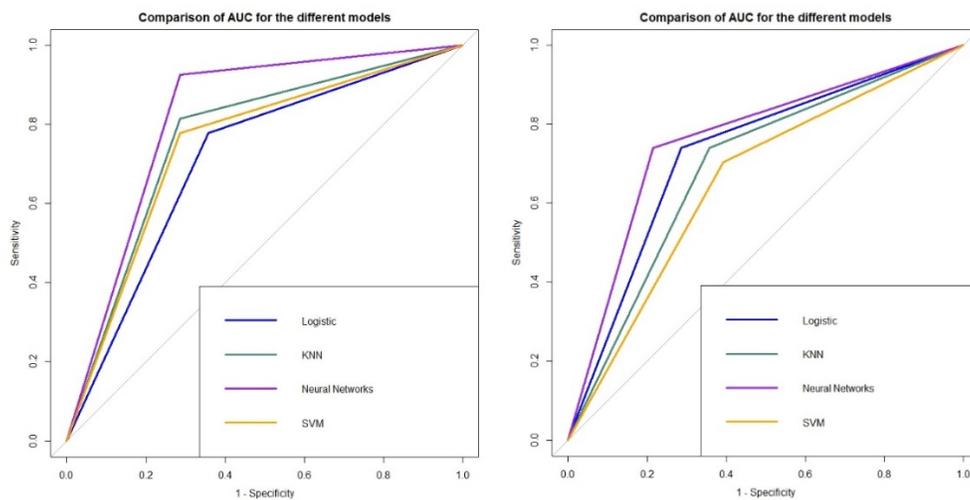


Figura 28. Curva ROC comparando los modelos (Forward - Backward) (PSG).

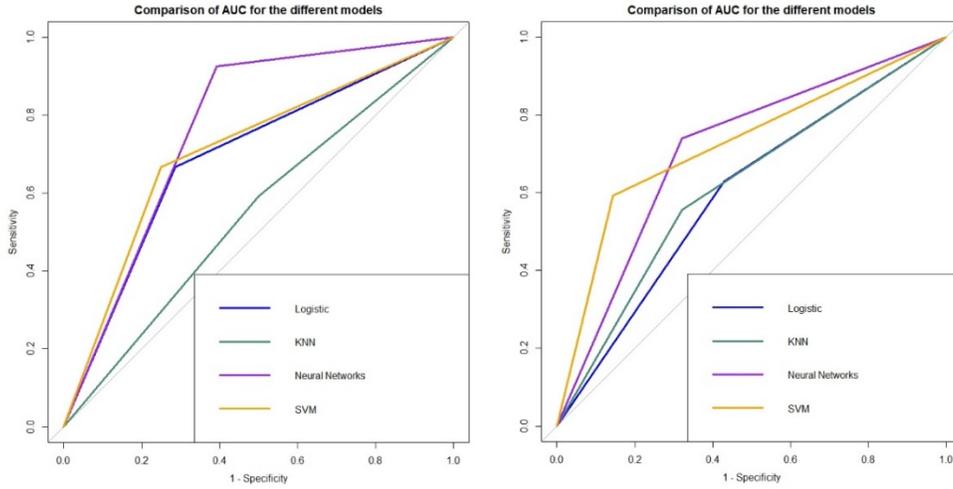


Figura 29. Curva ROC comparando los modelos (LASSO – Random Forest) (PSG).

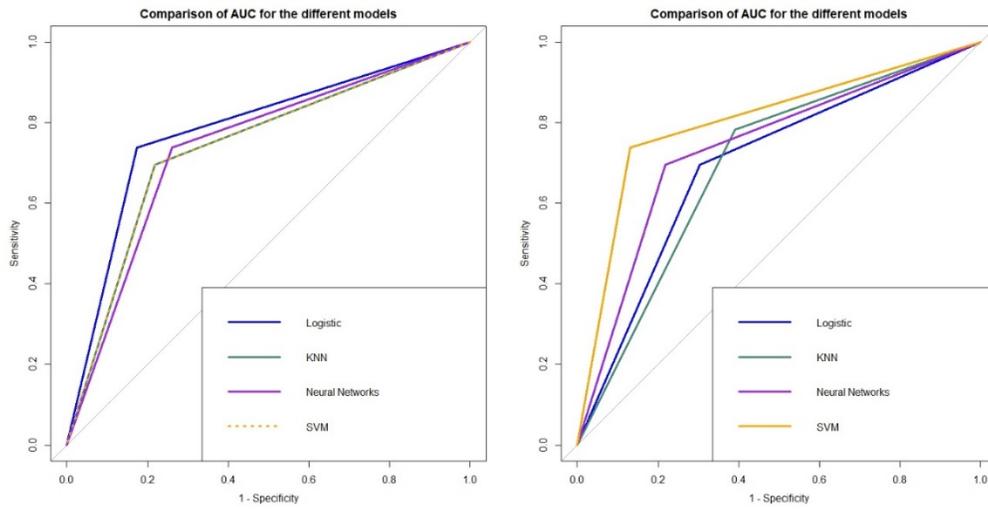


Figura 30. Curva ROC comparando los modelos (Foreward - Backward) (HRP).

