

## PROYECTO FIN DE GRADO

**TÍTULO:** Diseño e implementación de un cuadro de mando basado en inteligencia colectiva para la mejora de entornos de aprendizaje en el Campus Sur.

**AUTOR/A:** Mariana Vélez Torres

**TITULACIÓN:** Grado en Ingeniería de Sistemas de Telecomunicación

**DIRECTOR/A:** Sonia Herrero Luna

**TUTOR/A:** Margarita Martínez Núñez

**DEPARTAMENTO:** Ingeniería de Organización, Administración de Empresas y Estadística.

VºBº TUTOR/A

**Miembros del Tribunal Calificador:**

**PRESIDENTE/A:** Santiago Higuera de Frutos

**TUTOR/A:** Margarita Martínez Núñez

**SECRETARIO/A:** Margarita Millán Valenzuela

**Fecha de lectura:**

**Calificación:**

El Secretario/La Secretaria,



---

## Resumen

En entornos educativos, la inteligencia colectiva ayuda a comprender cómo la colaboración y el intercambio de conocimientos pueden influir en el aprendizaje de los estudiantes. Existen numerosos estudios que demuestran cómo el trabajo colaborativo mejora el aprendizaje al fomentar la participación de los estudiantes, promover una comprensión más profunda de los temas y aumentar la motivación intrínseca.

El objetivo de este proyecto, titulado “Diseño e implementación de un cuadro de mando basado en inteligencia colectiva para la mejora de entornos de aprendizaje en el Campus Sur”, es descubrir cómo las condiciones ambientales y las percepciones de los estudiantes pueden influir en su experiencia en el aula para mejorar su bienestar. Los objetivos específicos incluyen identificar las variables ambientales que mejoran el entorno educativo y la experiencia universitaria de los alumnos, analizar los factores que influyen en el bienestar de los estudiantes para promover su motivación y rendimiento académico, y contribuir al conocimiento para establecer premisas para futuros diseños de aulas universitarias más cómodas, agradables y sostenibles.

Desde el punto de vista tecnológico, el proyecto utiliza datos recolectados durante dos meses del invierno por tres sensores ubicados en un aula de la Universidad Politécnica de Madrid sobre la humedad, temperatura, luminosidad, ruido y CO<sub>2</sub>. A su vez, se realizaron encuestas a los estudiantes sobre su percepción del aula, de su estado anímico y de su estado físico durante las clases.

En cuanto al desarrollo técnico del proyecto, se limpiaron y se prepararon los datos de los sensores y las encuestas, para después diseñar ocho modelos finales de aprendizaje automático en los que se predecían diversas percepciones y estados de los estudiantes. Una vez se obtuvieron los resultados de los modelos de predicción, se diseñó un cuadro de mando en Power BI en el cual se pueden visualizar los resultados más importantes de cada modelo y la evolución y el promedio de las condiciones ambientales del aula estudiada.

Los resultados obtenidos muestran que la percepción de un ambiente cargado y una temperatura alta tiene un impacto positivo en la percepción general del aula. Sin embargo, el resultado más relevante de la investigación es que la humedad influye directamente en el bienestar emocional de los alumnos, haciendo que estén más contentos cuando la humedad es alta y más agobiados cuando ésta es baja.

Estos descubrimientos ayudan a promover prácticas más sostenibles en las aulas universitarias como una adecuada ventilación, una mejora en la regularización de la temperatura de las calefacciones durante el invierno y la adición de humidificadores. Con esto se pretende ayudar a optimizar los recursos energéticos de la universidad, reduciendo también costos económicos. Además, se espera crear un ambiente más confortable para el aprendizaje, con el fin de mejorar la experiencia universitaria a los estudiantes y su rendimiento académico.

---

---



---

## Abstract

In educational environments, collective intelligence aids the understanding of how collaboration and knowledge sharing can influence student learning. Numerous studies have demonstrated how collaborative work enhances learning by encouraging student engagement, promoting deeper understanding of academic topics, and increasing intrinsic motivation.

The goal of the Project, entitled "The design and implementation of a dashboard based on collective intelligence for the improvement of learning environments at Campus Sur", is to discover how environmental conditions and student perceptions can influence their classroom experience to improve their well-being. Specific objectives include identifying environmental variables that improve the educational environment and students' university experience, analyzing factors that influence students' well-being to enhance their motivation and academic performance, and contributing to knowledge to establish the premises for the future design of more comfortable, pleasant and sustainable university classrooms.

The project uses data collected over two winter months by three sensors, located in a classroom of the Polytechnic University of Madrid, on humidity, temperature, luminosity, noise and CO2. During this period, students were surveyed about their perception of the classroom, their mood and their physical condition during classes.

In terms of the development of the project, data collection from the sensors and surveys was followed by a process of cleaning and preparation of the data. Following this, eight final machine learning models were designed in order to predict various student perceptions and states. Once the results of the prediction models were obtained, a dashboard was designed in Power BI to visualize the most important results of each model, as well as the evolution and average of the environmental conditions of the studied classroom.

The results obtained show that the perception of a stuffy environment and a high temperature positively impacts the overall perception of the classroom. However, the most relevant result of the research is that humidity directly influences the emotional well-being of students, making them happier when humidity is high and more overwhelmed when humidity is low.

These findings help to promote more sustainable practices in university classrooms, such as adequate ventilation, improved temperature control of heating systems during the winter, and the addition of humidifiers. This is intended to help optimize the university's energy resources, while also reducing economic costs. In addition, the findings demonstrate how a more optimal learning environment can improve the students' university experience as well as their academic performance.

---



---

## Índice de figuras y ecuaciones

Figura 1: Ejemplo Cuadro de mando en Power BI [27] .....	13
Figura 2: Sensor SCD30 [28] .....	15
Figura 3: Sensor de luz ambiental VEML7700 [29] .....	15
Figura 4: Micrófono Condensador Electret KECG2240PBJ [30] .....	16
Figura 5: Árbol de Decisión .....	22
Figura 6: Diagrama algoritmo Random Forest [37] .....	24
Figura 7: Ejemplo de funciones de ajuste de hiperparámetros [39] .....	25
Ecuación 1: Fórmula R-Cuadrado o Coeficiente de Determinación .....	26
Ecuación 2: Fórmula MAE o Error Absoluto Medio .....	26
Ecuación 3: Fórmula RMSE o Raíz del Error Cuadrático Medio .....	27
Figura 8: Proceso de K-Folds Cross Validation [44] .....	28
Figura 9: Ejemplo Gráfico de Residuos .....	29
Figura 10: Ejemplo de Gráfico de Barras Random Forest [48] .....	31
Figura 11: Ejemplo de Gráfico SHAP Resumen [47] .....	31
Figura 12: Ejemplo de Gráfico de Dependencia [47] .....	32
Figura 13: Dataset de los Sensores .....	34
Figura 14: Dataset Final .....	35
Figura 15: Ejemplo Dataset Modelo 1 para Power BI .....	41
Figura 16: Menú Principal Power BI .....	42
Figura 17: Panel de Visualizaciones Power BI .....	42
Figura 18: Ejemplo Gráfico de Líneas Power BI .....	43
Figura 19: Ejemplo Tarjeta Power BI .....	43
Figura 20: Ejemplo Segmentación de Datos en Power BI .....	44
Figura 21: Ejemplo Gráfico de Barras Agrupadas en Power BI .....	44
Figura 22: Gráfico de residuos Modelo 1 con estados físicos .....	51
Figura 23: Importancia de Características Modelo 1 con estados físicos .....	52
Figura 24: Gráfica Características SHAP modelo 1 con estados físicos .....	53
Figura 25: Características RandomForest Modelo 1 sin estados Físicos .....	55
Figura 26: Características SHAP modelo 1 sin estados físicos .....	56
Figura 27: Gráfica de Residuos Modelo 2 con estados físicos .....	58
Figura 28: Gráfica Características SHAP modelo 2 con estados físicos .....	59
Figura 29: Características RandomForest Modelo 2 con estados Físicos .....	60
Figura 30: Gráfica Características SHAP modelo 2 sin estados físicos .....	61
Figura 31: Características RandomForest Modelo 2 con estados Físicos .....	62
Figura 32: Gráfica Comparativa Modelo 3 .....	64
Figura 33: Gráfica Características SHAP estado Contento .....	65
Figura 34: Gráfica Características SHAP estado Motivado .....	66
Figura 35: Gráfica Comparativa Modelo 4 .....	68
Figura 36: Gráfica Características SHAP estado Agobiado .....	69
Figura 37: Gráfica Características SHAP estado Estresado .....	70

---

Figura 38: Panel de Control Power BI .....	71
Figura 39: Resumen Sensores .....	73
Figura 40: Visualización Confort Ambiental .....	74
Figura 41: Objetivos de desarrollo sostenible .....	78

---

## Índice de Tablas

Tabla 1: Resultados Prueba 1 Modelo 1 con estados físicos .....	48
Tabla 2: Resultados Prueba 2 Modelo 1 con estados físicos .....	48
Tabla 3: Resultados Prueba 3 Modelo 1 con estados físicos .....	50
Tabla 4: Resultados prueba 1 Modelo 1 sin estados físicos .....	54
Tabla 5: Resultados prueba 2 Modelo 1 sin estados físicos .....	55
Tabla 6: Resultados Modelo 2 con estados físicos.....	57
Tabla 7: Resultados Modelo 2 sin estados físicos .....	61
Tabla 8: Resultados Modelo 3.....	63
Tabla 9: Resultados Modelo 4.....	68
Tabla 10: Criterios Ambientales.....	72
Tabla 11: Evaluación Tiempo Utilizado.....	75
Tabla 12: Tabla de Presupuestos .....	76

---

---

## Lista de acrónimos

**BPIE:** Building Performance Institute Europe (Instituto Europeo de Rendimiento de Edificios)

**CO<sub>2</sub>:** Dióxido de Carbono

**CSV:** Comma-Separated Values (Valores Separados por Comas)

**ENAE:** Escuela de Negocios y Administración de Empresas

**ETISIS:** Escuela Técnica Superior de Ingeniería de Sistemas Informáticos

**ETL:** Extract, Transform, Load (Extracción, Transformación y Carga)

**KPI:** Key Performance Indicator (Indicador Clave de Rendimiento)

**MAE:** Mean Absolute Error (Error Medio Absoluto)

**MDI:** Mean Decrease in Impurity (Disminución Media de Impureza)

**OOB:** Out-of-Bag (Fuera de la Bolsa)

**ODS:** Objetivos de Desarrollo Sostenible

**RMSE:** Root Mean Squared Error (Error Cuadrático Medio)

**SIoTCOM:** Sistemas IoT Conscientes de la Sostenibilidad y Dirigidos por Comunidades Sociales

**SQL:** Structured Query Language (Lenguaje de Consulta Estructurada)

**UPM:** Universidad Politécnica de Madrid

---

# Índice de contenidos

<b>Resumen.....</b>	<b>i</b>
<b>Abstract .....</b>	<b>v</b>
<b>Índice de figuras y ecuaciones .....</b>	<b>vii</b>
<b>Índice de Tablas .....</b>	<b>ix</b>
<b>Lista de acrónimos .....</b>	<b>xi</b>
<b>1. Introducción.....</b>	<b>1</b>
1.1 Marco y motivación del proyecto .....	1
1.2 Objetivos técnicos y académicos .....	2
1.3 Estructura de la memoria.....	3
<b>2. Marco tecnológico.....</b>	<b>5</b>
2.1 Antecedentes de la Inteligencia Colectiva .....	5
2.2 Calidad del Ambiente .....	7
2.3 Sensores .....	8
2.3.1 Utilidad de los Sensores.....	8
2.4 Herramientas de visualización para la toma de decisiones.....	9
2.4.1 Herramientas.....	10
2.4.2 Descripción de la herramienta escogida .....	11
<b>3. Especificaciones y restricciones de diseño.....</b>	<b>15</b>
3.1 Especificaciones .....	15
3.1.1 Especificaciones de Sensores.....	15
3.1.2 Especificaciones de Análisis.....	18
3.2 Restricciones .....	18
<b>4. Metodología .....</b>	<b>19</b>
4.1 Recopilación de datos .....	19
4.1.1 Datos de Sensores Ambientales .....	19
4.1.2 Datos de los Cuestionarios .....	19
4.2 Algoritmos y Metodologías Utilizadas .....	22
4.2.1 Algoritmos de Aprendizaje Automático .....	22
4.2.2 Evaluación del Modelo .....	26
4.2.3 Importancia de las Características .....	30
<b>5. Descripción de la solución propuesta.....</b>	<b>33</b>
5.1 Preprocesamiento y limpieza de los datos .....	33
5.1.1 Datos de los Sensores.....	33
5.1.2 Datos de los Cuestionarios .....	34
5.1.3 Agrupación de los datos.....	35
5.1.4 Preparación de los datos para el entrenamiento de los modelos.....	36
5.2 Modelos de Predicción.....	37
5.2.1 Selección de las variables objetivo.....	37
5.2.2 Modelos de predicción probados en la experimentación .....	38
5.3 Cuadro De Mando Power BI.....	41

---

5.3.1 Desarrollo del cuadro de mando.....	41
<b>6. Resultados .....</b>	<b>47</b>
6.1 Pruebas Iniciales .....	47
6.2 Modelo 1: Confort Ambiental.....	47
6.2.1 Modelo 1 con estados físicos .....	47
6.2.2 Modelo 1 sin estados físicos .....	53
6.3 Modelo 2: Percepción de la docencia .....	57
6.3.1 Modelo 2 con estados físicos .....	57
6.3.2 Modelo 2 sin estados físicos .....	60
6.4 Modelo 3: Estados Anímicos Positivos .....	62
6.4.1 Contento .....	64
6.4.2 Motivado.....	65
6.4.3 Pruebas adicionales.....	66
6.5 Modelo 4: Estados Anímicos Negativos .....	67
6.5.1 Agobiado.....	69
6.5.2 Estresado .....	70
6.5.2 Pruebas Adicionales .....	71
6.6 Implementación Cuadro de Mando en Power BI .....	71
6.6.1 Panel de Control .....	71
6.6.2 Sensores.....	72
6.6.3 Resultados de los modelos .....	73
6.6.4 Pruebas Adicionales .....	74
<b>7. Presupuesto .....</b>	<b>75</b>
<b>8. Impacto del proyecto.....</b>	<b>77</b>
<b>9. Conclusiones .....</b>	<b>81</b>
9.1 Conclusiones .....	81
9.2 Trabajos futuros.....	82
<b>10. Referencias .....</b>	<b>85</b>

---



---

# 1. Introducción

## 1.1 Marco y motivación del proyecto

Actualmente nos encontramos en un mundo en constante evolución, en el que optimizar recursos y mejorar la experiencia educativa se ha convertido en un desafío fundamental para las instituciones académicas.

El entorno universitario es un espacio dinámico donde la calidad del ambiente desempeña un papel importante en el éxito de los estudiantes y del profesorado. Un ambiente de aprendizaje óptimo está influenciado en gran medida por una buena iluminación, una temperatura agradable, una ventilación adecuada etc. Estos factores ambientales permiten mejorar el rendimiento académico y la concentración de los alumnos, al mismo tiempo que promueven la participación, el trabajo en equipo y aumentan la asistencia de los estudiantes a las clases.

Hoy en día, no existen indicadores específicos que aborden de manera simultánea la sostenibilidad y el rendimiento en el entorno de las aulas universitarias. Por ello, en este proyecto se pretende diseñar un indicador que integre datos de temperatura, luminosidad, humedad, CO<sub>2</sub> y ruido, capturados por tres sensores ubicados en una de las aulas de la Universidad Politécnica de Madrid, junto con las respuestas de unas encuestas que se realizarán a los estudiantes del aula mencionada, en las cuales expresarán sus percepciones y sensaciones con respecto al entorno.

La toma de decisiones del proyecto se llevará a cabo mediante la inteligencia colectiva, que es la capacidad intelectual que se manifiesta producto de la colaboración de varios individuos o comunidad, para abordar un problema en común [1]. Existen numerosos estudios que destacan la importancia de fomentar la interacción social y el trabajo colaborativo en entornos educativos, como estrategias efectivas para impulsar el aprendizaje y mejorar los resultados.

Gracias a la realización de este proyecto y a la comprensión de los datos proporcionados por los sensores ambientales y las percepciones de los estudiantes, se pretende encontrar qué variables ambientales pueden mejorar el entorno educativo y la experiencia universitaria de los alumnos. En definitiva, se explorará cómo la interacción social, complementada con el uso estratégico de la tecnología, puede realzar la efectividad académica y optimizar la toma de decisiones en ambientes educativos específicos, siendo en nuestro caso la universidad.

Una vez analizados los resultados obtenidos, se podrán establecer premisas para mejorar futuros diseños de aulas universitarias, creando espacios más cómodos, agradables y sostenibles.

Este trabajo de fin de grado se ha llevado a cabo gracias al apoyo del Proyecto del Plan Nacional IoT/COM (Sistemas IoT Conscientes de la Sostenibilidad y Dirigidos por Comunidades Sociales), ya que ha proporcionado la infraestructura y los recursos tecnológicos para la recolección de datos ambientales por los sensores y el análisis de las percepciones de los estudiantes.

---

## 1.2 Objetivos técnicos y académicos

Los objetivos de este proyecto fin de carrera son, desde el punto de vista técnico:

### Objetivo principal:

- Descubrir cómo las condiciones ambientales y las percepciones de los estudiantes pueden influir en su experiencia en el aula para mejorar su bienestar.

Para ello, se desarrollan diferentes objetivos específicos:

- **Identificar variables ambientales clave:** Encontrar qué variables ambientales mejoran el entorno educativo y la experiencia universitaria de los alumnos.
- **Analizar el impacto en el bienestar estudiantil:** Investigar los factores que afectan el bienestar de los estudiantes para promover su motivación y rendimiento académico.
- **Contribuir al conocimiento para establecer premisas para mejorar futuros diseños de aulas universitarias:** Proponer mejoras en el diseño de aulas para crear espacios más cómodos, agradables y sostenibles.

Desde el punto de vista académico, se adquieren las siguientes competencias y habilidades:

- **Adquisición de competencias en análisis de datos:** Desarrollar habilidades en análisis y procesamiento de datos.
- **Desarrollo de competencias de aprendizaje de machine learning:** Implementar y validar modelos predictivos utilizando técnicas de machine learning, específicamente usando el algoritmo RandomForest.
- **Competencias en visualización de datos:** Crear visualizaciones interactivas y un cuadro de mando en Power BI para facilitar la interpretación de los datos.
- **Autonomía en el desarrollo de proyectos:** Desarrollar la capacidad de trabajar de manera independiente, gestionando y ejecutando las diferentes fases del proyecto.