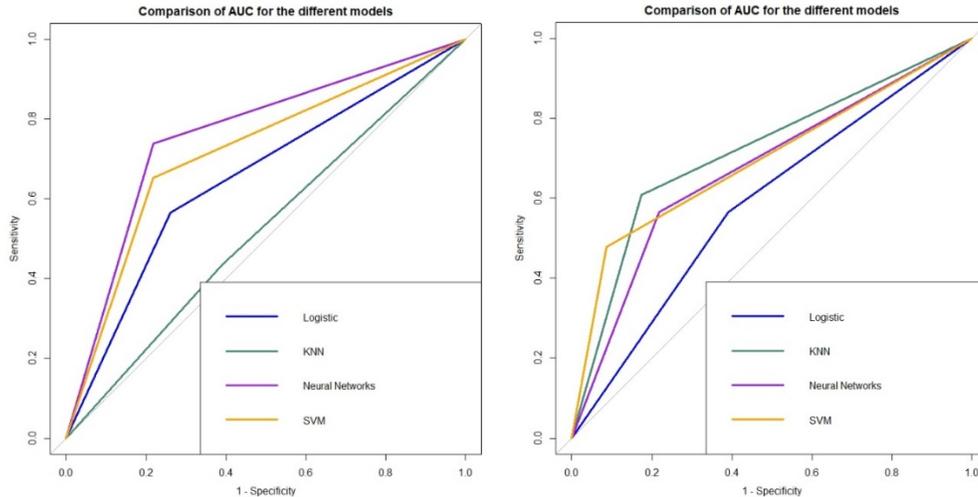


Figura 31. Curva ROC comparando los modelos (LASSO – Random Forest) (HRP).

En base a las curvas ROC obtenidas se puede tener una idea más clara en cuanto a mejor modelo se refiere, estas permiten una visualización de la especificidad y sensibilidad de cada modelo.

Se observa que el modelo que mejor resultados tuvo fue **Redes Neuronales**, con una precisión del 81.82%. Esto también se ve reflejado en una mayor área bajo la curva presente en el gráfico de la curva ROC.

6.3.4. Características indelebles

Después de realizar una evaluación de las características más importantes arrojadas por cada uno de los modelos, se creó un nuevo conjunto de datos con las variables que más se repiten en cada uno de estos (ver Tabla 17). Esto se realiza con el objetivo de generar un nuevo dataset el cual contenga variables que persisten en el estudio previamente realizado, de esta manera, se lleva a cabo un nuevo análisis, implementando de igual manera los modelos predictivos (ver Tabla 18).

Tabla 17. Predictores finales.

PREDICTORES	DESCRIPCIÓN
talla	Altura
ind_arousal_psg	Índice de arousal procedente de la polisomnografía
urico	Ácido úrico
triglicéridos	Triglicéridos
ta_media_24	Tensión arterial media durante las 24 horas del día
ta_sistolica_diurna	Tensión arterial sistólica diurna

Tabla 18. Validación de modelos en predictores finales.

MODELO	PRECISIÓN	SENSIBILIDAD	ESPECIFICIDAD	AUC
Regresión logística	78.18%	82.14%	74.07%	0.7833
KNN	56.36%	53.57%	59.26%	0.5682
Redes neuronales	83.64%	78.57%	88.89%	0.84
SVM	65.45%	46.43%	85.19	0.685

En la información arrojada por los modelos predictivos, se evidencia que el nuevo conjunto de datos logra conseguir una mejor respuesta en comparación a los métodos anteriormente mencionados con una precisión del 83,64%, de esta manera se consigue conformar el grupo de características finales para la predicción de adherencia a la CPAP.

6.3.5. Modelo no supervisado

Partiendo de los dos datasets especificados en la Tabla 12 y teniendo en cuenta que se hizo uso del método de **Clustering K-means**, se determinó el número ideal de clústeres para cada uno. En la Figura 32 se muestra el resultado obtenido por el método de la rodilla, en donde el número de clústeres ideales se determina en el punto en donde curva presenta un doble tal como el de una rodilla doblada.

En la Figura 33 se muestra método de la silueta. Para este caso, la misma gráfica indica con una línea de puntos verticales, el número de clústeres ideales para cada conjunto de datos.

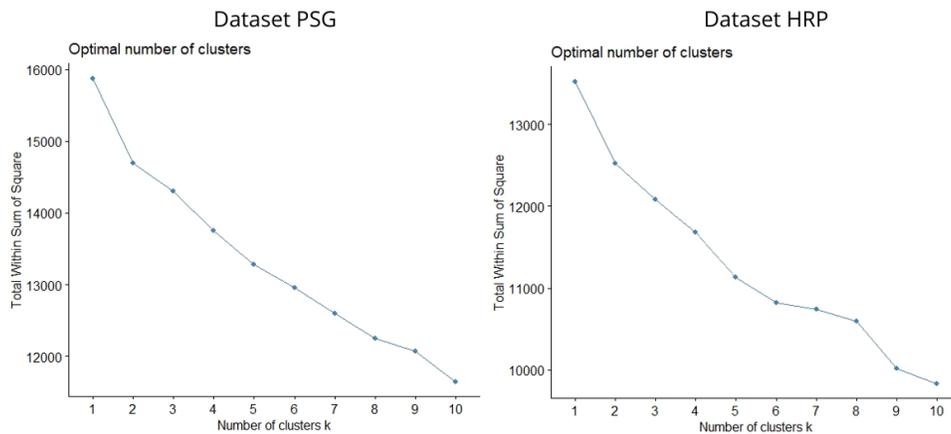


Figura 32. Número de clústeres – Método de la rodilla.

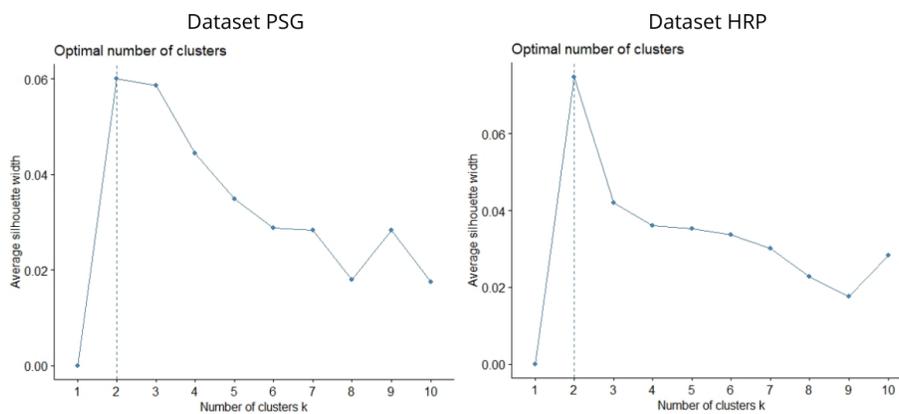


Figura 33. Número de clústeres – Método de la silueta.

Existen muchos métodos para la selección de clústeres (k) y estos muchas veces arrojan resultados diferentes. En este orden de ideas se utilizó un algoritmo en R capaz de implementar 30 métodos diferentes, de este modo, el algoritmo arroja un diagrama de barras como se muestra en la Figura 34, del número de métodos versus el número de clústeres elegidos por estos.

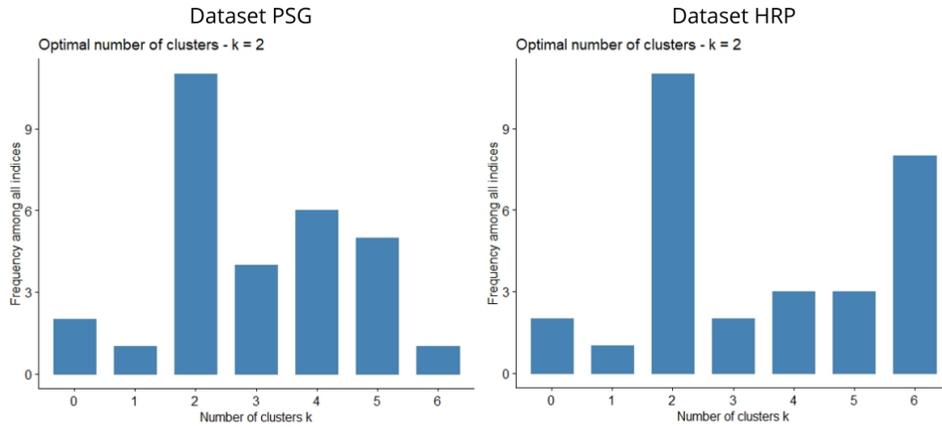


Figura 34. Número de clústeres – 30 métodos.

Sabiendo que los datasets aun presentan muchos predictores, se hace uso de la librería “FeatureImpCluster” y “flexclust”, que nos permiten realizar un análisis de los clústeres para un conjunto de datos en específico. Este análisis nos devuelve las la importancia de las variables en la elección de un número de clústeres (Tabla 19, Tabla 20), para nuestro caso el número de clústeres ideal es 2.

Tabla 19. Variables importantes - Clustering (PSG)

Importancia en la asignación de clústeres (k=2) (PSG)		
No	Predictores	Impacto (%)
1	tad1	1.28%
2	tas1	1.12%
3	ta_diastolica_diurna	1.12%
4	ta_media_24_dif	1.04%
5	ta_sistolica_24_dif	1.04%
6	ta_sistolica_diurna_dif	1.04%
7	talla	0.96%
8	talla_A	0.96%
9	gr_mapa_1	0.96%
10	ta_diastolica_diurna_dif	0.96%
11	ta_diastolica_nocturna	0.88%
12	ta_diastolica_24_dif	0.88%
13	epworth_final	0.80%
14	ta_diastolica_nocturna_dif	0.80%
15	peso	0.72%

Tabla 20. Variables importantes - Clustering (HRP)

Importancia en la asignación de clústeres (k=2) (HRP)		
No	Predictores	Impacto (%)
1	ia_pgr	1.27%
2	iah_pgr	1.09%
3	epworth	0.91%
4	epworth_A	0.91%
5	analogica	0.91%
6	epworth_inicio	0.91%
7	c6	0.63%
8	c6_bi	0.63%
9	aih30	0.63%
10	desviacion	0.54%
11	imc_nuevo	0.54%
12	imc_final	0.54%
13	tas_pico	0.45%
14	tas_valle	0.45%
15	fcard_noche	0.45%

Partiendo de las 15 variables más importantes descritas en las tablas anteriores y teniendo el número de 2 clústeres, se conformó dos nuevos datasets a los cuales se les implementó el método **K-means**.

El resultado de la implementación del modelo se muestra en la Figura 35, en donde se evidencia la distribución de cada clúster, es decir, la ubicación grafica para conjunto y la distancia entre cada observación (paciente). En la Tabla 21 y Tabla 22 se muestran las medias de las variables pertenecientes a cada clúster, esto hace referencia al promedio para cada característica en los 2 grupos seleccionados.