---

### 1.3 Estructura de la memoria

Para poder lograr los objetivos planteados, se van a desarrollar los siguientes apartados:

1. **Marco Tecnológico:** En este capítulo se proporciona una revisión de la literatura existente y el contexto tecnológico del proyecto. Se tratan los antecedentes de la inteligencia colectiva, la calidad ambiental, la utilidad de los sensores y las herramientas de visualización para la toma de decisiones.
2. **Especificaciones y Restricciones de Diseño:** Se detallan las especificaciones técnicas y las restricciones del proyecto. Se describen las características más técnicas de los sensores utilizados y las limitaciones encontradas durante el desarrollo del proyecto.
3. **Metodología:** Se añade este capítulo para poder definir más detalladamente la metodología utilizada para el desarrollo del proyecto. Se describe el proceso de recolección de datos, los algoritmos de aprendizaje automático empleados y las métricas de evaluación necesarias para este tipo de modelos de predicción.
4. **Descripción de la solución propuesta:** Se explica minuciosamente la solución propuesta con los pasos seguidos hasta su finalización, incluyendo el preprocesamiento y la limpieza de los datos, la creación de los modelos de predicción y el diseño del cuadro de mando en Power BI.
5. **Resultados:** Se presentan los resultados obtenidos de los diferentes modelos de predicción. Se incluyen las pruebas realizadas en cada modelo y la implementación de los resultados en el cuadro de mando de Power BI.
6. **Presupuesto:** Se proporciona una estimación detallada de los costos asociados con el desarrollo e implementación del proyecto.
7. **Impacto del Proyecto:** Se analizan las implicaciones de salud y seguridad, económicas, energéticas y tecnológicas del proyecto. También se describe la contribución del proyecto a los Objetivos de Desarrollo Sostenible (ODS).
8. **Conclusiones:** Se resumen los resultados más relevantes de la investigación mostrando el impacto y la eficacia de la solución propuesta. También se describe como estas conclusiones dan lugar a posibles proyectos futuros relacionados con este estudio.
9. **Referencias:** Se incluye una lista con las referencias bibliográficas utilizadas durante el proceso de investigación y desarrollo del proyecto.



---

## 2. Marco tecnológico

### 2.1 Antecedentes de la Inteligencia Colectiva

La inteligencia colectiva es la capacidad intelectual que se manifiesta producto de la colaboración de varios individuos o comunidad, para abordar un problema en común [1]. Según Williams y Malone, este último creador y director del MIT Center for Collective Intelligence, la inteligencia colectiva no es simplemente la suma de varias inteligencias individuales, sino de cómo esas inteligencias individuales interactúan y se combinan de manera sinérgica para lograr resultados que van más allá de lo que cada individuo podría lograr por sí solo [2] [3].

En el ámbito educativo, aunque el problema o resultado final es individual (como la nota de un examen), el entorno de aprendizaje se beneficia de la interacción y la colaboración entre los estudiantes, lo que conduce a mejores resultados para todos. De manera general, se observa que un grupo con un desempeño deficiente tiende a obtener calificaciones más bajas en comparación con un grupo con buen desempeño. Los profesores son conscientes de esta premisa mucho antes de conocer las calificaciones finales, ya que durante las clases observan si el grupo colabora, coopera, participa, y tiene disposición para aprender. Por otra parte, las condiciones ambientales (como el ruido o la temperatura que haya en el aula) también influyen en la predisposición de los estudiantes para colaborar y participar activamente.

Según el filósofo Perre Lévy [4], la inteligencia colectiva es “una inteligencia repartida en todas partes, valorizada constantemente, coordinada en tiempo real, que conduce a una movilización efectiva de las competencias”. En el contexto de Lévy, la inteligencia colectiva se refiere a la capacidad de las comunidades humanas para evolucionar hacia una mayor capacidad de cooperación e intercambio de conocimientos. Esto se logra a través de la interconexión de individuos mediante tecnologías digitales que permiten la conexión instantánea entre personas de diferentes ubicaciones geográficas, eliminando barreras de tiempo y espacio. Estas tecnologías proporcionan plataformas y herramientas que fomentan la colaboración en tiempo real, la coedición de documentos, la comunicación instantánea y la creación conjunta de contenido, lo que amplía en gran medida las posibilidades de interacción y trabajo conjunto entre individuos y comunidades a nivel global.

Esta accesibilidad universal a la información y la comunicación transforma la forma en que las personas colaboran, aprenden y crean conocimiento, potenciando así la inteligencia colectiva de la sociedad en su conjunto.

En entornos educativos, el concepto de inteligencia colectiva ha sido objeto de estudio para comprender cómo la colaboración y el intercambio de conocimientos pueden influir en el aprendizaje de los estudiantes. El trabajo colaborativo promueve una comprensión más profunda de los conceptos al permitir que los estudiantes discutan ideas, compartan perspectivas y resuelvan problemas juntos. Además, se enfatiza que la diversidad de

---

habilidades y experiencias en un grupo de estudiantes enriquece el aprendizaje, ofreciendo diferentes enfoques para abordar problemas académicos [5].

Robert Slavin [6] presenta evidencia de estudios que demuestran cómo el trabajo colaborativo mejora el aprendizaje al fomentar la participación de los estudiantes, promover una comprensión más profunda de los temas y aumentar la motivación intrínseca. Asimismo, destaca cómo la diversidad de perspectivas y habilidades en grupos cooperativos facilita el intercambio de ideas y la resolución creativa de problemas.

Numerosos investigadores han destacado la importancia de la interacción social en el aprendizaje colaborativo, subrayando cómo el intercambio de conocimientos y la colaboración entre estudiantes pueden influir positivamente en su proceso de aprendizaje [7]. Por otra parte, teóricos como Wenger [8] demostraron que las comunidades de aprendizaje son esenciales en la construcción colectiva de conocimiento, promoviendo la participación y la colectividad en la resolución de problemas.

En conjunto, estas investigaciones y teorías destacan la importancia de fomentar la interacción social y el trabajo colaborativo en entornos educativos como estrategias efectivas para impulsar el aprendizaje colectivo y mejorar los resultados educativos.

No obstante, fomentar la inteligencia colectiva en el aula también prepara a los estudiantes para enfrentar desafíos futuros en el ámbito laboral, gracias al cultivo de habilidades y competencias fundamentales para el éxito profesional. A través del trabajo en equipo, la comunicación efectiva y la colaboración en la generación de soluciones creativas, los estudiantes adquieren las destrezas necesarias para ser colaboradores efectivos en entornos laborales diversos y dinámicos. Como resultado, en el ámbito laboral, la inteligencia colectiva se traduce en un mayor rendimiento y productividad, permitiendo a los equipos abordar desafíos complejos y fomentar la innovación, creando soluciones más efectivas y creativas. Una de las herramientas más utilizadas en este entorno para entender la dinámica de grupo y la efectividad de la colaboración son las encuestas, ya que son un medio efectivo para recopilar datos sobre las percepciones, opiniones y experiencias de los miembros del equipo [2][9].

A lo largo de este proyecto, se explorará cómo la interacción social, complementada con el uso estratégico de la tecnología, puede realzar la efectividad académica y optimizar la toma de decisiones en ambientes educativos específicos, siendo en nuestro caso la universidad.

Las implicaciones prácticas derivadas de este proyecto podrían tener un impacto significativo en el bienestar y el rendimiento de los estudiantes universitarios. En primer lugar, al ayudar a comprender cómo las condiciones ambientales afectan al estado de ánimo de los estudiantes, la universidad podría implementar medidas para optimizar el diseño de las aulas y mejorar la calidad del entorno físico, ajustando la iluminación o la temperatura para crear entornos más confortables y propicios para el aprendizaje. En segundo lugar, conociendo cómo ciertos factores ambientales influyen en el estado de ánimo de los estudiantes, los profesores podrían

---

adaptar sus estrategias de enseñanza para fomentar un ambiente de aprendizaje más positivo y enriquecedor. Estas medidas podrían contribuir a mejorar el bienestar emocional de los estudiantes, así como su capacidad para concentrarse, aprender y tener un desempeño académico óptimo.

## **2.2 Calidad del Ambiente**

La calidad ambiental se refiere a las características del entorno que rodea a las personas, englobando todos los factores físicos y químicos que pueden tener un impacto directo o indirecto en su salud, bienestar y eficacia en actividades específicas. Este concepto trasciende la simple medición de parámetros individuales para abarcar una comprensión de cómo interactúan diversos elementos del ambiente, incluyendo la calidad del aire, la temperatura, la humedad, la luminosidad y el ruido, entre otros [10].

En espacios cerrados, como aulas y oficinas, donde las personas pasan una cantidad significativa de tiempo, la calidad ambiental adquiere una relevancia especial. Una buena calidad ambiental no solo contribuye a mantener y promover la salud física y mental de los ocupantes, sino que también mejora la concentración, la productividad y el rendimiento cognitivo. La calidad ambiental es especialmente crítica en entornos educativos, donde se necesitan condiciones óptimas para fomentar un aprendizaje efectivo y mantener altos niveles de atención y retención de conocimientos en los estudiantes [11].

Un estudio realizado en las aulas universitarias de Hong Kong [12] examina meticulosamente cómo varios factores ambientales (temperatura, concentración de CO<sub>2</sub>, nivel de iluminación y nivel de ruido de fondo) afectan la calidad ambiental interior (IEQ, por sus siglas en inglés) y, consecuentemente, el confort, la satisfacción y el rendimiento académico de los estudiantes. Utilizando un modelo de evaluación IEQ basado en evaluaciones comprensivas difusas y un proceso de jerarquía analítica, el estudio identifica que la calidad del ambiente térmico es el factor más crítico, seguido por la calidad del aire, iluminación y acústica. Estos hallazgos subrayan la importancia de gestionar y mejorar proactivamente cada uno de estos factores para optimizar los entornos de aprendizaje universitario.

Basándonos en la importancia resaltada por la investigación previa, el proyecto propuesto se enfocará en evaluar la calidad ambiental en la Universidad Politécnica de Madrid (UPM) y en cómo ésta afecta a los estudiantes. La comprensión de este impacto se logrará mediante la integración de la inteligencia colectiva, representada por el análisis de las percepciones de los estudiantes recogidos a través de encuestas, junto con la implementación de sensores que capturan las variables ambientales de su entorno.

Las variables ambientales medidas por los sensores serán explicadas en el apartado de especificaciones y restricciones.

---

## 2.3 Sensores

Un sensor es un dispositivo electrónico que mide la información física que encuentra en su entorno y la convierte en datos que codifica para poder ser interpretados por un ser humano o una máquina [13].

En su mayoría, los sensores son electrónicos y convierten los datos físicos recolectados en señales electrónicas para evaluar variables ambientales críticas. También existen sensores más sencillos como un termómetro de vidrio, que simplemente debe mostrar datos visuales.

Existen dos tipos de sensores electrónicos principales: sensores analógicos y digitales. Los sensores analógicos convierten los datos físicos capturados en una señal analógica. Son más precisos que los sensores digitales, ya que pueden darse cuenta de los cambios más leves de la variable física que captan. Sin embargo, los sensores digitales se limitan a un conjunto de valores finito (binarios). Los sensores funcionan convirtiendo un tipo de energía del ambiente detectado (por ejemplo, calor, luz, movimiento) en una señal eléctrica o datos.

### 2.3.1 Utilidad de los Sensores

Los sensores son necesarios en la actualidad debido a su capacidad de recopilar datos precisos y en tiempo real sobre nuestro entorno. Gracias a la evolución de la tecnología y el bajo costo de estos dispositivos, la utilización de sensores se ha popularizado y ahora tienen una amplia gama de aplicaciones, desde la monitorización del clima hasta el control de procesos industriales, la gestión de la salud, la seguridad, y la automatización del hogar, entre otros [14].

En un estudio reciente llevado a cabo en una escuela Melbourne, Australia [15], se ilustra el uso innovador de sensores junto con unos dispositivos portátiles que registraban datos fisiológicos para analizar cómo los entornos interiores y exteriores influyen el comportamiento y el estado mental de los ocupantes. El uso en conjunto de estos dispositivos permitió a los investigadores correlacionar la calidad ambiental con el confort térmico, el compromiso en el aprendizaje y las emociones de los estudiantes y el personal. Este estudio subraya la importancia de analizar las condiciones ambientales y fisiológicas en entornos educativos, para crear sistemas inteligentes que no solo mejoren el bienestar de los estudiantes y su rendimiento académico, sino también reduzcan el consumo energético.

En el estudio mencionado recalcan que, durante el diseño de los edificios, los ingenieros usaban simulaciones para predecir la calidad ambiental interior y el consumo energético.



---

Este punto es muy importante y ha de ser tenido en cuenta, ya que, actualmente los edificios son responsables de una tercera parte de la energía que se usa actualmente, así como de un 20% de las emisiones globales de CO<sub>2</sub> [16]. Esto se debe al uso de sistemas de calefacción, ventilación y aire acondicionado que se utilizan para regular el clima en espacios interiores. Se debe reducir este consumo energético para garantizar un futuro más sostenible.

Teniendo en cuenta que se ha invertido bastante energía y esfuerzo en proporcionar entornos adecuados a los ocupantes de los edificios, merece la pena investigar qué constituye exactamente su confort y bienestar.

Aunque el estudio aportó datos significativos y mejoró la comprensión de cómo el ambiente interior afecta el rendimiento académico y emocional, no concluyó con soluciones definitivas. Este resultado remarca la necesidad de nuestra investigación en la UPM, que aspira a extender estos hallazgos, integrando la inteligencia colectiva y la tecnología de los sensores para desarrollar estrategias efectivas que mejoren la calidad ambiental en espacios educativos, alineándose de tal manera con los objetivos globales de sostenibilidad y bienestar.

## **2.4 Herramientas de visualización para la toma de decisiones**

La visualización de datos [17] es el proceso por el que se convierte información cuantitativa y cualitativa en representaciones visuales y gráficos, mapas o diagramas. Esto permite a quienes toman decisiones observar de manera más intuitiva y comprender las tendencias, anomalías y patrones presentes en los datos.

Históricamente, esta técnica ha sido fundamental para revelar tendencias y patrones ocultos en grandes volúmenes de información. Un ejemplo icónico es el mapa creado por Charles Joseph Minard en 1869, que ilustró la campaña de Napoleón en Rusia de 1812 [18]. Este mapa empleó una representación gráfica innovadora para mostrar simultáneamente múltiples variables, como el número de tropas, la temperatura, la distancia recorrida y la ubicación geográfica, comunicando efectivamente las complejas interrelaciones históricas y geográficas.

Actualmente, la visualización de datos es más relevante que nunca por la gran cantidad de información que se genera diariamente. Las herramientas modernas de visualización permiten a los usuarios no solo ver datos en formas intuitivas, sino también interactuar con ellos para descubrir patrones profundos y hacer predicciones precisas. Estas capacidades son necesarias en áreas como el análisis de negocio, la investigación científica, la administración pública y la educación [19].

---

Existen diferentes métodos para realizar el análisis de datos en función de la información que se quiere investigar [20]:

1. **Análisis Descriptivo:** Visualiza datos históricos para identificar tendencias y patrones.
2. **Análisis Diagnóstico:** Ayuda a comprender las causas de los fenómenos observados. Se utiliza para resolver problemas concretos.
3. **Análisis Predictivo:** Utiliza modelos estadísticos de minería de datos, modelización predictiva y machine learning para prever futuros escenarios.
4. **Análisis Prescriptivo:** Combina el análisis predictivo y las reglas de decisión para sugerir acciones futuras.

#### 2.4.1 Herramientas

Existen muchas herramientas de visualización de datos disponibles, cada una con sus propias características. Para elegir una de ellas, se debe considerar el nivel de experiencia de la persona que la utilizará, los datos que se visualizarán, el precio del software o el tipo de gráficos y vistas que ofrece. Entre las más utilizadas se encuentran [20] [21]:

- **Tableau**

Tableau [22] es un software en visualización de datos que permite a los usuarios crear gráficos y tablas interactivas y dinámicas. Es utilizado por empresas y organizaciones para analizar, explorar y comunicar información. Tableau ofrece una amplia gama de opciones de visualización y facilita la creación de visualizaciones complejas sin necesidad de tener conocimientos profundos en programación o análisis de datos, permite una amplia conectividad con diversas fuentes de datos como archivos de texto, hojas de cálculo, fuentes de datos en línea etc.

Además de sus capacidades de visualización de datos, Tableau también ofrece características de análisis avanzadas, ya que los usuarios pueden integrarla con otras herramientas y servicios, como R y Python, para realizar análisis más estadísticos.

- **Power BI**

Microsoft Power BI [20] es una herramienta de visualización de datos creada en 2015 por Microsoft que facilita la creación de visualizaciones dinámicas e interactivas. Power BI proporciona funcionalidades avanzadas de inteligencia artificial y machine learning, lo que permite a los usuarios generar análisis predictivos y prescriptivos directamente desde la plataforma. Permite la conexión con una diversidad de fuentes de datos y ofrece la capacidad de transformar y modelar estos datos eficientemente. Power BI proporciona un amplio rango de tipos de visualización y se integra fácilmente con otras

---

herramientas de Microsoft. Además, asegura la protección y seguridad de los datos a través de su integración con Azure Active Directory.

- **Google Data Studio**

Google Data Studio [23] es una herramienta de Google gratuita para la creación de cuadros de mando y la visualización de datos. Es utilizada principalmente en el sector del marketing, ya que permite conectar y visualizar información de múltiples fuentes usadas en ese sector como Google Analytics y Google Ads.

Los usuarios pueden crear informes interactivos y personalizados, colaborar en tiempo real y compartir fácilmente sus dashboards. Además, se integra perfectamente con otras herramientas de Google, proporcionando opciones de personalización avanzadas y manteniendo altos estándares de seguridad y gobernanza de datos.

Es reconocida frente a otras herramientas porque es muy intuitiva y fácil de aprender a manejar, además destaca por su capacidad para transformar datos, limpiando y modelando la información para visualizarla. Sin embargo, para el cuadro de mando que se pretende desarrollar en este trabajo, presenta dificultades, debido a que la fuente principal es Google Sheets, herramienta que no está siendo usada para realizar la Base de Datos.

#### **2.4.2 Descripción de la herramienta escogida**

Para realizar el desarrollo del cuadro de mando, utilizaremos Microsoft Power BI como herramienta de visualización.

Power BI [24] es una colección de servicios de software, aplicaciones y conectores que funcionan en conjunto para convertir orígenes de datos sin relación entre sí en información coherente, interactiva y atractiva visualmente. Esta herramienta permite trabajar tanto con hojas de cálculo de Excel como con complejas bases de datos híbridas, tanto locales como en la nube. Facilita la conexión a fuentes de datos, ayuda a visualizar y descubrir aspectos clave de la información y compartirla con cualquier usuario deseado.

Consta de varios componentes que operan en conjunto, destacándose tres elementos fundamentales:

- a. **Power BI Desktop:** Aplicación de escritorio que permite extraer, transformar y visualizar datos mediante la creación de cuadros de mando interactivos.
- b. **Power BI Service:** Funciona como un servicio online (SaaS) con funcionalidades similares a Power BI Desktop, facilitando la extracción y la generación de informes a través de dashboards.
- c. **Power BI Mobile:** Aplicación diseñada para dispositivos móviles con sistemas iOS, Android y Microsoft.