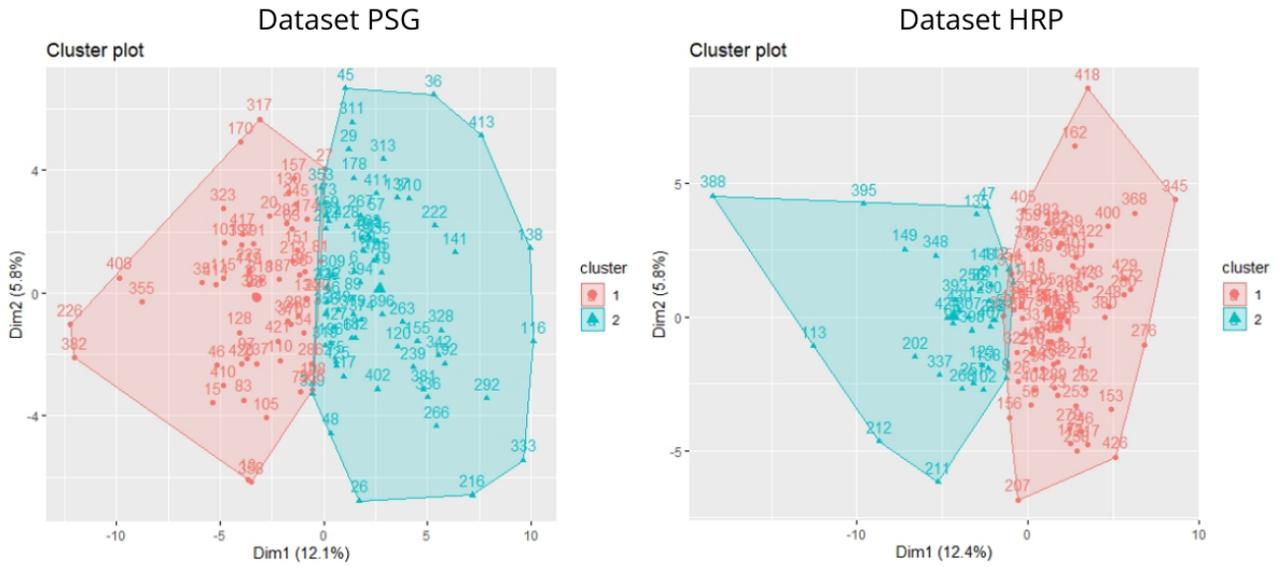


Figura 35. K-means.

Tabla 21. Promedio de las variables importantes (PSG)

Media de los predictores (PSG)			
No	Predictores	k1 (57)	k2 (68)
1	tad1	88.95	75.4
2	tas1	141	117.78
3	ta_diastolica_diurna	83.75	75.35
4	ta_media_24_dif	-4.49	1.4
5	ta_sistolica_24_dif	-5.05	2.68
6	ta_sistolica_diurna_dif	-4.37	2.16
7	talla	1.71	1.68
8	talla_A	1.71	1.68
9	gr_mapa_1	0.98	0.31
10	ta_diastolica_diurna_dif	-4.35	-0.01
11	ta_diastolica_nocturna	76.11	65.34
12	ta_diastolica_24_dif	-4.13	0.88
13	epworth_final	7.32	7.42
14	ta_diastolica_nocturna_dif	-3.45	2.14
15	peso	98.1	91.87

Tabla 22. Promedio de las variables importantes (HRP)

Media de los predictores (HRP)			
No	Predictores	k1 (76)	k2 (34)
1	ia_pgr	23.82	36.75
2	iah_pgr	39.05	58.58
3	epworth	13.51	14.36
4	epworth_A	13.54	41.47
5	analogica	57.26	59.38
6	epworth_inicio	13.66	14.47
7	c6	2.33	2.62
8	c6_bi	1.41	1.5
9	aih30	0.57	0.76
10	desviacion	9.53	12.46
11	imc_nuevo	32.06	33.96
12	imc_final	32.12	34.2
13	tas_pico	145.42	167.62
14	tas_valle	91.21	104.38
15	fcard_noche	67.72	70.82

6.3.6. Pruebas realizadas en la aplicación web

Para la dirección y uso de la aplicación web, se adquiere un dominio por parte de un tercero el cual tiene por nombre www.apneatreatmentprediction.online/ (ver Figura 36), y permite acceder a TREAT-OSA. En la Figura 37 se muestra la página en donde la persona encargada de realizar el estudio ingresa con sus credenciales y logra acceso a la página en la que se lleva a cabo la predicción.

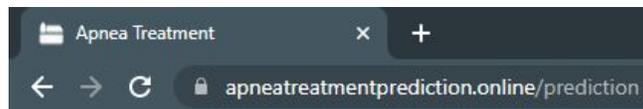


Figura 36. Dominio de TREAT-OSA.

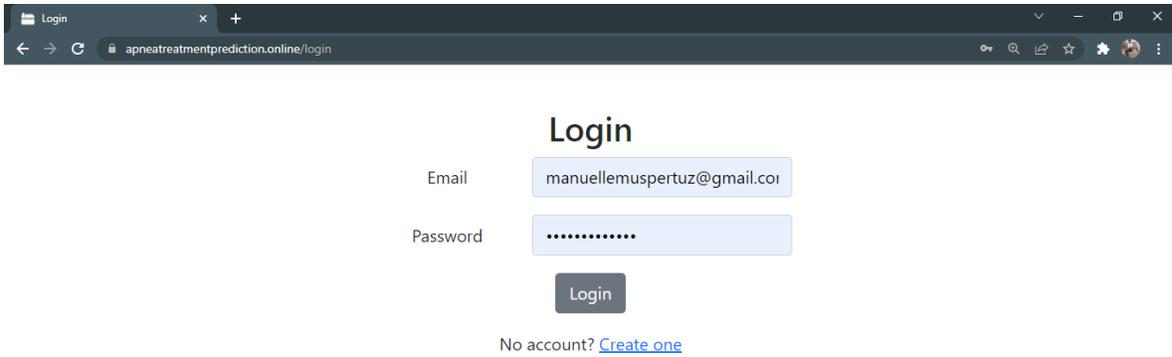


Figura 37. Página de inicio de sesión.

La aplicación cuenta con 3 secciones principales (ver Figura 38):

La sección 1 consta de la información que se requiere del paciente. En este apartado el médico o especialista del área, ingresa datos necesarios del evaluado, así como también aquellas características importantes en la implementación de los modelos predictivos.

La **sección 2** se especifica la media de cada una de las características ubicadas en el la sesión 1, provenientes del conjunto de datos de estudio previamente procesado.

La **sección 3** muestra el resultado de la predicción arrojada por la aplicación web. En este apartado el aplicativo predice si el paciente evaluado es o no es apto para el tratamiento de la CPAP, además calcula el índice de masa corporal (IMC) como información adicional e importante en el proceso.

The screenshot shows the TREAT-OSA web application interface. At the top, there is a navigation bar with 'TREAT-OSA', 'Home', 'Predecir', 'Perfil', and 'Cerrar sesión'. The user is identified as 'Manuel Lemus Pertuz'. The main content area is divided into three sections:

- Sección 1: Datos del paciente** (Patient Data): This section contains several input fields for personal, anthropometric, and laboratory data.

Categoría	Variable	Valor
Personales	Nombre completo	
	Edad (años)	50
Antropométricos	Altura (metros)	1.70
	Peso (Kg)	95
Analítico y Bioquímicos	Acido Úrico	6
	Triglicéridos	143
	Tensión arterial media (mmHg) en las 24 horas del día	83
	Tensión arterial sistólica diurna	125
Polisomnografía	Índice de arousal	43
- Sección 2: PROMEDIO DE LOS DATOS (REGISTRO BASE DEL ESTUDIO)** (Average of Data - Study Base Record): This section displays a summary table of the patient's data.

Edad	Peso	Altura	Acido Úrico
51	95.5	1.70	6
Triglicéridos	TA Media	TA Sistólica	Ind Arousal
143	92.52	125.4	42.95
- Sección 3: PREDICCIÓN CPAP** (CPAP Prediction): This section contains two buttons for prediction: 'Adherencia' and 'IMC'.

Figura 38. Sesiones de la aplicación web.

Finalmente se realizan una serie de pruebas en donde se ingresan pacientes aleatorios (ver Figura 39 y Figura 40) tomando como referencia el conjunto de datos reales arrojados por el estudio, como validación y correcto funcionamiento de TREAT-OSA.

The screenshot shows the TREAT-OSA web application interface with the same navigation bar and user profile. The patient data is now filled in, and the CPAP prediction results are displayed:

- Sección 1: Datos del paciente** (Patient Data):

Categoría	Variable	Valor
Personales	Nombre completo	Edgar Peña
	Edad (años)	46
Antropométricos	Altura (metros)	1.65
	Peso (Kg)	90
Analítico y Bioquímicos	Acido úrico	6.5
	Triglicéridos	148
	Tensión arterial media (mmHg) en las 24 horas del día	97
	Tensión arterial sistólica diurna	130
Polisomnografía	Índice de arousal	43
- PROMEDIO DE LOS DATOS (REGISTRO BASE DEL ESTUDIO)** (Average of Data - Study Base Record):

Edad	Peso	Altura	Acido Úrico
51	95.5	1.70	6
Triglicéridos	TA Media	TA Sistólica	Ind Arousal
143	92.52	125.4	42.95
- PREDICCIÓN CPAP** (CPAP Prediction):

Adherencia	IMC
Apto para CPAP	33.06 Obesidad Grado 1

Figura 39. Paciente apto para CPAP.

TREAT-OSA Home **Predcir** Perfil Cerrar sesión

Manuel Lemus Pertuz

Datos del paciente

Personales
* Nombre completo: Claudia Mejia
* Edad (años): 42

Antropométricos
* Altura (metros): 1.62
* Peso (Kg): 64

Análisis y Bioquímicos
* Acido úrico: 5.5
* Triglicéridos: 197
* Tensión arterial media (mmHg) en las 24 horas del día: 91
* Tensión arterial sistólica diurna: 118

Polisomnografía
* Índice de arousal: 15.02

PROMEDIO DE LOS DATOS (REGISTRO BASE DEL ESTUDIO)

Edad	Peso	Altura	Acido Úrico
51	95.5	1.70	6
Triglicéridos	TA Media	TA Sistólica	Ind Arousal
143	92.52	125.4	42.95

PREDICCIÓN CPAP

Adherencia No apto para CPAP	IMC 24.39 Normal
--	----------------------------

Predcir

Figura 40. Paciente no apto para CPAP.

CONCLUSIONES

El conjunto de datos original utilizado como base para el estudio presenta 335 variables de 450 pacientes sometidos a estudios. Una vez aplicados los procesos para la limpieza, se logran reducir en dos datasets, el primero con 135 variables de los pacientes sometidos a PSG y el segundo con 131 variables de los pacientes sometidos a HRP.

La selección de características corresponde a un conjunto de métodos que permiten extraer o determinar las variables más relevantes teniendo como referencia la variable a predecir. El método Forward y Backward arrojaron 7 variables, el Random Forest arrojó 10 y el LASSO hizo un top de 50 variables, de las cuales se seleccionaron las primeras 25. Indica cuales fueron las variables más relevantes que permitieron tener mayor desempeño.

Giulia Scioscia en su estudio, buscando las causas de la mala adherencia a la terapia CPAP, implementa una serie de modelos predictivos a través del método SVM (Support Vector Machine) que mostraron una precisión y sensibilidad del 68,6%, con un AUC del 72,9% [67]. Por otra parte, Silvia Sanromá en su estudio, diseñó una aplicación enfocada a la detección de la retinopatía diabética, consiguiendo una precisión del 80,8% con una sensibilidad del 80% y especificidad del 80,96% [68].

Para este estudio el modelo Redes Neuronales fue quien mejor resultados obtuvo en la mayoría de datasets con una precisión del 81,82%, sensibilidad del 71,43% y especificidad del 92,59%, pero con el análisis del nuevo conjunto de datos en donde están las características que persisten en los métodos de selección de características, se consigue mejores resultados, dando una precisión del 83,64%, en la predicción del tratamiento con una sensibilidad del 78,57%, especificidad del 88.89% y un área bajo la curva (AUC) correspondiente a la ROC de 84%.