---

Para este proyecto, se utilizará Power BI Desktop para desarrollar el cuadro de mando. Power BI Desktop utiliza un enfoque de transformación de datos conocido como ETL (extracción, transformación y carga), que se maneja de manera eficiente a través de Power Query. Esta herramienta simplifica el proceso de convertir datos de diferentes orígenes, ya sean simples o complejos, en visualizaciones claras y fácilmente interpretables en cuadros de mando [25].

Actualmente, Power BI se ha posicionado como líder en el ámbito de la Inteligencia Empresarial gracias a su notable flexibilidad y facilidad de uso, con presencia en más de 200 países, en cerca de 200.000 empresas y más de 5 millones de usuarios según datos de ENAE [26]

Se ha escogido esta herramienta debido a que facilita la integración de datos desde diversas fuentes como Excel, CSV, y SQL [24]. Esto es necesario para nuestro proyecto, ya que la base de datos capturados por los sensores están en formato CSV y las encuestas recogidas de los estudiantes en formato Excel. Además, Power BI tiene una gran capacidad para conectarse con múltiples plataformas y bases de datos, remarcando su utilidad como herramienta analítica en la nube.

Para hacer uso de Power Bi, es necesaria la creación de **informes** que recopilen y muestren los datos relevantes del modelo para estudiar problemas, identificar tendencias etc. En Power BI Desktop, un informe es una hoja que permite visualizar diversos objetos como gráficos de diferentes tipos y tablas, que representan análisis realizados sobre los datos importados. Los informes ofrecen una presentación más detallada y extensa de la información, incluyendo un análisis profundo de los datos.

Por otra parte, el **cuadro de mando** es una herramienta que contiene la información más importante, necesaria para lograr uno o más objetivos. En el cuadro de mando se resume la información esencial a través de una sola pantalla, ofreciendo en un vistazo rápido los indicadores clave de rendimiento (KPI) y los datos críticos del proyecto realizado.

Finalmente cabe destacar que Power BI es útil por la sencillez en la que esta creada, por su facilidad en compartir los informes, ya que se encuentran alojados en Azure y además por ser gratuita, lo que hace que muchas empresas e instituciones académicas opten por usar esta aplicación.

En la figura 1 se puede ver un ejemplo visual de un cuadro de mando implementado en Power BI:



Figura 1: Ejemplo Cuadro de mando en Power BI [27]

---

---

## 3. Especificaciones y restricciones de diseño

En este apartado, abordaremos las especificaciones y restricciones de diseño necesarias para el desarrollo efectivo del proyecto. Se detallarán tanto los componentes técnicos utilizados como las limitaciones a estos recursos.

### 3.1 Especificaciones

#### 3.1.1 Especificaciones de Sensores

Para la elaboración del proyecto, usaremos los sensores SCD30, VEML7700 y KECG2240PBJ proporcionados por la escuela Técnica Superior de Ingeniería de Sistemas Informáticos (ETSISI).

El módulo de sensores SCD30 de Sensirion (Figura 2) emplea tecnología de infrarrojos no dispersivos (NDIR) para realizar la medición de CO<sub>2</sub> y cuenta con sensores de temperatura y humedad integrados. La temperatura y la humedad ambiental se miden mediante el control y compensación de las fuentes externas de calor, sin necesidad de utilizar otros componentes adicionales.



Figura 2: Sensor SCD30 [28]

Por otra parte, utilizaremos el sensor de iluminación VEML7700 (Figura 3). Este sensor de luz ambiental ofrece mediciones con una resolución de 16 bits. Se trata de una pequeña caja transparente que incorpora un fotodiodo, un amplificador de bajo ruido y un conversor A/D de 16 bits, proporcionando lecturas de luz ambiental como valores digitales.

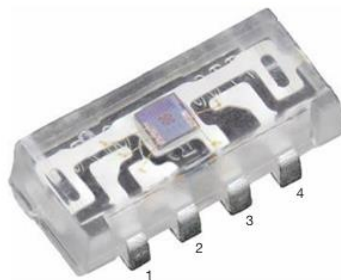


Figura 3: Sensor de luz ambiental VEML7700 [29]

---

Para medir el sonido, utilizaremos un micrófono de condensador electret KECG2240PBJ de Kingstate (Figura 4). El material electret, cargado permanentemente, permite generar una señal eléctrica proporcional al sonido captado sin necesidad de una fuente de alimentación externa, garantizando alta sensibilidad y estabilidad.



Figura 4: Micrófono Condensador Electret KECG2240PBJ [30]

Estos tres elementos se integraron en un mismo componente, lo que permitió un ahorro energético y económico. Además, fueron seleccionados para cubrir las medidas que cualifican la calidad interior del aula siguiendo la normativa europea del Building Performance Institute Europe (BPIE).

Las variables medidas de calidad ambiental por los sensores son las siguientes:

- **CO2 (Dióxido de Carbono):**
  - **Unidad de Medida:** Partes por millón (ppm).
  - **Rango de medida del sensor:** 0 a 40.000 ppm.
  - **Rango típico en ambientes interiores:** 400 ppm a 1.000 ppm.
  - **Importancia:** el dióxido de carbono es un indicador de la ventilación y de calidad del aire. Niveles elevados de CO2 pueden indicar una ventilación inadecuada y están asociados con una disminución en el rendimiento cognitivo y el bienestar [31].
- **Humedad Relativa:**
  - **Unidad de Medida:** Porcentaje (%).
  - **Rango de medida del sensor:** 0% al 100%
  - **Rango típico en ambientes interiores:** 30% al 60%.
  - **Importancia:** Se refiere al porcentaje de vapor de agua en el aire a una temperatura específica [32], niveles de humedad adecuados son necesarios para la salud a largo plazo. La humedad excesiva puede ocasionar el crecimiento de moho y otros alérgenos, mientras que la humedad baja puede causar irritación en los ojos y las vías respiratorias, falta de concentración e incluso procesos gripales.



- 
- **Temperatura:**
    - **Unidad de Medida:** Grados Centígrados (°C).
    - **Rango de medida del sensor:** -40°C hasta 70 °C
    - **Rango típico en ambientes interiores:** 17°C a 27°C.
    - **Importancia:** La temperatura tiene un gran impacto en el confort y la concentración. Las temperaturas que están fuera del rango de confort pueden llevar a una disminución en la productividad y aumentar la distracción, también pueden provocar somnolencia, estrés o irritabilidad [33].
  
  - **Luminosidad**
    - **Unidad de Medida:** Lux
    - **Rango de medida del sensor:** 0,01 lux hasta 167.000 lux
    - **Rango típico en ambientes interiores:** 300 lux a 500 lux.
    - **Importancia:** Una iluminación adecuada es necesaria para realizar actividades visuales con comodidad y eficiencia. La iluminación insuficiente puede causar fatiga visual y disminuir la concentración, mientras que una iluminación excesiva puede causar deslumbramiento y, por tanto, malestar [33].
  
  - **Ruido**
    - **Unidad de Medida:** Decibelios (dB)
    - **Rango de medida del sensor:** 30 dB hasta 130 dB
    - **Rango típico en ambientes interiores:** 40 dB a 55dB
    - **Importancia:** El nivel de ruido puede afectar significativamente la concentración y el bienestar. Niveles altos de ruido pueden causar estrés, dificultades de comunicación, pérdida de concentración y afectar negativamente al rendimiento académico [33].

---

### 3.1.2 Especificaciones de Análisis

#### Preparación de los datos

- **Importación de Datos:** Se hará uso de Python para la importación de los datos en estructuras de datos de la librería *Pandas* para su manipulación y análisis posterior.
- **Limpieza de los datos:** Identificación y manejo de los valores faltantes, datos erróneos o inconsistentes en el conjunto de datos.
- **Reestructuración de los datos:** Reestructuración de los datos según sea necesario para que se ajusten a los requisitos del análisis.

#### Análisis de relaciones

- **Identificación de Relaciones:** Identificación de relaciones entre las variables ambientales y las percepciones de los estudiantes.
- **Visualización de correlaciones:** Visualización de correlaciones utilizando diferentes gráficos de dispersión, mapas de calor etc.

#### Power BI

- **Cuadro de Mando:** Creación de un cuadro de mando interactivo que presente los diferentes KPIs y correlaciones entre las variables.

## 3.2 Restricciones

- **Volumen de datos:** Disponibilidad de Datos de solo dos meses.
- **Precisión de los KPIs:** Las correlaciones deben de ser precisas para informar conclusiones confiables.
- **Regulaciones de Privacidad:** Los datos recopilados por los sensores y las encuestas deben cumplir con regulaciones de privacidad y protección de datos para garantizar la confidencialidad y seguridad de la información recopilada.

---

## 4. Metodología

Previamente a los apartados de desarrollo de la solución y resultados, es necesario entender la metodología que ha sido implementada para la elaboración de este proyecto. En esta sección se describirán detalladamente los datos que han sido utilizados para el estudio y los diferentes modelos de machine learning que se han implementado. También se explicarán los diferentes parámetros y gráficas que serán utilizados para exponer los resultados.

### 4.1 Recopilación de datos

La recopilación de datos para este proyecto se realizó en dos fases principales:

1. Obtención de datos de sensores ambientales
2. Recolección de encuestas realizadas a los estudiantes.

En los siguientes apartados se describen los detalles de cada conjunto de datos.

#### 4.1.1 Datos de Sensores Ambientales

Los datos de los sensores ambientales fueron tomados para medir las condiciones del entorno durante las clases. Para realizar el estudio, los sensores se colocaron en el aula 3301 de la ETSISI. Las variables medidas son las siguientes:

- **Temperature (Temperatura):** Medida en grados Celsius. (°C)
- **CO2 (Dióxido de Carbono):** Medida en partes por millón (ppm).
- **Luminosity (Luminosidad):** Medida en lux.
- **Humidity (Humedad):** Medida en porcentaje (%).
- **Noise (Ruido):** Medida en decibelios (dB)

#### 4.1.2 Datos de los Cuestionarios

Se administraron encuestas a los alumnos al final de cada clase para obtener datos sobre sus percepciones y estados físicos y anímicos. Las preguntas de las encuestas se clasifican en dos tipos: preguntas tipo Likert y preguntas binarias.

##### Preguntas tipo Likert

Las preguntas tipo Likert utilizan una escala del 1 al 5, con valores específicos para cada pregunta. Las preguntas que se formularon con esta escala son:

#### 1. La docencia impartida en la clase me ha parecido buena.

- 1: Totalmente en desacuerdo
- 5: Totalmente de acuerdo

#### 2. Mi estado anímico durante la clase ha sido bueno.

- 1: Totalmente en desacuerdo
- 5: Totalmente de acuerdo

---

**3. Mi estado físico durante la clase ha sido bueno.**

- 1: Totalmente en desacuerdo
- 5: Totalmente de acuerdo

**4. La luminosidad durante la clase ha sido adecuada**

- 1: Nada adecuada
- 5: Totalmente adecuada

**5. El nivel de ruido durante la clase ha sido:**

- 1: Nada adecuado (mucho ruido)
- 5: Totalmente adecuado (poco ruido)

**6. La temperatura del aula durante la clase ha sido:**

- 1: Nada adecuada
- 5: Totalmente adecuada

**7. ¿Se ha notado un ambiente cargado?**

- 1: Ambiente nada cargado
- 5: Ambiente muy cargado

**8. En cuanto, a la ocupación del aula, dirías que está:**

- 1: Vacía
- 5: Llena

**9. En general, durante la clase, el aula me ha parecido:**

- 1: Nada confortable
- 5: Totalmente confortable

**Preguntas Binarias**

Las preguntas binarias permiten respuestas de “Sí”, “No” y “Parcialmente”. Las preguntas que se formularon de forma binaria son:

- 1. ¿Están las luces apagadas?**
- 2. ¿Están las cortinas abiertas?**
- 3. ¿Está el aire acondicionado encendido?**
- 4. ¿Está la calefacción encendida?**
- 5. ¿Están las ventanas abiertas?**

---

### **Estados Anímicos y Físicos**

Para poder tener una visión más completa de los estados anímicos y físicos de los alumnos, se incluyeron preguntas en las que podían seleccionar múltiples respuestas.

**1. Con respecto a tu estado ANÍMICO ... ¿Cómo te has sentido? \* Puedes seleccionar VARIAS respuestas:**

- Eufórico
- Motivado
- Contento
- Normal
- Cansado
- Con Ansiedad
- Agobiado
- Estresado
- Triste
- Aburrido
- Dormido

**2. Con respecto a tu estado FÍSICO... ¿Cómo te has sentido? \* Puedes seleccionar VARIAS respuestas:**

- Enérgico/a
- Normal
- Agotado/a
- Enfermo/a
- Con hambre o sed
- Con dolor de cuello
- Con dolor de cabeza
- Con dolor de espalda
- Con sueño

Por otro lado, la encuesta también pedía el nombre y apellido del alumno, el número de aula (3301), el puesto que ocupaban de la clase y la hora de inicio y final de la clase que había tenido justo antes de rellenarla. Cabe destacar que los alumnos obtuvieron una recompensa en forma de créditos por participar y contestar las preguntas durante el semestre.

## 4.2 Algoritmos y Metodologías Utilizadas

En este apartado se explicarán los algoritmos de aprendizaje automático empleados para analizar los datos recolectados y a su vez encontrar modelos que sean capaces de realizar predicciones de aquellos atributos más relevantes. Los métodos de evaluación de estos modelos de machine learning y las diferentes formas de captar las características importantes también serán descritas, con el objetivo de poder extraer la máxima información de los datos obtenidos de las diferentes fuentes anteriormente mencionadas.

### 4.2.1 Algoritmos de Aprendizaje Automático

#### Árbol de Decisión

Un árbol de decisión [34] es un algoritmo de aprendizaje automático que se utiliza para tareas de clasificación o regresión con múltiples variables con el fin de predecir una variable objetivo discreta o numérica. Como su propio nombre indica, la estructura de un árbol consta de:

- **Nodo raíz:** Es el punto inicial del árbol de decisión, no tiene ramas entrantes. Las ramas salientes del nodo raíz alimentan a los nodos internos o de decisión.
- **Nodos de decisión:** Evalúan las características disponibles para formar subconjuntos homogéneos, que se indican mediante los nodos hoja. En estos nodos se analizan los parámetros con los que se quiere realizar la predicción de la variable objetivo.
- **Nodos hoja:** Representan los resultados que puede tomar la variable objetivo dentro del modelo.

El aprendizaje basado en árboles de decisiones emplea una estrategia de divide y vencerás, buscando los puntos de división óptimos dentro de un árbol. Este proceso se repite de forma recursiva desde la raíz hacia las hojas o hasta que la mayoría de los registros se hayan clasificado bajo etiquetas de clase específicas. La complejidad del árbol influye en la pureza de los nodos hoja: los árboles más pequeños tienden a tener nodos hoja más puros, mientras que, en los árboles más grandes, mantener esa pureza es más difícil.

A continuación, se puede ver claramente un árbol de decisión con sus terminologías:

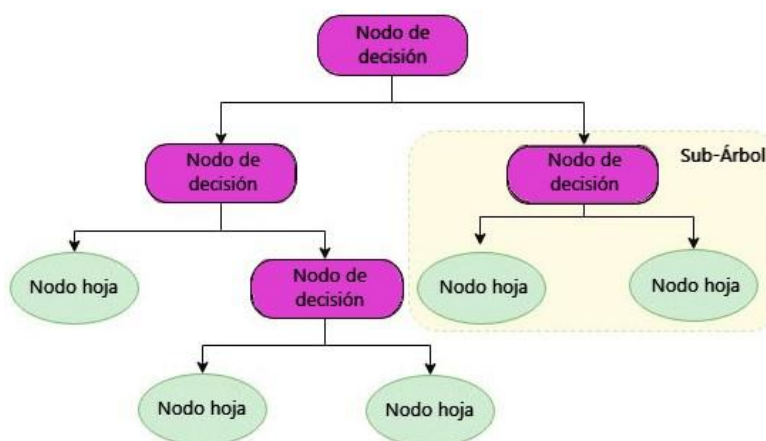


Figura 5: Árbol de Decisión

---

## Random Forest (Bosque Aleatorio)

El modelo Random Forest [35] es un algoritmo de aprendizaje automático supervisado que pertenece a las técnicas de ensemble, es decir, técnicas que combinan múltiples modelos para mejorar las predicciones. Se le llama “Bosque aleatorio” debido a que implica la construcción de un bosque de árboles de decisión (Figura 6), donde cada árbol se entrena utilizando una muestra aleatoria del conjunto de datos original. Las predicciones finales se obtienen promediando las predicciones individuales de todos los árboles en el caso de la regresión, o mediante votación en el caso de la clasificación.

El funcionamiento [36] de este algoritmo se puede dividir en varias fases:

- **Bootstrap Aggregating (Bagging):** Se seleccionan muestras aleatorias con reemplazo del conjunto de datos de entrenamiento, es decir, se toman diferentes muestras aleatorias de los datos originales y se entrena un modelo en cada una de estas muestras.
- **Construcción de Árboles de Decisión (Bootstrap Sampling):** Se construye y se entrena un árbol de decisión para cada muestra de datos diversos, seleccionando un subconjunto aleatorio de características en cada división del árbol. Esta técnica, conocida como bootstrap sampling, ayuda a reducir la correlación entre los árboles.
- **Predicción y Resultado Final:** En problemas de clasificación, cada árbol “vota” por la clase correspondiente a una nueva instancia de datos y la clase con la mayoría de los votos se convierte en la predicción final del modelo. En problemas de regresión, se toma el promedio de todas las predicciones de los árboles para obtener el resultado final. Esta combinación de resultados de múltiples árboles permite al modelo manejar mejor la variabilidad de los datos y proporcionar predicciones más precisas y estables.

Las ventajas del modelo de Random Forest incluyen su capacidad para manejar grandes conjuntos de datos con múltiples características y la facilidad de determinar la importancia de características. Sin embargo, presenta desventajas como un procesamiento más lento y el requerimiento de más recursos debido al manejo de las estructuras de datos empleadas, además la interpretación de sus predicciones tiene una mayor complejidad en comparación con un árbol de decisión.

Para este trabajo se ha utilizado el algoritmo Random Forest debido a su capacidad para manejar conjuntos de datos complejos que tienen tanto variables continuas (las medidas de los sensores) como variables categóricas (percepciones de los estudiantes).

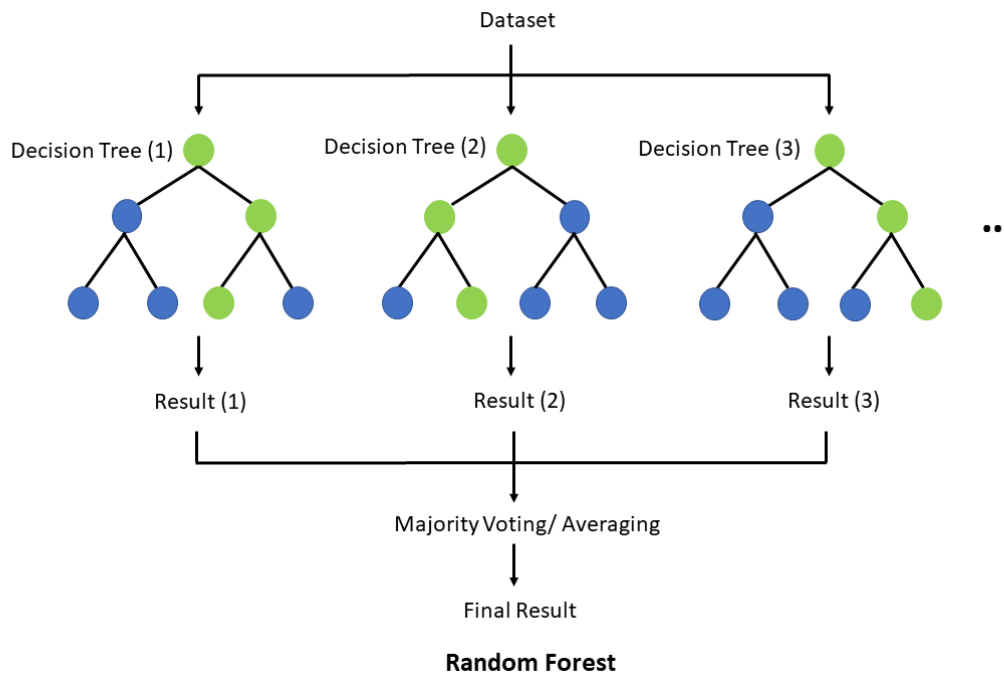


Figura 6: Diagrama algoritmo Random Forest [37]

### Ajuste de Hiperparámetros

El algoritmo Random Forest dispone de una serie de hiperparámetros [38] para ajustar el modelo en base a las características del conjunto de datos. Estos parámetros se tienen en cuenta durante el proceso de entrenamiento y por ello se pueden obtener resultados más precisos. Los hiperparámetros pueden ser modificados y ajustados a medida que se va experimentando con el modelo y el conjunto de datos empleado, para así reducir el error y el riesgo de sobreentrenamiento:

- **n\_estimators:** Este parámetro establece el número de árboles que hay en el bosque. Un valor alto aumenta la precisión del modelo, pero también incrementa el tiempo de entrenamiento y predicción y los recursos computacionales necesarios. Un valor elevado de este parámetro también puede dar lugar a crear un modelo adaptado al conjunto de datos y tener problemas a la hora de predecir datos de prueba, es decir, puede dar lugar a sobreentrenamiento.
- **max\_depth:** Establece la profundidad máxima de cada árbol en el bosque.
- **min\_samples\_leaf:** Establece el número mínimo de muestras necesarias para dividir una hoja.
- **min\_samples\_split:** Establece el mínimo número de muestras para dividir un nodo interno (nodo de decisión).
- **max\_features:** Este parámetro indica el número máximo de características a considerar en cada división de datos para entrenar los árboles de decisión. Un valor alto permite seleccionar más atributos en cada división, lo que hace que se aumente la diversidad de los árboles y reduce la correlación entre ellos.



También se pueden ajustar los hiperparámetros, utilizando un método llamado error fuera de bolsa (out-of-bag error, OOB) característico de Random Forest o emplear validación cruzada. El error OOB se obtiene al dejar fuera una muestra del conjunto de datos para la validación del modelo, proporcionando una estimación precisa del error sin necesidad de un conjunto de validación separado.

Para ajustar los hiperparámetros en modelos RandomForest se pueden utilizar funciones como **GridSearchCV()** y **RandomForestCV()** de la librería de Python **Scikit-learn**. Se diferencian en la forma en la que buscan en el espacio los hiperparámetros [39].

Grid Search evalúa el rendimiento de un modelo a través de los diferentes parámetros que se introducen en la cuadrícula predefinida de parámetros y extrae los mejores valores y combinaciones. Por otra parte, la búsqueda por randomForest muestrea los hiperparámetros aleatoriamente a partir de una distribución. En la figura 7 se puede ver claramente el funcionamiento de cada uno:

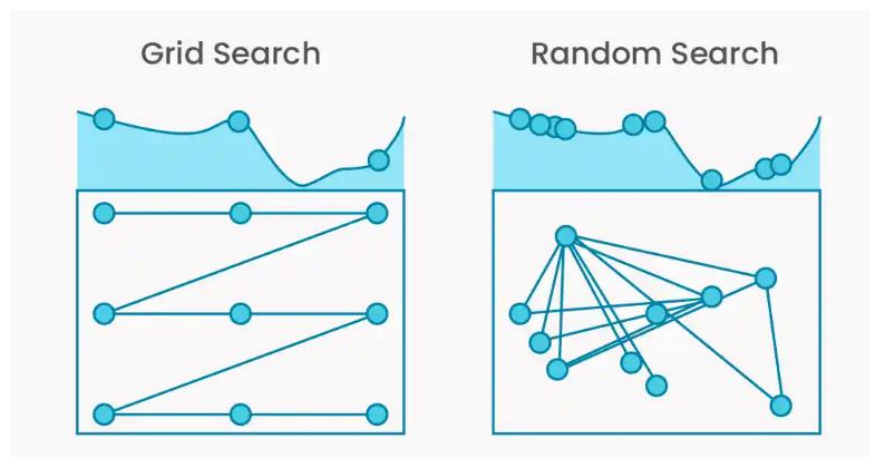


Figura 7: Ejemplo de funciones de ajuste de hiperparámetros [39]

El overfitting [40] (Sobreajuste o sobre entrenamiento) es un término utilizado para evaluar algoritmos de aprendizaje automático los cuales tienen un ajuste muy específico para el conjunto de datos empleado en el proceso de entrenamiento. Es decir, cuando un modelo presenta señales de overfitting al añadir datos diferentes a los de entrenamiento en los resultados pierde precisión.

---

#### 4.2.2 Evaluación del Modelo

Se deben de establecer unos criterios de comparación homogéneos para poder realizar un análisis y definir que algoritmo de los probados obtiene mejores resultados. Además, se pueden utilizar para evaluar el rendimiento de cada modelo de predicción. A continuación, se detallan las diferentes métricas de evaluación utilizadas en este trabajo:

##### **Coeficiente de Determinación ( $R^2$ )**

El coeficiente R-cuadrado [41] es una medida estadística utilizada comúnmente para evaluar la capacidad productiva de un modelo de regresión. Mide la proporción de la varianza explicada por el modelo en relación con la varianza total, en otras palabras, indica en forma de porcentaje se ajusta un modelo a la variación de la variable a predecir. Se puede utilizar como métrica para la comparación, ya que no varía dependiendo del modelo utilizado. En la Ecuación 1 se puede ver cómo es la fórmula de esta métrica:

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{total}} = 1 - \frac{\sum_i (y_i - y_{Regression})^2}{\sum_i (y_i - \bar{y}_i)^2}$$

**Ecuación 1: Fórmula R-Cuadrado o Coeficiente de Determinación**

- $y_i$  : Representa los valores observados
- $y_{Regression}$  : Representa las predicciones del modelo
- $\bar{y}_i$  : Representa la media de los valores observados

El coeficiente R-cuadrado varía entre 0 y 1 normalmente. Un valor cercano a 1 indica un ajuste fuerte, por lo que las predicciones del modelo están cercanas a los datos reales. Por lo contrario, un valor cercano a 0 indica un ajuste débil, asique el modelo no explica ninguna de las variaciones de los datos de salida [42]. A continuación, se muestran unos rangos de cómo se podría evaluar un modelo en base a su coeficiente de determinación:

- $R^2 < 0,5 \rightarrow$  **Ajuste débil**
- $0,5 \leq R^2 \leq 0,8 \rightarrow$  **Ajuste moderado**
- $R^2 > 0,8 \rightarrow$  **Ajuste fuerte**

##### **Error Absoluto Medio (MAE):**

El Error absoluto medio (Ecuación 2) es una medida estadística que representa el promedio de la diferencia absoluta entre el valor observado y los valores predichos [42].

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

**Ecuación 2: Fórmula MAE o Error Absoluto Medio**

- $n$ : Representa el número de observaciones
- $y_i$  : Representa el valor real
- $\hat{y}_i$  : Representa el valor predicho

---

Todos los errores se ponderan en la misma escala, debido a que se toman valores absolutos y sirve como medida uniforme para comprobar como de bien funciona el modelo.

### Raíz del Error Cuadrático Medio (RMSE)

El RMSE [42], proveniente del término en inglés **Root Mean Squared Error** (Ecuación 3), se calcula como la raíz cuadrada de la media de los errores al cuadrado (MSE) entre las predicciones y los valores reales de los datos de prueba. Mide el promedio de los errores y se ocupa de las desviaciones del valor real. Cuanto menor sea el RMSE, mejor será el modelo y, por tanto, sus predicciones.

$$RMSE = \sqrt{\frac{1}{n} \sum_i (y_i - y_{Regression})^2}$$

Ecuación 3: Fórmula RMSE o Raíz del Error Cuadrático Medio

- **n**: Representa el número de observaciones
- **$y_i$**  : Representa el valor real
- **$y_{Regression}$**  : Representa las predicciones del modelo

Como ventaja, proporciona una medida de la precisión del modelo fácil de interpretar, ya que está en las mismas unidades que la variable respuesta. Además, es sensible a errores grandes, ya que los errores al cuadrado se suman.

### Validación cruzada

La validación cruzada, o cross-validation [43], es un método utilizado para evaluar el rendimiento de un modelo predictivo de Machine Learning. Este método requiere separar previamente una parte de los datos del conjunto de entrenamiento, los cuales no se utilizarán para entrenar el modelo, sino para probarlo y validarlo.

#### K-Folds- Cross Validation (Validación Cruzada)

Esta técnica garantiza que todas las observaciones de la serie de datos original puedan aparecer tanto en la serie de datos de entrenamiento como en la de prueba, lo cual es bastante útil cuando se dispone de datos limitados, como en el caso de este proyecto.

Su funcionamiento sigue los siguientes pasos:

1. **División del conjunto de datos**: separa aleatoriamente el conjunto de datos en K partes iguales, denominadas “folds”.
2. **Selección un valor de K**: Elige un valor de K en función del tamaño del conjunto de datos, comúnmente entre 5 y 10.

3. **Entrenamiento y validación:** para cada uno de los K-folds:

- Ajusta el modelo utilizando K-1 folds como conjunto de entrenamiento.
- El fold restante se utiliza como conjunto de prueba para validar el modelo.
- Registra las puntuaciones y los errores obtenidos.

4. **Cálculo de la métrica de rendimiento:** Se repite el proceso hasta que cada K-fold haya sido utilizado como conjunto de prueba. Posteriormente se calcula la media de las puntuaciones registradas para obtener la métrica de rendimiento del modelo.

Este proceso se puede hacer manualmente o haciendo uso de las funciones de la librería **Scikit-Learn** de Python:

- **cross\_val\_score:** Indica la puntuación de cada fold de prueba.
- **cross\_val\_predict:** Ofrece las predicciones para cada observación cuando forma parte del conjunto de prueba.

El proceso de K-Folds Cross validation se ilustra en la Figura 8, donde se muestra cómo se dividen los datos en diferentes folds y como cada fold se utiliza una vez como conjunto de prueba mientras los demás se utilizan para entrenamiento.

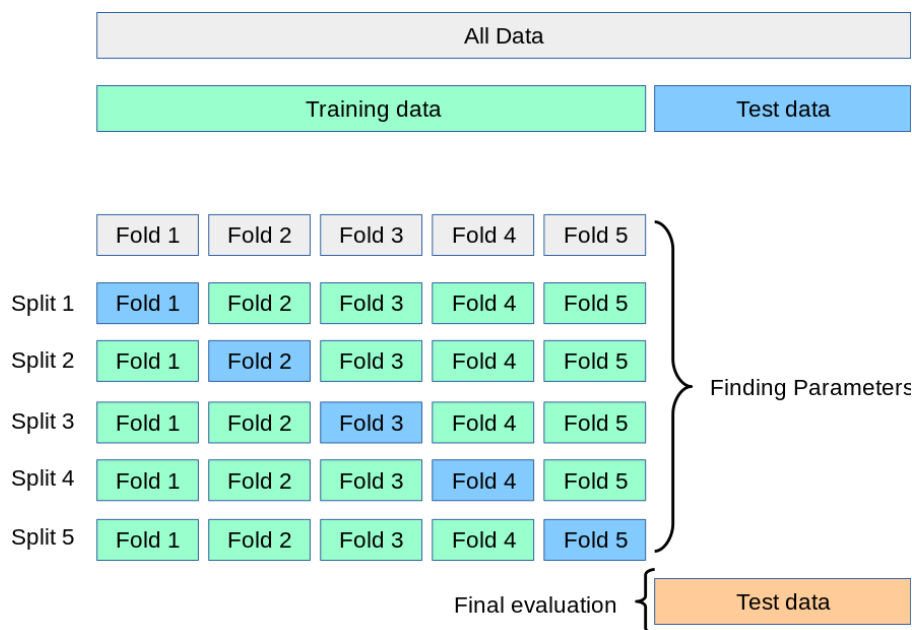


Figura 8: Proceso de K-Folds Cross Validation [44]

## Evaluación del modelo con Gráficas de Residuos

Otra manera de evaluar la robustez de un modelo es haciendo uso de sus residuos. Los residuos son los errores del modelo, la diferencia entre los valores observados y los valores predichos por el modelo de regresión. Sirven para evaluar si los datos obtenidos en el modelo de predicción son asimétricos o incluyen valores atípicos [45].

Se utilizará una gráfica de **residuos vs ajustes** para comprobar que los residuos se distribuyen aleatoriamente y tienen una varianza constante. Para que el modelo sea bueno los residuos deben estar ubicados de manera aleatoria alrededor del 0.

Patrones que indican si el modelo no cumple:

Patrón	Indicación
Dispersión en abanico o irregular de los residuos en los valores ajustados	Varianza no constante
Curvilíneo	Un término de orden superior faltante
Un punto que está alejado de cero	Un valor atípico
Un punto que está lejos de los otros puntos en la dirección x	Un punto influyente

Generalmente, en la gráfica de residuos vs ajustes (Figura 9) los residuos se muestran en el eje Y y los valores predichos o ajustados en el eje X

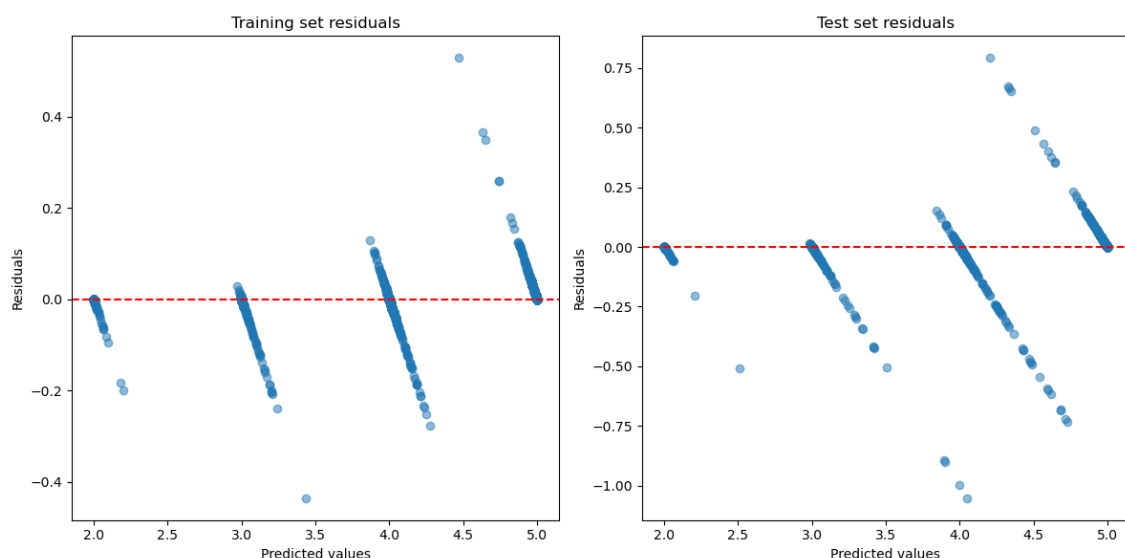


Figura 9: Ejemplo Gráfico de Residuos

---

### 4.2.3 Importancia de las Características

La importancia de las Características o importancia de variables de predicción [46] es un concepto utilizado en el campo de machine learning que sirve para comprender que características (variables) contribuyen más en las predicciones del modelo. Esta técnica resulta muy útil porque facilita la interpretación de los modelos e incluso ayuda a la optimización. Al identificar y eliminar las características más irrelevantes, se simplifica el modelo y se reduce el tiempo de entrenamiento y en algunos casos aumenta la precisión al reducir el ruido causado por aquellas variables que no aportan al modelo.

Métodos para calcular la importancia de características:

- **Reducción Promedio de la Impureza (Mean Decrease Impurity, MDI):** Se utiliza principalmente en modelos basados en árboles, como el Random Forest. El algoritmo calcula la disminución promedio de la impureza causada por la división en una característica particular. Aquellas características con mayor reducción de impurezas son las que se consideran más importantes.
- **Importancia de la permutación (Permutation Importance):** Este método mezcla aleatoriamente los valores de una sola característica y mide el impacto de esta en el rendimiento del modelo. En otras palabras, muestra cómo se vería afectado el modelo si se le impidiera aprender de esa característica. Cuanto mayor sea la caída del rendimiento, más importante es la característica.
- **SHAP (explicaciones de aditivos de Shapley):** Este método es muy útil para explicar la salida de cualquier modelo a través de valores Shapley. Utiliza un enfoque basado en la teoría de juegos cooperativos, donde se mide la contribución de cada jugador (en este caso, cada característica) al resultado final. En machine learning, a cada característica se le asigna un valor de importancia que representa su contribución al modelo [47].

Los valores Shapley [47] son capaces de mostrar cómo afecta cada característica a la predicción final, la importancia de cada característica en comparación con otras y la dependencia del modelo de la interacción entre características. Esto permite tener una interpretación más detallada de las influencias de cada variable en el modelo, haciendo que sea más fácil comprender como funcionan.

### Visualización de la Importancia de Características

Para visualizar la importancia de características de manera clara y legible, existen diferentes gráficos que se pueden utilizar en pyhon:

- **Gráfico de Barras:** Haciendo uso de un diagrama de barras simple (Figura 10) se puede mostrar las puntuaciones de importancia de las características principales. Se implementan en Python usando bibliotecas como *Matplotlib* o *Seaborn*.