La aplicación web diseñada e implementada se convierte en una eficiente herramienta virtual que le servirá al personal médico a tener un tratamiento oportuno con pocas variables, además que proporciona en una plataforma familiar (sitios web, aplicaciones web) el análisis e implementación de técnicas de aprendizaje automático en la predicción del tratamiento de pacientes con AOS.

BIBLIOGRAFÍA

- [1] M. K. Uçar, M. R. Bozkurt, C. Bilgin, and K. Polat, "Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques," *Neural Comput. Appl.*, vol. 29, no. 8, pp. 1–16, 2018, doi: 10.1007/s00521-016-2365-x.
- [2] F. Hajipour, M. J. Jozani, A. Elwali, and Z. Moussavi, "Regularized logistic regression for obstructive sleep apnea screening during wakefulness using daytime tracheal breathing sounds and anthropometric information," *Med. Biol. Eng. Comput.*, vol. 57, no. 12, pp. 2641–2655, 2019, doi: 10.1007/s11517-019-02052-4.
- [3] C. Mencar *et al.*, "Application of machine learning to predict obstructive sleep apnea syndrome severity," *Health Informatics J.*, vol. 26, no. 1, pp. 298–317, 2020, doi: 10.1177/1460458218824725.
- [4] R. Stretch *et al.*, "Predicting nondiagnostic home sleep apnea tests using machine learning," *J. Clin. Sleep Med.*, vol. 15, no. 11, pp. 1599–1608, 2019, doi: 10.5664/jcsm.8020.
- [5] L. Gaspar *et al.*, "Long-Term Continuous Positive Airway Pressure Treatment Ameliorates Biological Clock Disruptions in Obstructive Sleep Apnea," *SSRN Electron. J.*, vol. 65, 2020, doi: 10.2139/ssrn.3730010.
- [6] D. Álvarez *et al.*, "A machine learning-based test for adult sleep apnoea screening at home using oximetry and airflow," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-62223-4.
- [7] J. T. Maurer, "Early diagnosis of sleep related breathing disorders.," *GMS Curr. Top. Otorhinolaryngol. Head Neck Surg.*, vol. 7, p. Doc03, 2008, [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22073090><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3199834>.
- [8] D. P. White, "Pathogenesis of obstructive and central sleep apnea," *Am. J. Respir. Crit. Care Med.*, vol. 172, no. 11, pp. 1363–1370, 2005, doi: 10.1164/rccm.200412-1631SO.
- [9] T. I. Morgenthaler, V. Kagramanov, V. Hanak, and P. A. Decker, "Complex sleep apnea syndrome: Is it a unique clinical syndrome?," *Sleep*, vol. 29, no. 9, pp. 1203–1209, 2006, doi: 10.1093/sleep/29.9.1203.
- [10] T. Young, P. E. Peppard, and D. J. Gottlieb, "Epidemiology of obstructive sleep apnea: A population health perspective," *Am. J. Respir. Crit. Care Med.*, vol. 165, no. 9, pp. 1217–1239, 2002, doi: 10.1164/rccm.2109080.

- [11] J. Stradling and R. J. O. Davies, "Controversies In Sleep Medicine : Hypertension Editorial Sleep Apnea and Hypertension-What a Mess !," vol. 20, no. 9, 1997.
- [12] M. Ip, K. F. Chung, K. N. Chan, S. P. Lam, and K. Lee, "Previously unrecognized obstructive sleep apnea in Chinese subjects with essential hypertension," *Lung*, vol. 177, no. 6, pp. 391–400, 1999, doi: 10.1007/PL00007656.
- [13] A. T. Mulgrew *et al.*, "The impact of obstructive sleep apnea and daytime sleepiness on work limitation," *Sleep Med.*, vol. 9, no. 1, pp. 42–53, 2007, doi: 10.1016/j.sleep.2007.01.009.
- [14] S. R. Patel, D. P. White, A. Malhotra, M. L. Stanchina, and N. T. Ayas, "Continuous positive airway pressure therapy for treating sleepiness in a diverse population with obstructive sleep apnea results of a meta-analysis," *Arch. Intern. Med.*, vol. 163, no. 5, pp. 565–571, 2003, doi: 10.1001/archinte.163.5.565.
- [15] F. Bozkurt, M. K. Uçar, C. Bilgin, and A. Zengin, "Sleep–wake stage detection with single channel ECG and hybrid machine learning model in patients with obstructive sleep apnea," *Phys. Eng. Sci. Med.*, vol. 44, no. 1, pp. 63–77, 2021, doi: 10.1007/s13246-020-00953-5.
- [16] M. H. Miraz and M. Ali, *Blockchain enabled enhanced IoT ecosystem security*, vol. 200. 2018.
- [17] T. E. Weaver, "Novel aspects of CPAP treatment and interventions to improve cpap adherence," *J. Clin. Med.*, vol. 8, no. 12, 2019, doi: 10.3390/jcm8122220.
- [18] A. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine," *N. Engl. J. Med.*, vol. 380, no. 14, pp. 1347–1358, 2019, doi: 10.1056/nejmra1814259.
- [19] V. K. Kapur, "Obstructive sleep apnea: Diagnosis, epidemiology, and economics," *Respir. Care*, vol. 55, no. 9, pp. 1155–1167, 2010.
- [20] O. Guglielmi, A. I. S. Gómez, B. Jurado-Gómez, and G. Buéla-Casal, "Efectos del síndrome de apneas-hipopneas del sueño sobre la calidad de vida y la somnolencia diurna*," *Univ. Psychol.*, vol. 12, no. 2, pp. 601–611, 2013, doi: 10.11144/Javeriana.UPSY12-2.esah.
- [21] F. Zerah-Lancner *et al.*, "Predictive value of pulmonary function parameters for sleep apnea syndrome," *Am. J. Respir. Crit. Care Med.*, vol. 162, no. 6, pp. 2208–2212, 2000, doi: 10.1164/ajrccm.162.6.2002002.
- [22] J. H. Chen and S. M. Asch, "Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations," *N. Engl. J. Med.*, vol. 376, no. 26, pp. 2507–2509, 2017, doi: 10.1056/nejmp1702071.