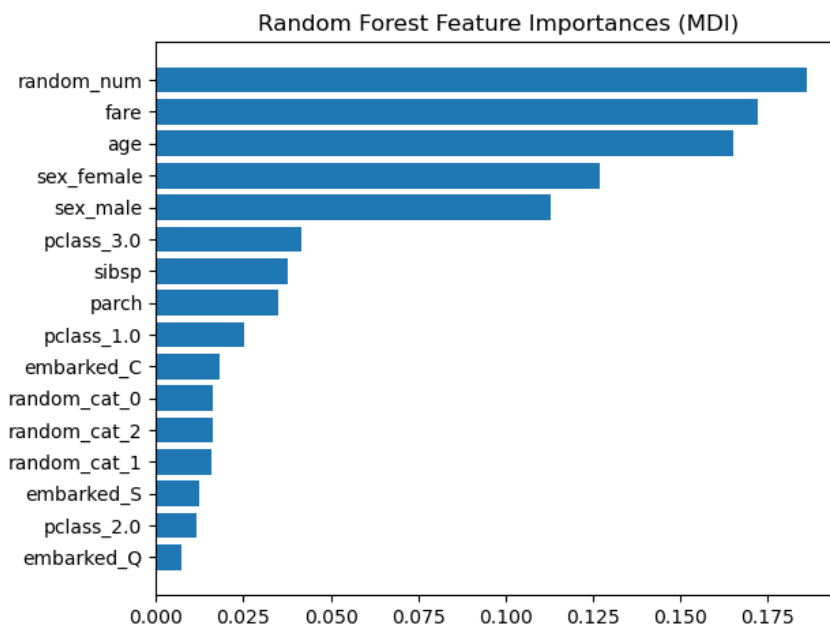


Figura 10: Ejemplo de Gráfico de Barras Random Forest [48]

- **Gráficos SHAP:** los valores SHAP pueden visualizarse de diferentes maneras:

#### Gráficos Resumen (summary plots)

En los gráficos resumen (Figura 11) se muestra la distribución de los valores SHAP para cada característica, proporcionando una visión tanto de la importancia de cada característica individual como del impacto de estas en la predicción total. Se utiliza la librería *shap* de Python junto con *matplotlib*.

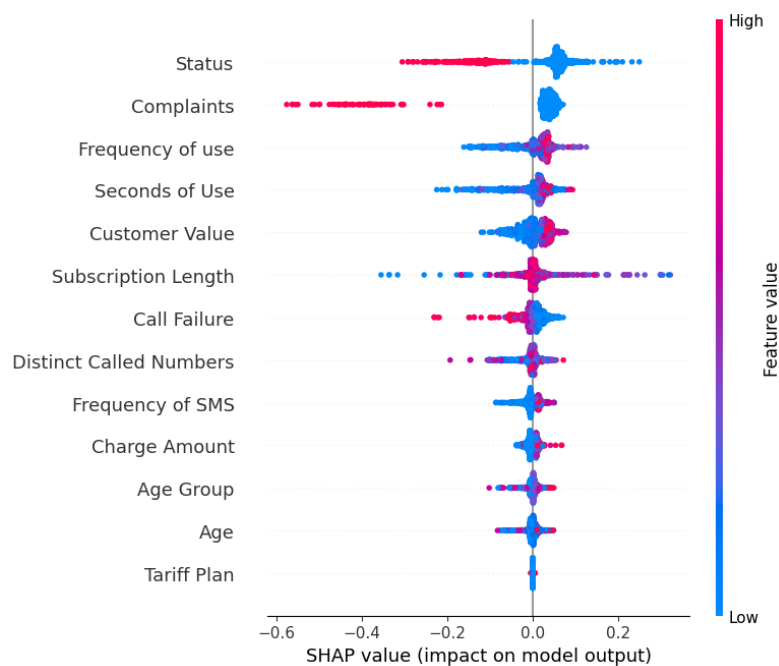


Figura 11: Ejemplo de Gráfico SHAP Resumen [47]

Comprensión de la gráfica:

- **Eje X:** Representa los valores SHAP, indican el grado de influencia de cada característica en las probabilidades logarítmicas
- **Eje Y:** Muestra el nombre de las características ordenadas de mayor a menor importancia, de arriba hacia abajo
- **Color de los puntos:** Refleja el valor de la característica correspondiente, donde el rojo indica valores altos y el azul valores bajos

### Gráficos de Dependencia Parcial (dependence plots)

Un gráfico de dependencia (figura 12) es un tipo de gráfico de dispersión que muestra cómo las predicciones del modelo se ven afectadas por una característica específica, permitiendo observar las diferentes relaciones y posibles interacciones entre esa característica y las predicciones. También sirven para identificar interacciones entre múltiples características de un mismo modelo.

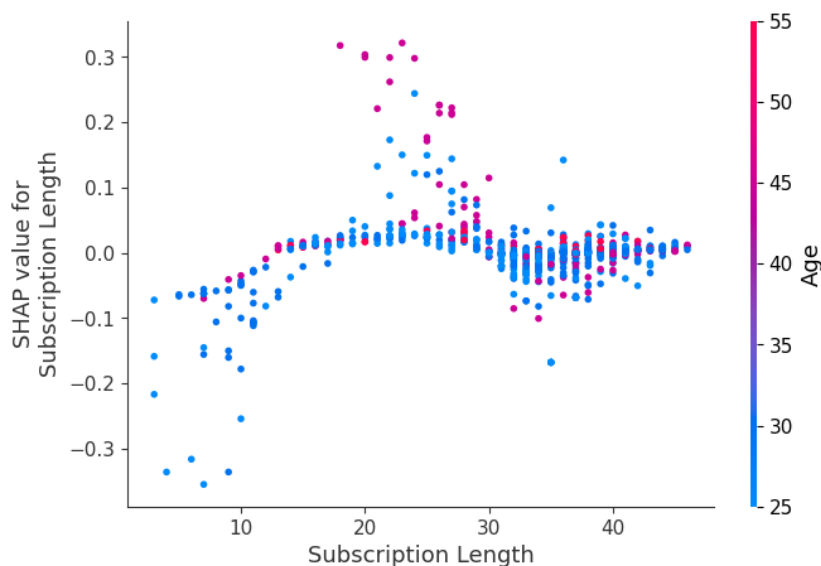


Figura 12: Ejemplo de Gráfico de Dependencia [47]

En el ejemplo proporcionado en la figura 12 se puede observar cómo se visualiza un gráfico de dependencia:

- **Eje X:** Representa la característica específica que se está analizando, en este caso “Subscription Length” (Duración de la suscripción).
- **Eje Y:** Muestra los valores SHAP correspondientes a la característica e indica cuánto influye en la predicción del modelo.
- **Colores de los puntos:** Los puntos de este tipo de gráficos están coloreados según otra característica del modelo, en este caso “Age” (Edad). De este modo se puede ver como interactúan las dos características.



---

## 5. Descripción de la solución propuesta

En este punto se detallará la solución propuesta para identificar qué factores influyen en la satisfacción y el rendimiento de los estudiantes con el objetivo de potenciar la sostenibilidad y el bienestar en entornos educativos. La solución se ha dividido en tres partes principales:

1. Tratamiento de los datos.
2. Desarrollo de modelos de machine learning.
3. Implementación de un cuadro de mando en Power BI.

De esta manera, se entenderá de una forma más clara el proceso seguido hasta alcanzar las conclusiones.

### 5.1 Preprocesamiento y limpieza de los datos

Como se ha comentado anteriormente, para la elaboración de este estudio se cuenta con datos de tres sensores colocados en el aula 3301 de la ETSISI y de unos cuestionarios realizados a los estudiantes al finalizar cada clase. Para poder implementar estos datos a los modelos de machine learning es necesario limpiarlos y prepararlos previamente.

#### 5.1.1 Datos de los Sensores

Los datos de los sensores ambientales fueron recolectados diariamente y almacenados en diferentes archivos CSV, de modo que por cada día se obtuvieron cinco archivos, uno para cada variable medida (temperatura, humedad, CO<sub>2</sub>, ruido y luminosidad). Para este estudio, se utilizaron los datos de noviembre y diciembre de 2023. Cada archivo CSV contiene dos columnas: la primera indica la hora en la que se tomó la muestra y la segunda el valor correspondiente a la medida recogida. Cabe destacar que los datos se recolectaron durante las 24 horas del día, con sensores tomando muestras cada cinco minutos, haciendo un total de 13.803 datos recolectados por los sensores.

Los archivos se entregaron en dos carpetas distintas, una para noviembre y otra para diciembre, aunque no se tienen muestras del día 1 al 5 de diciembre debido a un error que causó el apagado de los sensores.

#### Limpieza y transformación de los datos

Para asegurar que los datos que se van a emplear para generar los modelos no tienen ningún error, se deben corregir aquellos datos atípicos para evitar sesgos o posibles errores en la predicción. Se siguieron los siguientes pasos utilizando funciones de la librería “**Pandas**” del lenguaje de programación Python [49]:

##### 1. Agrupación de los CSVs:

Para utilizar los datos de los sensores en los modelos de predicción es necesario juntar todos los archivos CSVs explicados anteriormente, para ello se añadió una columna “Time” que combina la fecha extraída del nombre del archivo con la hora registrada en cada fila. Finalmente se unificaron todos los archivos en un mismo CSV.

## 2. Eliminación de Duplicados:

Se utilizó la función **`drop_duplicates()`** para eliminar los registros duplicados (en caso de que hubiera), ya que pueden distorsionar los resultados y necesitamos asegurar la máxima precisión en los datos.

## 3. Corrección de Formatos:

Se aseguró que la columna “Time”, en la que se encontraba la fecha y la hora de cada medida estuviera en el formato adecuado mediante la función **`pd.to_datetime()`**.

En la figura 13 se puede ver una parte del archivo “Sensores” tras las modificaciones comentadas:

```
Time, Temperature, CO2, Luminosity, Humidity, Noise
01-11-23 00:00, 25.198535919189453, 2671.0845947265625, 0.0, 46.74575805664063, 11.600191020965577
01-11-23 00:05, 25.182425181070965, 2650.85673421224, 0.0, 46.71780904134115, 11.649442005157471
01-11-23 00:10, 25.16213099161784, 2622.1601806640624, 0.0, 46.69873555501302, 11.762839031219482
01-11-23 00:15, 25.13618469238281, 2602.0242919921875, 0.0, 46.6851806640625, 11.69280792872111
01-11-23 00:20, 25.11028289794922, 2577.66806640625, 0.0, 46.66468302408854, 11.633920987447103
01-11-23 00:25, 25.091590881347656, 2550.6621988932293, 0.0, 46.6363525390625, 11.65589057604472
01-11-23 00:30, 25.070006052653, 2525.8310099283854, 0.0, 46.61654154459635, 11.660097964604695
```

Figura 13: Dataset de los Sensores

**Dimensión del Dataset *Sensores*:** 13804 filas y 6 columnas

### 5.1.2 Datos de los Cuestionarios

Las encuestas fueron rellenas por los estudiantes al terminar cada clase y se almacenaron en dos archivos Excel, uno para las encuestas de turno de mañana (clases de 9:00 a 13:00) y otro para el turno de tarde (clases de 15:00 a 19:00). Las preguntas de las encuestas se desarrollaron en la sección “Metodología” y capturan información sobre las percepciones de los estudiantes respecto a las condiciones ambientales del aula y de su propio estado físico o anímico. Se recolectaron un total de 384 encuestas.

#### Limpieza y transformación de los datos

Al tener dos archivos Excel para los cuestionarios, el primer paso fue juntarlos en uno solo para obtener una sola base de datos con las encuestas. Esta agrupación se hizo directamente en Excel copiando los datos de ambos archivos (“Recuento3301M” y “Recuento3301T”) en un nuevo Excel al que se nombró “datos\_3301”.

El cuestionario incluía la fecha y hora en que el alumno lo había relleno y en columnas separadas la fecha, hora de inicio de clase y hora final de clase. Para que estos datos tuvieran el mismo formato que la columna “Time” del archivo de los sensores, se crearon dos nuevas columnas: “Inicio”, donde se registraba la fecha y la hora inicial de la clase y “Final”, con la fecha y la hora final de la clase. Estos nuevos cambios se realizaron en un nuevo archivo Excel llamado “Encuestas2”.

**Dimensionalidad del Dataset “Encuestas2”:** 385 filas y 27 columnas.

Se eliminó una de las columnas de la encuesta en la que se pedía al alumno identificar la temperatura que percibía en grados, ya que muchos no contestaron a la pregunta y por otra parte resulta irrelevante para este estudio.

### 5.1.3 Agrupación de los datos

Una vez se tenían los diferentes datasets de los sensores y de las encuestas por separado (“Sensores” y “Encuestas2”), se procedió a su unificación, para ello se utilizó nuevamente Python siguiendo estos pasos:

1. **Lectura de Datos:** Se cargaron los datos de sensores y de las encuestas utilizando las funciones `pd.read_csv()` y `pd.read_excel()` respectivamente.
2. **Conversión de Fechas y Horas:** Se convirtieron las columnas que incluían Fecha y Hora de ambos Datasets en el mismo formato ‘%d-%m-%y %H: %M’ usando la función `pd.to_datetime()` para que la unificación de ambos resulte más fácil.
3. **Filtrado de los datos por intervalos de tiempo:** Se filtraron los datos para comparar si la columna “Time” del archivo “Sensores” (fecha y hora de cada medición) caía entre las columnas “Inicio” y “Final” (intervalo de horario de cada clase) del archivo “Encuestas2”. De esta manera sólo se tienen en cuenta los datos de los sensores que han sido tomados durante las horas de clase.
4. **Unificación de los Datos:** Tras filtrar los datos, se unificaron utilizando la función `pd.concat()` y se guardaron en un archivo CSV al que se nombró como “datos\_combiandos3301”.

**Dimensionalidad del Dataset final “datos\_combiandos3301”:** 9673 filas y 32 columnas.

```
Time, Temperature, CO2, Luminosity, Humidity, Noise, Marca temporal, "Apellidos, Nombre:", Aula:, Fecha:, Hora Inicio Clase, Hora Fin Clase, Puesto del aula
en el que estás sentado:, Inicio, Final, ¿Están las luces apagadas?, ¿Están las cortinas abiertas?, La luminosidad durante la clase ha sido, Nivel de
ruido durante la clase, ¿Está el aire acondicionado encendido?, "En caso afirmativo, indica los grados si los puedes identificar:", ¿Está la
calefacción encendida?, ¿Están las ventanas abiertas?, La temperatura del aula durante la clase ha sido, ¿Se ha notado un ambiente cargado?, "En
cuanto a ocupación del aula, dirías que el aula está:", "En general, durante la clase, el aula me ha parecido...", La docencia impartida en la
clase me ha parecido buena:, Mi estado ANÍMICO durante la clase ha sido bueno, "Con respecto a tu estado ANÍMICO ... ¿Cómo te has sentido?
* Puedes seleccionar VARIAS respuestas", Mi estado FÍSICO durante la clase ha sido bueno, "Con respecto a tu estado FÍSICO... ¿Cómo te has
sentido?
* Puedes seleccionar VARIAS respuestas"
2023-11-07 09:00:00, 25.119272867838543, 613.8704201027199, 790.818115234375, 36.795835141782405, 34.96238009134928, 2023/11/07 9:29:28 a. m.
CET, Mario , 3301, 2023-11-07, 09:00:00, 11:00:00, Segunda fila mitad , 2023-11-07 09:00:00, 2023-11-07 11:00:00, No, 5.0, 5.5, No, No, No, 5, 3, 3, 5, Totalmente
de acuerdo, Totalmente de acuerdo, Motivado, Contento, Totalmente de acuerdo, Normal, Con hambre o sed, Con dolor de cuello, Con dolor de espalda
2023-11-07 09:05:00, 25.13931937839674, 632.4879667862602, 772.3608915909477, 36.38866258704144, 48.65076019452966, 2023/11/07 9:29:28 a. m.
CET, Mario , 3301, 2023-11-07, 09:00:00, 11:00:00, Segunda fila mitad , 2023-11-07 09:00:00, 2023-11-07 11:00:00, No, 5.0, 5.5, No, No, No, 5, 3, 3, 5, Totalmente
de acuerdo, Totalmente de acuerdo, Motivado, Contento, Totalmente de acuerdo, Normal, Con hambre o sed, Con dolor de cuello, Con dolor de espalda
```

Figura 14: Dataset Final

Como se comentó anteriormente, no se tienen muestras de los sensores del día 1 al 5 de diciembre de 2023. Aunque se podrían haber rellenado manualmente haciendo aproximaciones basadas en los días anteriores y posteriores, para este proyecto es necesario tener los datos exactos con el fin de tener resultados más firmes de la percepción estudiantil del aula. En consecuencia, los cuestionarios de esos días fueron descartados.

---

#### 5.1.4 Preparación de los datos para el entrenamiento de los modelos

Para proceder a la ejecución de los modelos de predicción, se deben tener todos los datos en el mismo formato. Los datos de los sensores están en formato numérico, por lo que no requieren una conversión adicional. Sin embargo, los datos de las encuestas se deben transformar para ser tratados adecuadamente.

##### Conversión de Preguntas Binarias

Las preguntas binarias se convirtieron en valores numéricos siguiendo la siguiente escala:

- **Sí:** 1
- **No:** 0
- **Parcialmente:** 0.5

##### Conversión de Preguntas en Escala Likert

Algunas preguntas de la encuesta ya estaban directamente en formato numérico, por lo que el alumno solo debía seleccionar un valor en el rango del 1 al 5. Estas se dejaron en el formato original. No obstante, en otras preguntas los encuestados debían escoger entre opciones que varían desde “Totalmente de acuerdo” hasta “Totalmente en desacuerdo”. Para estas últimas se utilizó la siguiente escala:

- **Totalmente de acuerdo:** 5
- **De acuerdo:** 4
- **Ni de acuerdo ni en desacuerdo:** 3
- **En desacuerdo:** 2
- **Totalmente en desacuerdo:** 1

##### Conversión de Preguntas sobre el Estado Anímico y Físico

Las preguntas cualitativas sobre el estado anímico y físico se trataron de una manera diferente. El proceso utilizado se llama **codificación one-hot** (one-hot encoding) [50], en el que primero se crea una columna para cada posible respuesta y luego se convierten los resultados a formato binario (1 o 0). Se marca un 1 en la columna correspondiente al estado seleccionado y un 0 en las que no se han escogido.

---

## 5.2 Modelos de Predicción

En este apartado se van a describir los pasos seguidos para desarrollar los diferentes modelos de aprendizaje automático que ayudarán a predecir las variables objetivo de este proyecto.

### 5.2.1 Selección de las variables objetivo

Para crear un modelo de predicción, se debe escoger previamente qué variables objetivo se desean predecir. Para seleccionarlás, es necesario tener en cuenta el objetivo de la investigación, que en este caso es identificar qué factores influyen en la satisfacción y el rendimiento de los estudiantes para potenciar la sostenibilidad y el bienestar en entornos educativos.

Las variables objetivo deben estar dentro del conjunto de datos que se utilizará para crear el modelo, en esta investigación es el Dataset “datos\_combinados3301” con datos de los sensores y de las encuestas.

Teniendo en cuenta el objetivo, las variables seleccionadas para crear un modelo que las prediga son las siguientes:

#### 1- En general, durante la clase, el aula me ha parecido...

Esta variable mide la percepción general de los estudiantes sobre el entorno del aula. Un ambiente del aula percibido positivamente puede relacionarse con la satisfacción y el bienestar de los estudiantes. Al evaluar esta variable, también se puede observar como las condiciones ambientales influyen en esta percepción. Por ejemplo, una temperatura muy alta podría estar relacionada con un valor bajo de esta variable indicando una percepción negativa del entorno.

#### 2- La docencia impartida en la clase me ha parecido buena

Se ha escogido esta variable objetivo para la segunda investigación, debido a que la calidad de la docencia puede ser un buen indicador sobre la satisfacción de los estudiantes. Si a los alumnos les parece buena la docencia impartida, puede mejorar su motivación y por tanto su rendimiento académico.

#### 3- Estados anímicos positivos: Contento, Motivado y Eufórico

Para evaluar los estados anímicos positivos de los estudiantes se han realizado tres modelos de predicción diferentes, uno para predecir la variable Contento, otro para Motivado y el tercero para Eufórico. Se han seleccionado estos estados ya que pueden ayudar a identificar las condiciones o factores que fomentan un entorno educativo positivo.

---

#### 4- Estados anímicos negativos: Agobiado, Estresado y con Ansiedad

Al igual que en el caso anterior, se han realizado tres modelos para los estados anímicos negativos. Es importante saber qué factores o que variables influyen en los alumnos que experimentan estas sensaciones, ya que pueden afectar gravemente a su rendimiento académico. Al identificar que les hace sentirse de esa manera, se podrían implementar estrategias para mitigar esos efectos negativos. Por ejemplo, si se encuentra que niveles de ruido elevados están relacionados con el estrés, se podrían hacer ajustes en el aula para mejorar estas condiciones.

##### 5.2.2 Modelos de predicción probados en la experimentación

Para entender como se ha ejecutado la parte técnica de la investigación, se van a detallar los pasos que se han seguido para desarrollar los modelos de predicción en Python:

###### 1. Importación de librerías y Carga de Datos:

Se utilizarán la librería de Python, en concreto: *pandas*, *numpy*, *scikit-learn*, *scipy*, *shap* y *matplotlib*.

###### 2. Preprocesamiento del conjunto de datos

###### 3. Definición de las variables:

- Las variables ambientales de los sensores y las respuestas de las encuestas se definen como variables independientes (X).
- Las variables objetivo funcionan como variables dependientes (Y).

Es importante tener en cuenta que la variable objetivo no debe estar en el conjunto de características de entrenamiento.

###### 4. División de los datos en conjuntos de entrenamiento y prueba

**5. Normalización de los datos:** Se normalizan las variables independientes utilizando *StandardScaler()* para poder asegurar que todas las variables tienen una escala comparable.

**6. Ajuste de Hiperparámetros:** Los hiperparámetros se ajustarán haciendo uso de las funciones *GridSearchCV()* y *RandomizedSearchCV()*.

###### 7. Evaluación del modelo

Se ha decidido emplear modelos de predicción de diferentes tipos con el objetivo de analizar sus cualidades y ver cuál es el que mejor predicción es capaz de realizar para el conjunto de datos descrito anteriormente, se hará una comparación de los resultados con Lasso, Ridge y RandomForest.

---

## Modelos de predicción finales

En este subapartado se desarrollarán los diferentes pasos realizados hasta obtener los modelos finales, los cuales se explicarán más detalladamente en el apartado “Resultados”, donde también se podrán ver los datos finales obtenidos.

En primer lugar, se hizo un modelo cuya variable objetivo era “Mi estado anímico durante la clase ha sido bueno”. Se probaron diferentes modelos de predicción como Lasso, Ridge, Bagging y Randomforest, obteniendo este último los mejores resultados. Este modelo se descartó a pesar de dar buenos resultados debido a que se consideró mejor investigar los estados anímicos positivos y negativos por separado.

Por lo tanto, se han considerado finalmente los siguientes modelos como los modelos con mayor precisión y serán los que van a ser empleados para mostrar en el cuadro de mando:

### Modelo 1: Confort Ambiental

- **Variable Objetivo:** En general, durante la clase, el aula me ha parecido...
- **Algoritmo empleado:** RandomForest, Lasso, Ridge y Bagging RandomForest
- **Objetivo de predicción:** Evaluar como las condiciones ambientales influyen en la percepción del aula de los estudiantes.
- **Variables empleadas para la predicción:** Se realizaron dos modelos diferentes:
  1. Modelo 1 con estados físicos
  2. Modelo 1 sin estados físicos

Los estados físicos considerados son: Enérgico/a, Normal, Agotado/a, Enfermo/a, Con hambre o sed, Con dolor de cuello, Con dolor de cabeza, Con dolor de espalda y Con sueño.

- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Descripción de la fase de entrenamiento:** Para la fase de entrenamiento se fueron ajustando los hiperparámetros iterativamente para optimizar los resultados y escoger los más precisos.

Se evaluaron los resultados utilizando las métricas descritas en el apartado “Metodología” ( $R^2$ , MAE, RMSE y validación cruzada) además se compararon los diferentes modelos de predicción.

### Modelo 2: Percepción de la Docencia

- **Variable Objetivo:** La docencia impartida en la clase me ha parecido buena
- **Algoritmo empleado:** RandomForest
- **Objetivo de predicción:** Evaluar como las condiciones ambientales y los estados anímicos y físicos de los estudiantes influyen en su percepción de la docencia
- **Variables empleadas para la predicción:** Se realizaron dos modelos diferentes:
  1. Modelo 2 con estados físicos
  2. Modelo 2 sin estados físicos

---

Los estados físicos considerados son: Enérgico/a, Normal, Agotado/a, Enfermo/a, Con hambre o sed, Con dolor de cuello, Con dolor de cabeza, Con dolor de espalda y Con sueño.

- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Descripción de la fase de entrenamiento:** Para este modelo se realizaron nuevamente ajustes en los hiperparámetros hasta obtener los resultados más robustos. Sin embargo, se requirieron menos pruebas que en el modelo 1.

### **Modelo 3: Estados anímicos Positivos**

- **Variable objetivo:** Estados anímicos positivos
- **Algoritmo empleado:** RandomForest
- **Objetivo de predicción:** Evaluar qué factores ambientales y anímicos afectan en los estados anímicos positivos de los alumnos
- **Variables empleadas para la predicción:** Se realizaron tres modelos diferentes:
  1. Predicción de la variable Contento
  2. Predicción de la variable Motivado
  3. Predicción de la variable Eufórico
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Descripción de la fase de entrenamiento:** Para este modelo se realizaron diferentes pruebas, se evaluaron los resultados incluyendo las variables físicas, sin incluirlas y eliminando las variables menos importantes tras visualizar las gráficas de importancia de características RandomForest. Finalmente, los resultados más precisos fueron para los modelos Contento y Motivado excluyendo las variables físicas, descartando por tanto el modelo que predice la variable Eufórico.

### **Modelo 4: Estados anímicos Negativos**

- **Algoritmo empleado:** RandomForest
- **Objetivo de predicción:** Evaluar como las condiciones ambientales y los estados anímicos influyen en los estados anímicos negativos de los estudiantes.
- **Variables empleadas para la predicción:** Se realizaron tres modelos diferentes:
  1. Predicción de la variable Estresado
  2. Predicción de la variable Agobiado
  3. Predicción de la variable Con Ansiedad
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Descripción de la fase de entrenamiento:** Para este modelo se realizaron diferentes pruebas, se evaluaron los resultados incluyendo las variables físicas, sin incluirlas y eliminando las variables menos importantes tras visualizar las gráficas de importancia de características RandomForest. Finalmente, los resultados más precisos fueron para los modelos Estresado y Agobiado excluyendo las variables físicas, descartando por tanto el modelo que predice la variable Con Ansiedad.



## 5.3 Cuadro De Mando Power BI

Para este proyecto se ha realizado un cuadro de mando interactivo en Power BI en el que se pueden visualizar los resultados de los cuatro modelos mencionados y sus respectivas variantes. También se ha realizado una evaluación de los datos ambientales recolectados por los sensores.

En este apartado se detallarán los pasos necesarios para generar el cuadro de mando y los diferentes elementos utilizados.

### 5.3.1 Desarrollo del cuadro de mando

Para crear un cuadro de mando en Power BI desktop se debe tener el conjunto de datos previamente preparado, esto incluye la limpieza de los datos y la transformación de estos para que estén en el formato adecuado para el análisis.

#### Preparación de los Datos

Para mostrar las métricas de los sensores se ha utilizado el conjunto de datos descrito anteriormente “datos\_combinados3301”, que incluye los datos de los sensores y las percepciones de los estudiantes. Se ha tenido que transformar los datos del CSV debido a que Power BI no era capaz de entender los decimales. Para ello se reemplazaron los puntos del CSV por comas en aquellas columnas que contenían valores numéricos decimales haciendo uso de Python.

Para mostrar los cuatro modelos de predicción, se han guardado los resultados finales ( $R^2$ , MAE, RMSE y validación cruzada) junto con la media de los valores absolutos de las 5 características SHAP más importantes de cada modelo en un archivo CSV. Ambos resultados se han combinado en un mismo dataset para facilitar su implementación en Power BI.

Por ejemplo, en la figura 15 se puede ver el dataset final del Modelo 1, con las cinco características SHAP más importantes para el modelo con estados físicos y sin estados físicos y sus respectivos resultados:

```
característica,importancia,estados_fisicos,R2_Train,R2_Test,MAE_Train,RMSE_Train,MAE_Test,RMSE_Test,Cross_Val_R2_Mean
¿Se ha notado un ambiente cargado?,"0,167675",con estados físicos,"0,995989","0,988249","0,027358","0,053209","0,044166","0,090305","0,985955"
La temperatura del aula durante la clase ha sido,"0,140887",con estados físicos,"0,995989","0,988249","0,027358","0,053209","0,044166","0,090305"
La luminosidad durante la clase ha sido,"0,105628",con estados físicos,"0,995989","0,988249","0,027358","0,053209","0,044166","0,090305","0,985955"
La docencia impartida en la clase me ha parecido buena:,"0,081153",con estados físicos,"0,995989","0,988249","0,027358","0,053209","0,044166","0,090305","0,985955"
Mi estado ANÍMICO durante la clase ha sido bueno,"0,058519",con estados físicos,"0,995989","0,988249","0,027358","0,053209","0,044166","0,090305"
¿Se ha notado un ambiente cargado?,"0,301532",sin estados físicos,"0,996924","0,988321","0,014479","0,046601","0,027780","0,090027","0,986542"
La temperatura del aula durante la clase ha sido,"0,236558",sin estados físicos,"0,996924","0,988321","0,014479","0,046601","0,027780","0,090027"
La docencia impartida en la clase me ha parecido buena:,"0,110205",sin estados físicos,"0,996924","0,988321","0,014479","0,046601","0,027780","0,090027","0,986542"
"En cuanto a ocupación del aula, dirías que el aula está:","0,070497",sin estados físicos,"0,996924","0,988321","0,014479","0,046601","0,027780","0,090027","0,986542"
La luminosidad durante la clase ha sido,"0,062246",sin estados físicos,"0,996924","0,988321","0,014479","0,046601","0,027780","0,090027","0,986542"
```

Figura 15: Ejemplo Dataset Modelo 1 para Power BI

Una vez los datos estaban preparados se procedió a introducirlos en Power BI Desktop. Se generó un informe nuevo en blanco y se importaron los datos del CSV utilizando la opción “Agregar datos al informe” (figura 16) que aparecía en el menú principal de la aplicación.

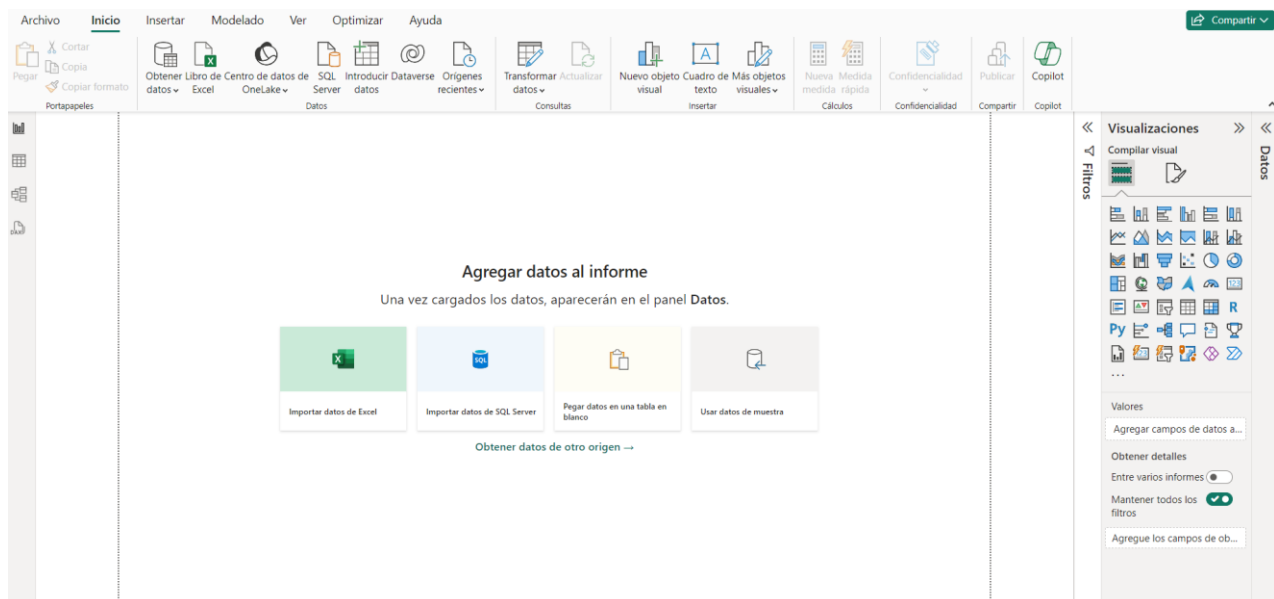


Figura 16: Menú Principal Power BI

## Diseño del Cuadro de Mando

Una vez se tienen los datos cargados en la aplicación, se procede a realizar el diseño del Cuadro de mando, que en este proyecto cuenta con 5 páginas diferentes que serán descritas en el apartado de “Resultados”.

Para la creación de un cuadro de mando se utilizan diferentes objetos visuales para mostrar los resultados de manera que sean legibles y fáciles de entender para el usuario. La selección de los objetos visuales se realiza en el lado derecho de la pantalla, donde se encuentra el panel de visualizaciones. En ese panel (Figura 17) se incluyen los diferentes objetos disponibles y la posibilidad de descargar más si fuera necesario.

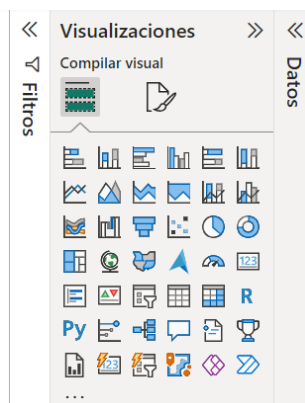


Figura 17: Panel de Visualizaciones Power BI

Los objetos visuales utilizados en este estudio son los siguientes:

- **Gráficos de Líneas**

- **Descripción:** Este tipo de gráficos muestran la evolución de una variable continua o tendencias a lo largo del tiempo.
- **Uso en el proyecto:** En este proyecto se utilizarán para visualizar como cada variable ambiental recogida por los sensores varía durante el tiempo. Muestran el promedio diario de CO<sub>2</sub>, luminosidad, ruido, temperatura y humedad a lo largo de los dos meses estudiados.
- **Ejemplo visual:** En la Figura 18 se muestra un ejemplo del icono del objeto visual que se debe seleccionar en Power BI para utilizar los gráficos de líneas junto con la tarjeta que muestra el promedio de humedad a lo largo del tiempo.

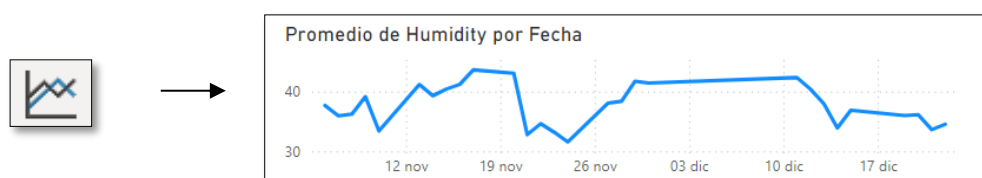


Figura 18: Ejemplo Gráfico de Líneas Power BI

- **Tarjetas de Métricas**

- **Descripción:** Las tarjetas métricas sirven para destacar diferentes indicadores de manera simple y visual.
- **Uso en el proyecto:** En este cuadro de mando se utilizan para mostrar el promedio total de las variables ambientales recolectadas durante los meses de noviembre y diciembre de 2023. A su vez, también se utilizan para mostrar los resultados de las pruebas de cada modelo de modo en un simple vistazo se pueda comprobar la precisión del modelo. Se presentan métricas de R<sup>2</sup>, MAE, y RMSE para cada modelo de predicción.
- **Ejemplo visual:** En la Figura 19 se muestra un ejemplo del icono del objeto visual que se debe seleccionar en Power BI para utilizar las tarjetas junto con la tarjeta que muestra el promedio total de humedad.



Figura 19: Ejemplo Tarjeta Power BI

- **Segmentación de datos**

- **Descripción:** Este objeto visual sirve para añadir la parte interactiva al cuadro de mando
- **Uso en el proyecto:** En este cuadro de mando la segmentación de datos se utiliza para permitir al usuario elegir entre las diferentes variantes que tiene cada modelo. Por ejemplo, en el modelo 1 sale un desplegable en el que se puede elegir entre Modelo 1 con estados físicos o Modelo 1 sin estados físicos. Esto hace que sea fácil la comparación y el análisis de los resultados.
- **Ejemplo visual:** En la Figura 20 se puede observar un ejemplo del icono del objeto visual que se debe seleccionar en Power BI para agregar segmentación de datos junto con el ejemplo del modelo 1 mencionado anteriormente.

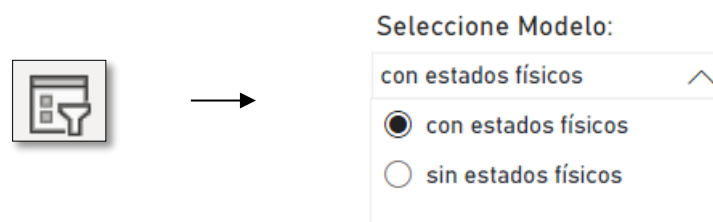


Figura 20: Ejemplo Segmentación de Datos en Power BI

- **Gráfico de Barras Agrupadas**

- **Descripción:** Permiten comparar series de datos entre diferentes categorías.
- **Uso en el proyecto:** En este cuadro de mando, las gráficas de barras agrupadas se utilizan para mostrar el impacto de las cinco características SHAP más importantes de cada modelo con el fin de poder visualizar en que grado afectan en la predicción.
- **Ejemplo visual:** En la Figura 21 se puede observar un ejemplo del icono del objeto visual que se debe seleccionar en Power BI para agregar el gráfico de barras agrupadas junto con el ejemplo descrito anteriormente.