- [23] N. Savage, "Better Medicine Through Machine Learning," *Commun. ACM*, vol. 55, no. 1, pp. 17–19, 2012, doi: 10.1145/2063176.2063182.
- [24] F. Cabitza, R. Rasoini, and G. F. Gensini, "Unintended consequences of machine learning in medicine," *JAMA - J. Am. Med. Assoc.*, vol. 318, no. 6, pp. 517–518, 2017, doi: 10.1001/jama.2017.7797.
- [25] M. He, X. Yan, J. Zhou, and G. Xie, "Traditional Chinese medicine database and application on the web," *J. Chem. Inf. Comput. Sci.*, vol. 41, no. 2, pp. 273–277, 2001, doi: 10.1021/ci0003101.
- [26] B. Yadav, S. Ch, S. Mathur, and J. Adamowski, "Estimation of in-situ bioremediation system cost using a hybrid Extreme Learning Machine (ELM)-particle swarm optimization approach," *J. Hydrol.*, vol. 543, pp. 373–385, 2016, doi: 10.1016/j.jhydrol.2016.10.013.
- [27] J. Zhou, C. L. Theesfeld, K. Yao, K. M. Chen, A. K. Wong, and O. G. Troyanskaya, "Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk," *Nat. Genet.*, vol. 50, no. 8, pp. 1171–1179, 2018, doi: 10.1038/s41588-018-0160-6.
- [28] B. Ustun, M. B. Westover, C. Rudin, and M. T. Bianchi, "Clinical prediction models for sleep apnea: The importance of medical history over symptoms," *J. Clin. Sleep Med.*, vol. 12, no. 2, pp. 161–168, 2016, doi: 10.5664/jcsm.5476.
- [29] G. C. Gutiérrez-Tobal, D. Álvarez, A. Crespo, F. Del Campo, and R. Hornero, "Evaluation of Machine-Learning Approaches to Estimate Sleep Apnea Severity From At-Home Oximetry Recordings," *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 2, pp. 882–892, 2019, doi: 10.1109/JBHI.2018.2823384.
- [30] "コモンマーモセットにおける 持続皮下輸液方法の技術検討 石井 一 1) ・ 中村 克樹 3) ・ 石橋 英俊 2)," pp. 2–3.
- [31] S. Tsuiki *et al.*, "Machine learning for image-based detection of patients with obstructive sleep apnea: an exploratory study," *Sleep Breath.*, 2021, doi: 10.1007/s11325-021-02301-7.
- [32] B. Adeniyi and G. E. Erhabor, "Obstructive Sleep Apnea," *West Afr. J. Med.*, vol. 37, no. 7, pp. 713–714, 2020, doi: 10.4081/ni.2011.e15.
- [33] J. S. Floras, "Sleep apnea and cardiovascular risk," *J. Cardiol.*, vol. 63, no. 1, pp. 3–8, 2014, doi: 10.1016/j.jjcc.2013.08.009.
- [34] S. Javaheri *et al.*, "Sleep Apnea: Types, Mechanisms, and Clinical Cardiovascular Consequences," *J. Am. Coll. Cardiol.*, vol. 69, no. 7, pp. 841–858, 2017, doi:

- 10.1016/j.jacc.2016.11.069.
- [35] T. D. Bradley and J. S. Floras, "Sleep apnea and heart failure: Part I: Obstructive sleep apnea," *Circulation*, vol. 107, no. 12, pp. 1671–1678, 2003, doi: 10.1161/01.CIR.0000061757.12581.15.
- [36] P. Tung *et al.*, "Obstructive and central sleep apnea and the risk of incident atrial fibrillation in a community cohort of men and women," *J. Am. Heart Assoc.*, vol. 6, no. 7, 2017, doi: 10.1161/JAHA.116.004500.
- [37] A. E. Ortíz Naretto and Et al, "Apnea central del sueño y respiración periódica," *Rev. Am. Med. Respir.*, vol. 20, pp. 150–161, 2020.
- [38] G. Díaz Santos and S. Guerrero, "Apnea central del sueño con patrón de respiración Cheynes-Stokes," *Rev. Colomb. Neumol.*, vol. 29, no. 1, p. 26, 2017, doi: 10.30789/rcneumologia.v29.n1.2017.235.
- [39] D. P. White, "Sleep apnea," *Proc. Am. Thorac. Soc.*, vol. 3, no. 1, pp. 124–128, 2006, doi: 10.1513/pats.200510-116JH.
- [40] N. M. Punjabi, "The epidemiology of adult obstructive sleep apnea," *Proc. Am. Thorac. Soc.*, vol. 5, no. 2, pp. 136–143, 2008, doi: 10.1513/pats.200709-155MG.
- [41] C. R. Laratta, N. T. Ayas, M. Povitz, and S. R. Pendharkar, "Diagnosis and treatment of obstructive sleep apnea in adults," *Cmaj*, vol. 189, no. 48, pp. E1481–E1488, 2017, doi: 10.1503/cmaj.170296.
- [42] J. V. Rundo and R. Downey, "Chapter 25 - Polysomnography," in *Clinical Neurophysiology: Basis and Technical Aspects*, vol. 160, K. H. Levin and P. Chauvel, Eds. Elsevier, 2019, pp. 381–392.
- [43] A. S. Daftary, H. E. Jalou, L. Shively, J. E. Slaven, and S. D. Davis, "Polysomnography reference values in healthy newborns," *J. Clin. Sleep Med.*, vol. 15, no. 3, pp. 437–443, 2019, doi: 10.5664/jcsm.7670.
- [44] A. Dres, E. Borsini, M. Bosio, S. Quadrelli, J. Campos, and T. Décima, "Poligrafía respiratoria en el diagnóstico de los trastornos respiratorios durante el sueño Una herramienta necesaria para el neumonólogo," pp. 152–160, 2012.
- [45] T. Gharibeh and R. Mehra, "Obstructive sleep apnea syndrome: Natural history, diagnosis, and emerging treatment options," *Nat. Sci. Sleep*, vol. 2, pp. 233–255, 2010, doi: 10.2147/NSS.S6844.
- [46] B. W. Rotenberg, D. Murariu, and K. P. Pang, "Trends in CPAP adherence over twenty years of data collection : a flattened curve," *J. Otolaryngol. - Head Neck Surg.*, pp. 1–9, 2016, doi: 10.1186/s40463-016-0156-0.