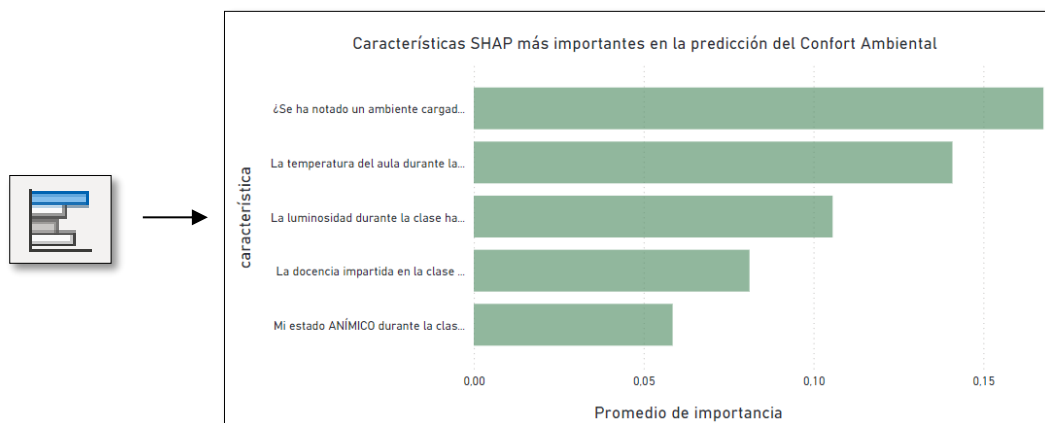


Figura 21: Ejemplo Gráfico de Barras Agrupadas en Power BI

---

Además de los mencionados objetos visuales, también se han añadido Botones para crear un menú principal, y se ha utilizado el objeto visual de Python integrado en Power Bi para realizar diferentes pruebas que finalmente no forman parte del cuadro de mando definitivo.



---

## 6. Resultados

En este apartado se van a describir las diferentes pruebas que se han realizado en cada modelo hasta llegar a los cuatro modelos definitivos. Además, se mostrarán los diferentes resultados y las gráficas obtenidas para la toma de decisiones. Por último, se desarrollará el cuadro de mando final en el que se podrán visualizar los diferentes modelos de predicción y los datos recolectados por los sensores.

### 6.1 Pruebas Iniciales

Antes de comenzar a construir los modelos de predicción empleando algoritmos con mayor complejidad se realizó una prueba inicial utilizando un análisis de correlación lineal presentado en un mapa de calor. Sin embargo, los resultados mostraron que apenas había correlaciones significativas entre las características, a su vez entre las variables de tipo numérico tampoco había correlación. Esto se debe a que en este estudio se tienen datos cuantitativos (de los sensores) y cualitativos (percepciones de los estudiantes), que no deben estar linealmente relacionados.

Debido a la baja correlación lineal, se procedió a probar modelos de regresión no lineal. A continuación, se muestran los diferentes modelos realizados:

### 6.2 Modelo 1: Confort Ambiental

Este modelo intenta predecir la variable **“En general, durante la clase, el aula me ha parecido...”** con el fin de evaluar como las condiciones ambientales influyen en la percepción del aula de los estudiantes.

Se han realizado los modelos de predicción siguiendo los pasos descritos en el anterior apartado obteniendo:

#### 6.2.1 Modelo 1 con estados físicos

Se evalúa la variable objetivo incluyendo los estados físicos: Enérgico/a, Normal, Agotado/a, Enfermo/a, Con hambre o sed, Con dolor de cuello, Con dolor de cabeza, Con dolor de espalda y Con sueño.

##### Prueba 1

- **Algoritmo empleado:** RandomForest
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Hiperparámetros utilizados:** *RandomizedSearchCV()*
  - max\_depth: 30
  - max\_features: 'auto'
  - min\_samples\_leaf: 1
  - min\_samples\_split: 2
  - n\_estimators: 246

- **Resultados obtenidos:**

Métrica	Valor
R2 score (Test)	0.991668649276295
Mean Absolute Error (MAE)	0.01898886666330478
Root Mean Squared Error (RMSE)	0.07603766648950563

Tabla 1: Resultados Prueba 1 Modelo 1 con estados físicos

Se puede observar por el alto valor de R2 Y los bajos valores de los errores que el modelo predice bastante bien.

Para comprobar si presenta overfitting, se procede a añadir los resultados de prueba, entrenamiento y validación cruzada, de esta manera se puede hacer un análisis más completo de los resultados. Además, se ajustan los hiperparámetros para probar si se obtienen mejores valores.

## **Prueba 2**

- **Algoritmo empleado:** RandomForest
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Hiperparámetros utilizados:** Mediante *RandomizedSearchCV()* los parámetros más óptimos son:
  - max\_depth: 30
  - max\_features: log2
  - min\_samples\_leaf: 2
  - min\_samples\_split: 8
  - n\_estimators: 195
- **Resultados obtenidos:**

Métrica	Valor
R2 score (Train)	0.9959891740828941
R2 score (Test)	0.9882489276774349
Mean Absolute Error (Train)	0.02735847445452891
Mean Absolute Error (Test)	0.04416579541403714
Root Mean Squared Error (Train)	0.053209100739558576
Root Mean Squared Error (Test)	0.09030458919724703
Cross-validated R2 scores	[0.98666164, 0.98569775, 0.98585539, 0.98420266, 0.98735798]
Mean R2 score (Cross-validated)	0.9859550842322615

Tabla 2: Resultados Prueba 2 Modelo 1 con estados físicos



---

Para comprobar que el modelo no está sobreajustado hay que fijarse en los resultados de  $R^2$  del conjunto de entrenamiento (0.995989) y del conjunto de prueba (0.988249), este último valor sugiere que el modelo generaliza bien a datos no vistos. Hay una diferencia pequeña entre ambos resultados, lo que indica que el modelo no está muy sobreajustado.

En cuanto a los errores, deben de ser bajos para asegurar una buena predicción, en este caso el MAE del conjunto de entrenamiento es bajo (0.027358), por lo que el modelo predice valores cercanos a los reales en promedio. En el conjunto de prueba se ha obtenido un MAE ligeramente mayor (0.044166), sin embargo, sigue siendo un valor bajo y podemos decir que indica buenas predicciones. Los errores RMSE de entrenamiento y prueba han salido también bajos por lo que no se presentan signos de sobreajuste.

Por último, hay que fijarse en los resultados de validación cruzada, la media de ellos es (0.985955) que es bastante cercana al  $R^2$  del conjunto de prueba (0.988249), reafirmando que el modelo parece generalizar bien la predicción.

A pesar de que el modelo obtiene muy buenos resultados y podría ser el definitivo se decide hacer una prueba más comparándolo con otros modelos de predicción diferentes a RandomForest:

### **Prueba 3**

- **Algoritmo empleado:** RandomForest, Ridge, Lasso y Bagging.
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Hiperparámetros utilizados:**

**Lasso:** 'alpha': 0.01

**Ridge:** 'alpha': 10.0

**Bagging:** Mediante *GridSearchCv()* los parámetros más óptimos son:

- n\_estimators: 50
- max\_samples: 1.0
- max\_features: 1.0

**RandomForest:** Mediante *RandomizedSearchCV()* los parámetros más óptimos son:

- max\_depth: 30
- max\_features: log2
- min\_samples\_leaf: 2
- min\_samples\_split: 8
- n\_estimators: 195

- **Resultados obtenidos:**

Modelo	R2 Score (Test)	Mean R2 Score (Cross- validated)	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
Random Forest	0.988248927677	0.985955084232	0.0441657954140	0.0903045891972
Bagging	0.987961209028	0.988228661226	0.0232675121699	0.091403433229
Ridge	0.616260618688	N/A	N/A	N/A
Lasso	0.61537043354	N/A	N/A	N/A

Tabla 3: Resultados Prueba 3 Modelo 1 con estados físicos

- **Justificación del modelo seleccionado:**

Una vez se tienen todos los resultados de esta prueba, se procede a su análisis detallado para seleccionar el mejor modelo.

#### **Coeficiente de Determinación ( $R^2$ )**

Los valores de  $R^2$  para los algoritmos Lasso (0.6153) y Ridge (0.61626) son bastante bajos en comparación con los demás, por lo que estos dos algoritmos quedan descartados.

El algoritmo RandomForest tiene el  $R^2$  más alto del conjunto de prueba (0.98824) seguido por el Bagging (0.98796) siendo ambos bastante altos.

#### **Errores MAE Y RMSE**

El error MAE es menor en el algoritmo Bagging (0.02326) que en el RandomForest (0.04416), por lo que las predicciones del algoritmo Bagging parecen estar más cerca de valores reales. En cuanto al RMSE, en Bagging (0.0914) es algo superior que en el RandomForest (0.09030), lo que describe que este último maneja mejor los errores grandes.

#### **Validación Cruzada**

La media de la validación cruzada de ambos modelos es bastante alta y próxima al valor de  $R^2$  del conjunto de datos de prueba, siendo algo superior en el modelo Bagging.

Aunque ambos algoritmos (Bagging y RandomForest) podrían ser seleccionados como finalistas debido a que tienen muy buenos resultados en el modelo, se selecciona el modelo RandomForest como el mejor, ya que tiene mayor precisión y maneja bien los errores.

## Modelo Final

Finalmente, el algoritmo seleccionado para representar al Modelo 1 con estados físicos es el modelo Randomforest descrito en la segunda prueba. Para comprobar que generaliza bien la predicción y no tiene un gran overfitting se sometió a dos pruebas adicionales:

1. Se introdujo un conjunto de datos aleatorio y se sacaron predicciones que mostraban valores dentro del rango de la variable objetivo, sin mostrar valores atípicos. Cabe destacar que al no tener más datos que los utilizados para generar el modelo, se utilizaron estos mismos para hacer la prueba con datos aleatorios.
2. Se analizaron los gráficos de residuos del conjunto de prueba y entrenamiento:

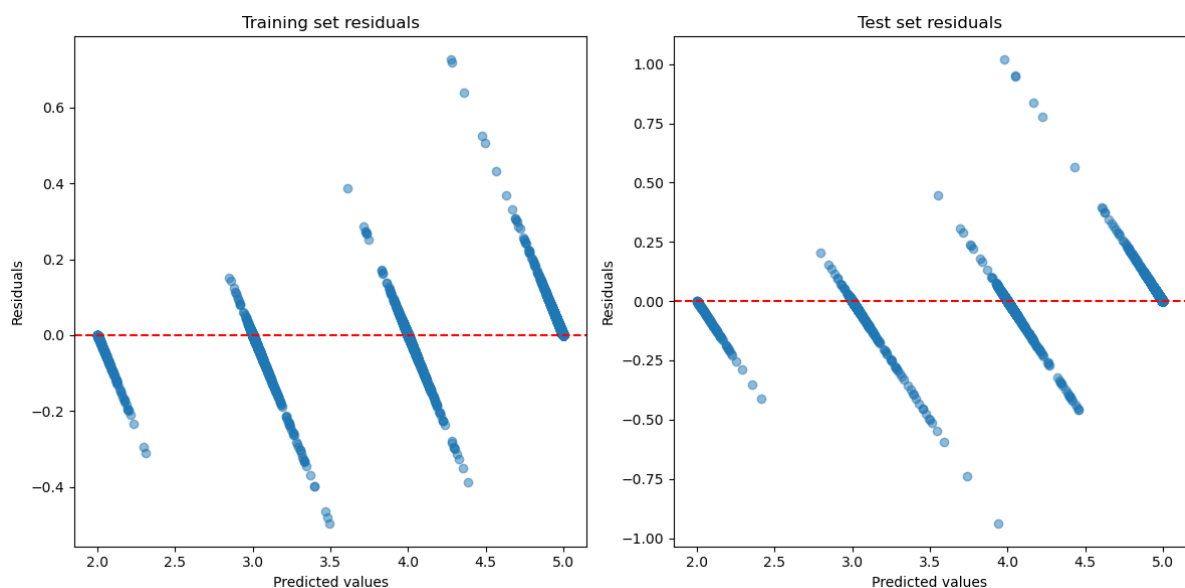


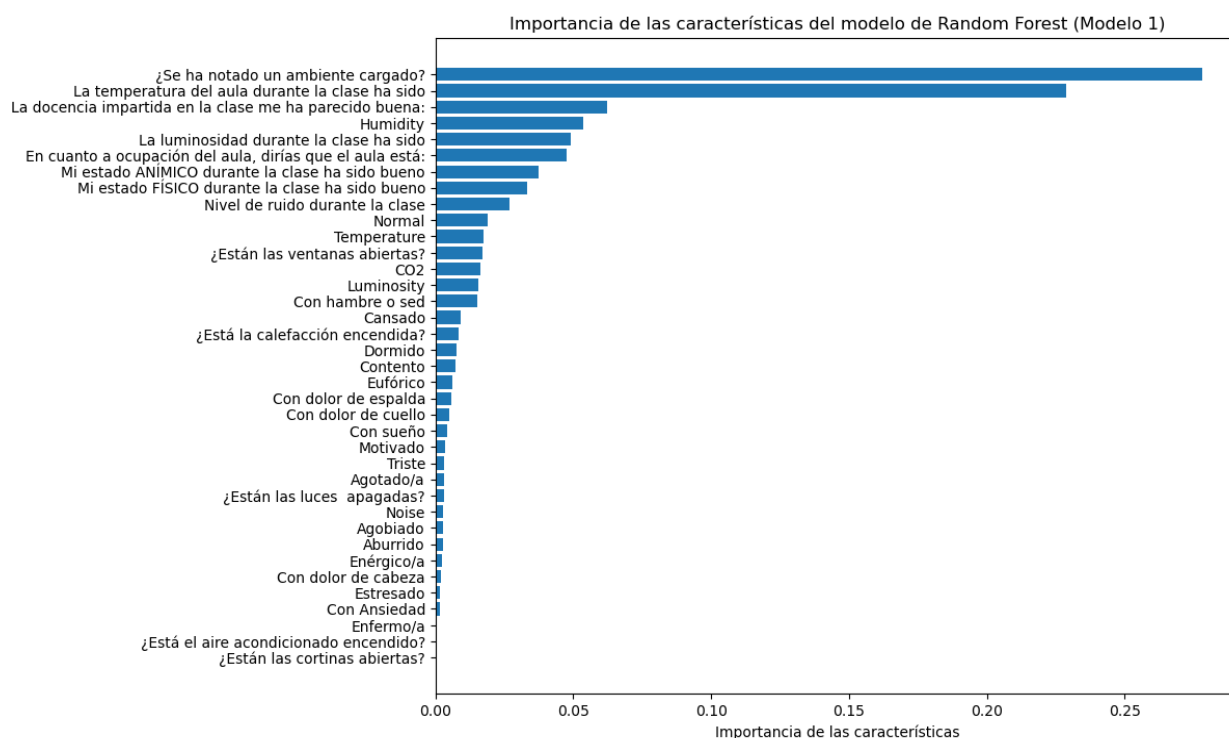
Figura 22: Gráfico de residuos Modelo 1 con estados físicos

Los residuos de ambos conjuntos de datos muestran una distribución cercana a 0, lo que indica que si se trata de un buen modelo. Sin embargo, los residuos en los extremos muestran algo de dispersión, indicando que quizás se estén haciendo peores predicciones en algunos casos específicos lo que podría ser un área de mejora.

A continuación, se muestran los resultados de la importancia de características representada en dos gráficas:

## Importancia de Características RandomForest

En la Figura 23 se puede ver la gráfica de barras correspondiente a la importancia de características del modelo 1 con estados físicos, en la cual se pueden apreciar todas las variables que participan en la predicción de la variable objetivo **“En general, durante la clase, el aula me ha parecido...”** en orden de importancia de arriba abajo.



**Figura 23: Importancia de Características Modelo 1 con estados físicos**

Las variables que más influyen en esta predicción son la percepción del ambiente cargado, de la temperatura del aula y de la docencia.

### Importancia de Características SHAP

En la Figura 24 se puede ver una gráfica SHAP con las características más importantes del modelo y como afectan individualmente al mismo, se van a comentar las más significativas:

- La percepción de un ambiente cargado tiene el mayor impacto en el modelo. Los valores altos (indicando un ambiente más cargado) tienen un impacto positivo en la predicción, por lo que los alumnos que perciben el ambiente más cargado tienden a tener una mejor percepción del aula. Lo mismo ocurre con la percepción de la temperatura, los alumnos prefieren una temperatura alta en el aula.
- Con la percepción de la luminosidad y de la docencia ocurre algo similar, los alumnos prefieren aulas bien iluminadas y cuanto mejor perciben la docencia más les gusta el aula. Esto tiene sentido, ya que los datos fueron tomados en invierno (cielos muy oscuros) y generalmente si no te sientes a gusto con la clase impartida, no te gustará tampoco el aula en la que te encuentras.
- En cuanto a las variables recogidas por los sensores, no toman un valor importante en el modelo.

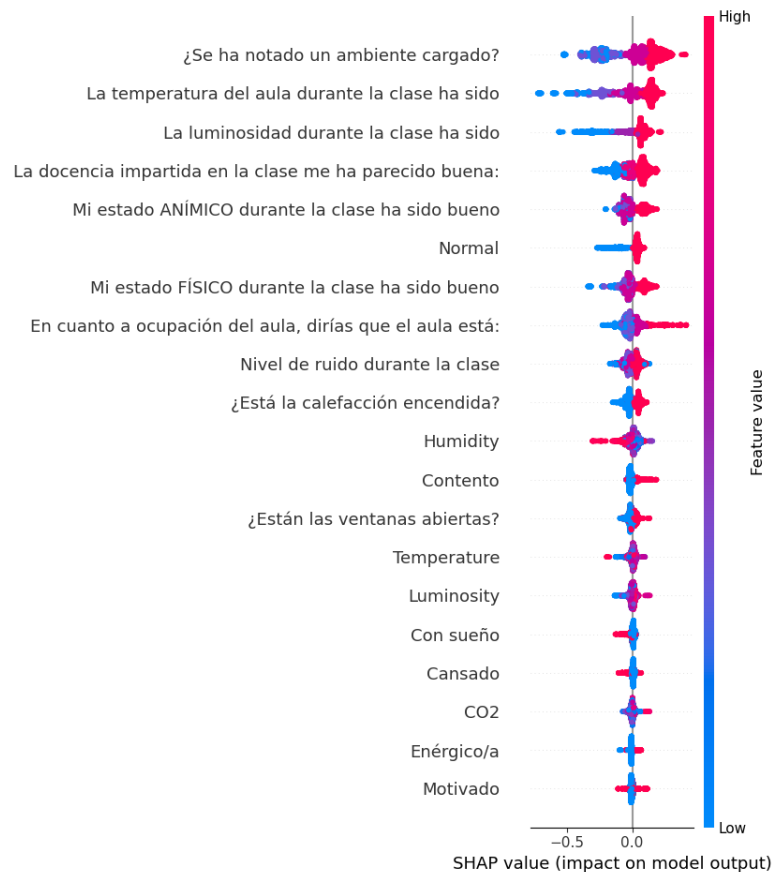


Figura 24: Gráfica Características SHAP modelo 1 con estados físicos

Dado que los estados físicos se encuentran abajo en la escala de importancia, se procede a evaluar el modelo sin ellos.