- [47] D. J. Bratton, T. Gaisl, A. M. Wons, and M. Kohler, "CPAP vs mandibular advancement devices and blood pressure in patients with obstructive sleep apnea a systematic review and meta-analysis," *JAMA - J. Am. Med. Assoc.*, vol. 314, no. 21, pp. 2280–2293, 2015, doi: 10.1001/jama.2015.16303.
- [48] T. G. Quinnell and A. L. Clutterbuck-James, "Alternatives to continuous positive airway pressure 2: Mandibular advancement devices compared," *Curr. Opin. Pulm. Med.*, vol. 20, no. 6, pp. 595–600, 2014, doi: 10.1097/MCP.000000000000096.
- [49] W. J. Randerath *et al.*, "Non-CPAP therapies in obstructive sleep apnoea," *Eur. Respir. J.*, vol. 37, no. 5, pp. 1000–1028, 2011, doi: 10.1183/09031936.00099710.
- [50] G. E. De Vries *et al.*, "Usage of positional therapy in adults with obstructive sleep apnea," *J. Clin. Sleep Med.*, vol. 11, no. 2, pp. 131–137, 2015, doi: 10.5664/jcsm.4458.
- [51] P. M. Baptista, "CirugÃ\~a como tratamiento de la apnea obstructiva del sueÃ\~o," *An. Sist. Sanit. Navar.*, vol. 30, pp. 75–88, 2007, [Online]. Available: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272007000200006&nrm=iso.
- [52] A. V Zinchuk *et al.*, "Polysomnographic phenotypes and their cardiovascular implications in obstructive sleep apnoea," *Thorax*, p. thoraxjnl-2017-210431, 2017, doi: 10.1136/thoraxjnl-2017-210431.
- [53] K. Sarkhosh, N. J. Switzer, M. El-Hadi, D. W. Birch, X. Shi, and S. Karmali, "The Impact of Bariatric Surgery on Obstructive Sleep Apnea: A Systematic Review," *Obes. Surg.*, vol. 23, no. 3, pp. 414–423, 2013, doi: 10.1007/s11695-012-0862-2.
- [54] R. Kohavi and H. John, "Artificial Intelligence Wrappers for feature subset selection," vol. 97, no. 97, pp. 273–324, 2011.
- [55] V. Kumar and S. Minz, "Feature Selection : A literature Review," vol. 4, no. 3, 2014, doi: 10.6029/smartcr.2014.03.007.
- [56] D. Hammami, T. S. Lee, T. B. M. J. Ouarda, and J. Lee, "Predictor selection for downscaling GCM data with LASSO," vol. 117, no. September, pp. 1–11, 2012, doi: 10.1029/2012JD017864.
- [57] R. J. Tibshirani, "A General Framework for Fast Stagewise Algorithms," vol. 16, pp. 2543–2588, 2015.
- [58] S. H. Shah, Y. Angel, R. Houborg, S. Ali, and M. F. McCabe, "A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat," *Remote Sens.*, vol. 11, no. 8, 2019, doi: 10.3390/rs11080963.

- [59] D. Minghui, Z. Shaopeng, and L. Ming, "Based on Machine Learning," no. Icmra, pp. 676–680, 2015.
- [60] *Machine Learning in Radiation Oncology*. 2015.
- [61] M. Awad and R. Khanna, *Machine Learning in Action: Examples*. 2015.
- [62] P. K. Kimes, Y. Liu, D. Neil Hayes, and J. S. Marron, "Statistical significance for hierarchical clustering," *Biometrics*, vol. 73, no. 3, pp. 811–821, 2017, doi: 10.1111/biom.12647.
- [63] M. R. Valarezo Pardo, J. A. Honores Tapia, A. S. Gómez Moreno, and L. F. Vines Sánchez, "COMPARACIÓN DE TENDENCIAS TECNOLÓGICAS EN APLICACIONES WEB.," *3C Technol.*, vol. 7, no. 3, 2018.
- [64] S. Kobayashi, "Open Source Software Development on Medical Domain," *Mod. Inf. Syst.*, 2012, doi: 10.5772/38117.
- [65] A. N. Randomized, C. Trial, M. I. Asensio-cruz, M. Cabello, and M. Angeles, "Primary Care Physicians Can Comprehensively Manage Patients with Sleep Apnea," vol. 198, no. 5, pp. 648–656, 2018, doi: 10.1164/rccm.201710-2061OC.
- [66] F. Sierra, "Rincón epidemiológico," vol. 18, no. 3, pp. 180–182, 2003.
- [67] G. Scioscia *et al.*, "Machine learning-based prediction of adherence to continuous positive airway pressure (CPAP) in obstructive sleep apnea (OSA).," *Inform. Health Soc. Care*, pp. 1–9, Nov. 2021, doi: 10.1080/17538157.2021.1990300.
- [68] F. D. Project, "Assessment of diabetic retinopathy risk using random forests," 2016.