### 6.2.2 Modelo 1 sin estados físicos

A continuación, se evalúa la variable objetivo “**En general, durante la clase, el aula me ha parecido...**” excluyendo los estados físicos: Enérgico/a, Normal, Agotado/a, Enfermo/a, Con hambre o sed, Con dolor de cuello, Con dolor de cabeza, Con dolor de espalda y Con sueño.

#### Prueba 1

- **Algoritmo empleado:** RandomForest  
Debido a que, en los anteriores modelos de predicción, el algoritmo que ofrecía mejores resultados era el RandomForest, se utilizará para evaluar el resto de los modelos.
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Hiperparámetros utilizados:** *RandomizedSearchCV()*
  - max\_depth: 20
  - max\_features: 1.0
  - min\_samples\_leaf: 5
  - min\_samples\_split: 6
  - n\_estimators: 203

- **Resultados obtenidos:**

Métrica	Valor
R2 score (Train)	0.993979369636473
R2 score (Test)	0.984415953226372
Mean Absolute Error (Train)	0.023952976539395435
Mean Absolute Error (Test)	0.03768594009068343
Root Mean Squared Error (Train)	0.06519135308268469
Root Mean Squared Error (Test)	0.10399469684723739
Cross-validated R2 scores	[0.97899587, 0.9807396, 0.98398157, 0.9789632, 0.98363524]
Mean R2 score (Cross-validated)	0.9812630963913674

Tabla 4: Resultados prueba 1 Modelo 1 sin estados físicos

Los resultados de esta primera prueba nos indican que el modelo funciona bastante bien, la diferencia entre los valores de entrenamiento y de prueba de cada métrica es bastante baja, además, la media de validación cruzada (0.9812) tiene un valor similar al  $R^2$  de prueba (0.9844), lo cual nos indica que el modelo no parece tener un gran overfitting. Por otra parte, cabe destacar que los valores de  $R^2$  son elevados (explican gran parte de la variabilidad de conjunto de datos) y los errores son bajos. Este modelo generaliza bien la predicción, sin embargo, se va a ajustar los hiperparámetros para comprobar si se obtienen mejores resultados.

## **Prueba 2**

- **Algoritmo empleado:** RandomForest
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Hiperparámetros utilizados:** *RandomizedSearchCV()*
  - max\_depth: 20
  - max\_features: 1.0
  - min\_samples\_leaf: 2
  - min\_samples\_split: 8
  - n\_estimators: 149

- **Resultados obtenidos:**

En la Tabla 5 se pueden ver los resultados de este modelo, con un valor  $R^2$  test (0.9883) superior que en la prueba 1 (0.9844). La media de la validación cruzada es bastante próxima al  $R^2$  de prueba, por lo que el modelo parece tener menos sobreajuste. Por otra parte, los errores MAE y RMSE son más bajos. Al tener un  $R^2$  próximo a 1 y errores bastante bajos, podemos decir que el modelo realiza predicciones precisas.

Métrica	Valor
R2 score (Train)	0.996923578439321
R2 score (Test)	0.9883210427831663
Mean Absolute Error (Train)	0.014479355659489684
Mean Absolute Error (Test)	0.027780349957870026
Root Mean Squared Error (Train)	0.04660064588862445
Root Mean Squared Error (Test)	0.09002706783749649
Cross-validated R2 scores	[0.98477288, 0.98677715, 0.98854714, 0.98479117, 0.98782116]
Mean R2 score (Cross-validated)	0.9865419023849948

Tabla 5: Resultados prueba 2 Modelo 1 sin estados físicos

### Modelo Final

Finalmente, el modelo seleccionado para representar al Modelo 1 sin estados físicos es el modelo Randomforest descrito en la segunda prueba.

A continuación, se muestran los resultados de la importancia de características representada en dos gráficas:

### Importancia de Características RandomForest

En la figura 25 se pueden ver las características RandomForest en orden de importancia, de tal modo que se puede interpretar fácilmente que variables afectan más a la predicción, siendo en este caso la percepción del ambiente cargado y la percepción de la temperatura las más predominantes.

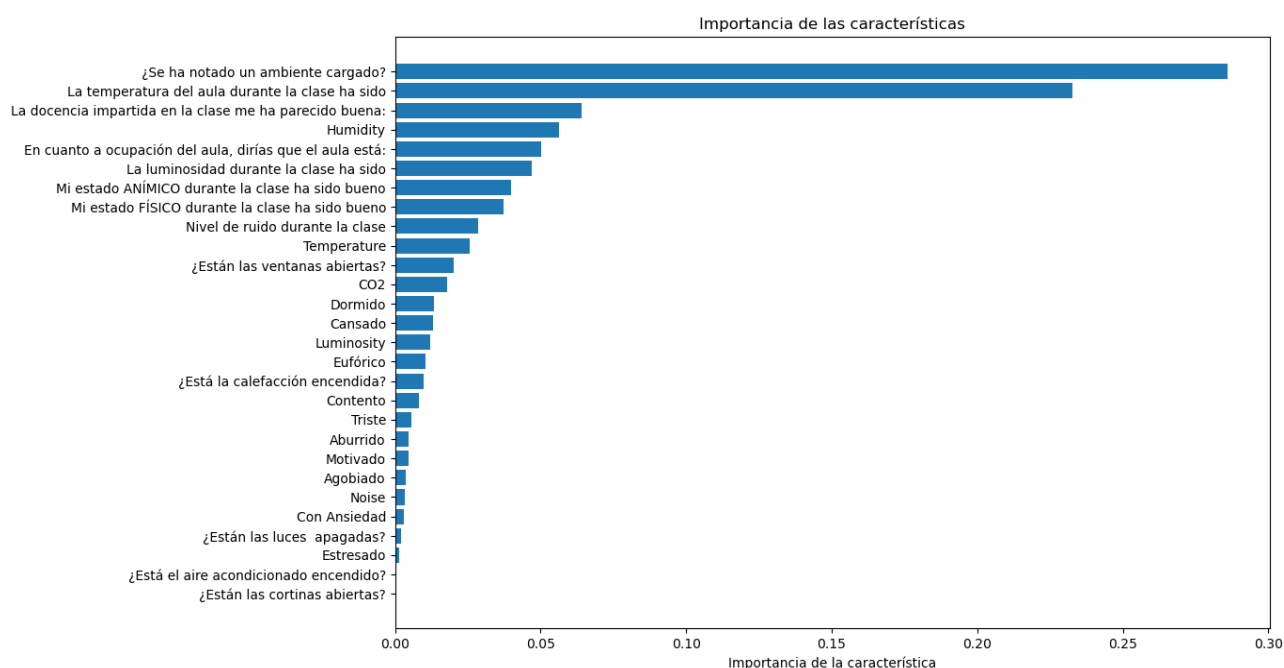


Figura 25: Características RandomForest Modelo 1 sin estados Físicos

## Importancia de Características SHAP

La gráfica de Características SHAP (Figura 26) es similar que la del modelo con las variables físicas, por lo que deducimos que estas no afectan apenas en la predicción.

Se ve una clara diferencia en el orden de importancias, en este caso la percepción de la docencia está en el tercer puesto y aparece la percepción de la ocupación del aula. Cuando los alumnos sienten que el aula esta más llena, les agrada más estar en ella, lo cual podemos relacionar con el ambiente cargado que ocupa el primer puesto.

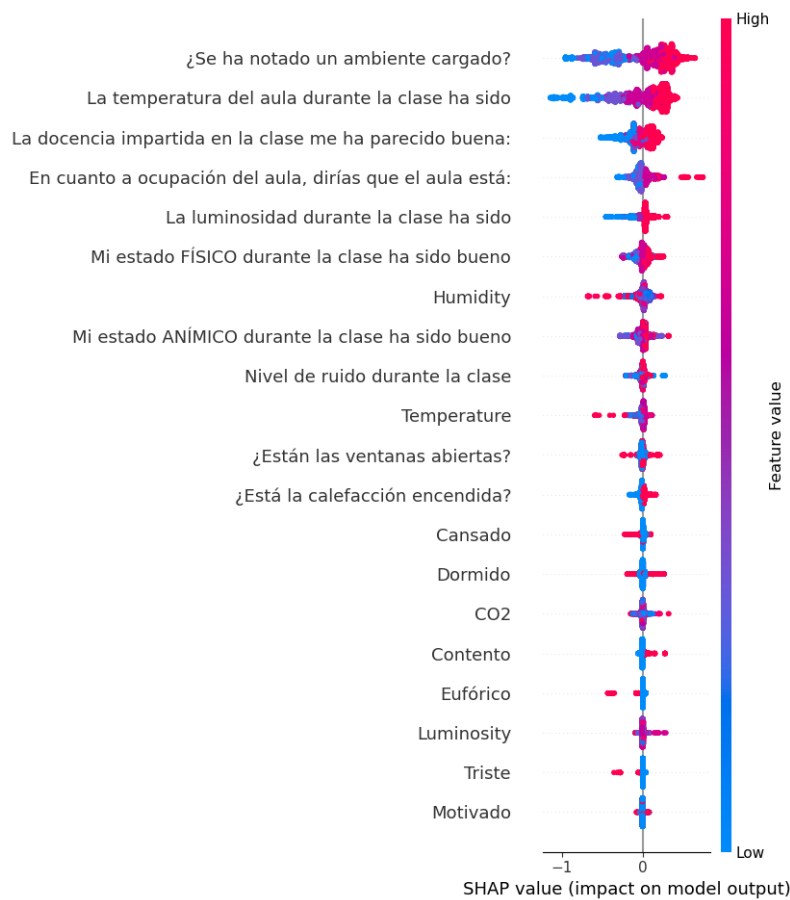


Figura 26: Características SHAP modelo 1 sin estados físicos



## 6.3 Modelo 2: Percepción de la docencia

Este modelo intenta predecir la variable “**La docencia impartida en la clase me ha parecido buena**” para intentar predecir que variables afectan a los alumnos sobre su percepción de calidad de la docencia durante las clases.

Los modelos de predicción se han desarrollado siguiendo los pasos descritos en el capítulo 5:

### 6.3.1 Modelo 2 con estados físicos

A continuación, se mostrarán las pruebas realizadas hasta obtener el modelo definitivo incluyendo los estados físicos: Enérgico/a, Normal, Agotado/a, Enfermo/a, Con hambre o sed, Con dolor de cuello, Con dolor de cabeza, Con dolor de espalda y Con sueño.

#### Prueba 1

- **Algoritmo empleado:** RandomForest
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Hiperparámetros utilizados:** *RandomizedSearchCV()*
  - max\_depth: 20
  - max\_features: 1.0
  - min\_samples\_leaf: 5
  - min\_samples\_split: 6
  - n\_estimators: 203
- **Resultados obtenidos:**

Los resultados generados por este modelo se pueden visualizar en la Tabla 6, donde el  $R^2$  de entrenamiento y prueba tienen poca diferencia entre ellos y son valores altos, esto indica que el modelo generaliza muy bien los datos. Comparando el  $R^2$  de prueba con la media de la validación cruzada vemos que no varían mucho los resultados, reafirmando que el modelo tiene buen rendimiento en diversos conjuntos de datos.

Para finalizar el análisis, los errores MAE Y RMSE, que miden la diferencia entre los valores observados y los predichos, son bajos. Junto con el alto valor de  $R^2$  podemos decir que este modelo generaliza bien las predicciones y no presenta signos de sobreajuste.

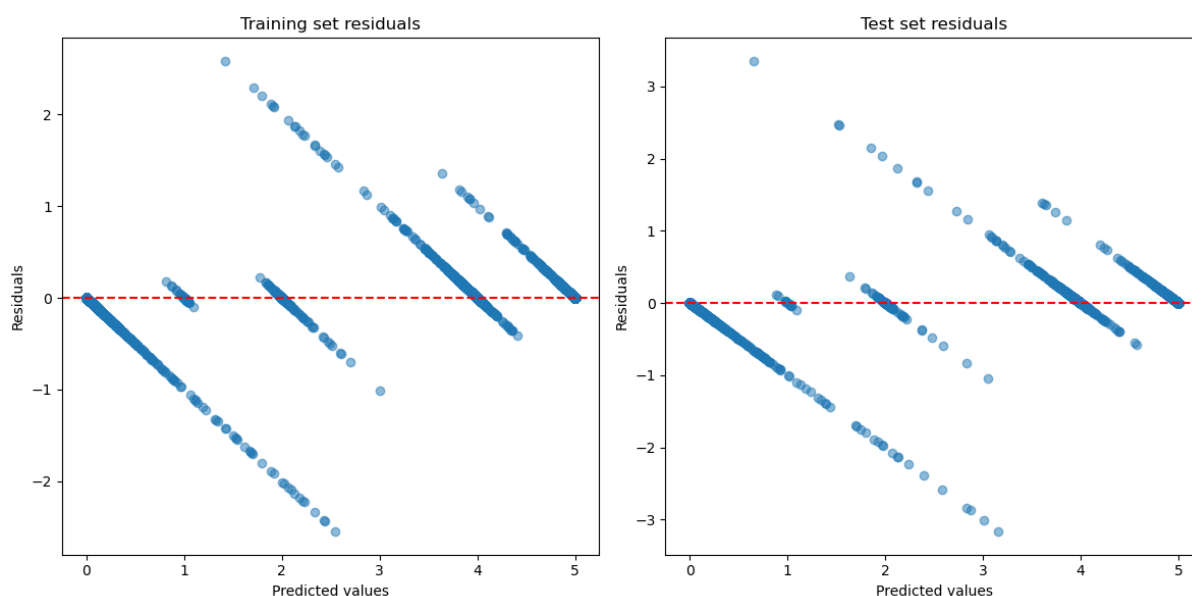
Métrica	Valor
R2 score (Train, new)	0.9898713056382135
R2 score (Test, new)	0.9816417453563154
Mean Absolute Error (Train, new)	0.06455633027106496
Mean Absolute Error (Test, new)	0.09479383584009791
Root Mean Squared Error (Train, new)	0.20607626350613673
Root Mean Squared Error (Test, new)	0.28253611553998265
Cross-validated R2 scores	[0.97066117, 0.97062054, 0.97598136, 0.97578729, 0.97391185]
Mean R2 score (Cross-validated)	0.9733924425215218

Tabla 6: Resultados Modelo 2 con estados físicos

## **Modelo final**

Una vez evaluada la prueba realizada anteriormente, se confirma que el modelo es robusto y se puede seleccionar para representar al modelo 2 con estados físicos.

Aunque el modelo parece funcionar muy bien, se hace una prueba adicional analizando sus residuos:



**Figura 27: Gráfica de Residuos Modelo 2 con estados físicos**

Se puede comprobar que los puntos están dispersos alrededor del valor 0, además no parece haber un patrón claro en los residuos, podemos decir que captura bien las relaciones. Sin embargo, podemos observar una línea que muestra una ligera tendencia hacia valores más bajos a medida que se aumentan los valores ajustados. Esto podría indicar un pequeño sesgo en las predicciones del modelo que podría considerarse un área de mejora.

## **Importancia de Características SHAP**

La gráfica de Características SHAP (Figura 28) nos muestra las características más influyentes en la percepción de la docencia en orden descendente.

- La variable que más influye en esta predicción es “Mi estado ANIMICO durante la clase ha sido bueno”, indicando que los alumnos que tienen buen estado anímico tienen muy buena percepción de la docencia.
- Al igual que en el modelo 1 la percepción de la luminosidad es muy influyente en la predicción. Cuando la luminosidad es adecuada, los alumnos perciben muy bien la docencia impartida, por lo contrario, cuando esta es inadecuada tienen una percepción negativa. Sin embargo, comparándolo con la luminosidad recogida por el sensor, esta última parece no afectar al modelo.

- Curiosamente, valores altos de un ambiente cargado tienen un impacto positivo, lo que sugiere que una percepción de un ambiente cargado no necesariamente afecta negativamente a la percepción de la docencia.
- Cabe destacar que en este modelo aparecen los estados anímicos como influyentes, por ejemplo “Aburrido”, que nos muestra que los alumnos que están aburridos perciben negativamente la docencia.
- En cuanto a las variables ambientales, se puede ver que no tienen mucha importancia, pero niveles altos de CO2 afectan negativamente a la predicción. El ruido no es influyente, ya que no aparece.

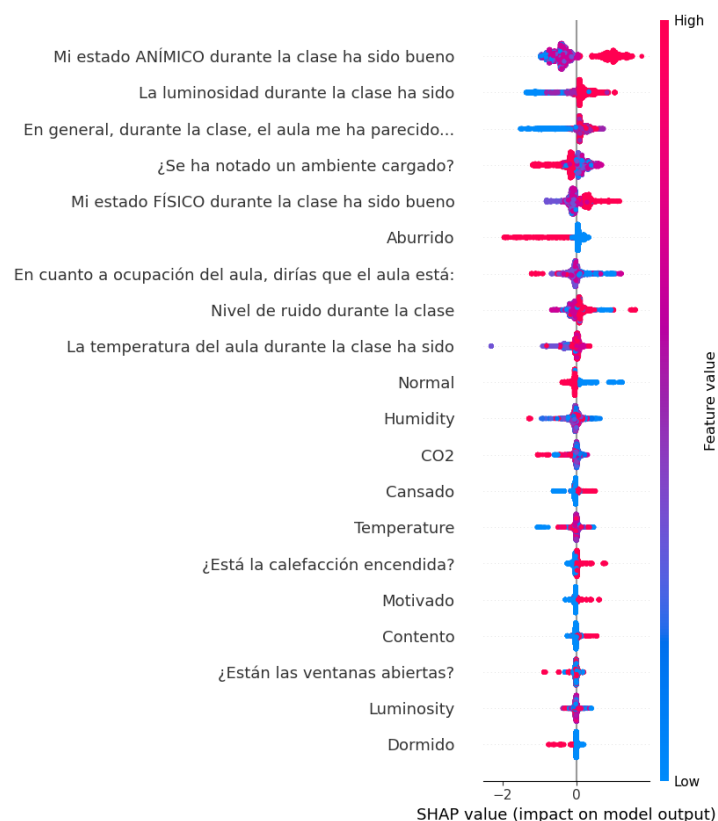


Figura 28: Gráfica Características SHAP modelo 2 con estados físicos

### Importancia de Características RandomForest

En la Figura 29 se pueden ver las características RandomForest del modelo en orden de importancia de arriba abajo. La variable que más afecta al modelo es la percepción del estado anímico de los estudiantes, con un grado muy superior a las demás. En segundo lugar, la percepción de la luminosidad. Se puede apreciar que los estados físicos no influyen en el modelo, por lo que es conveniente quitarlos para comprobar si se puede mejorar la eficiencia de este.

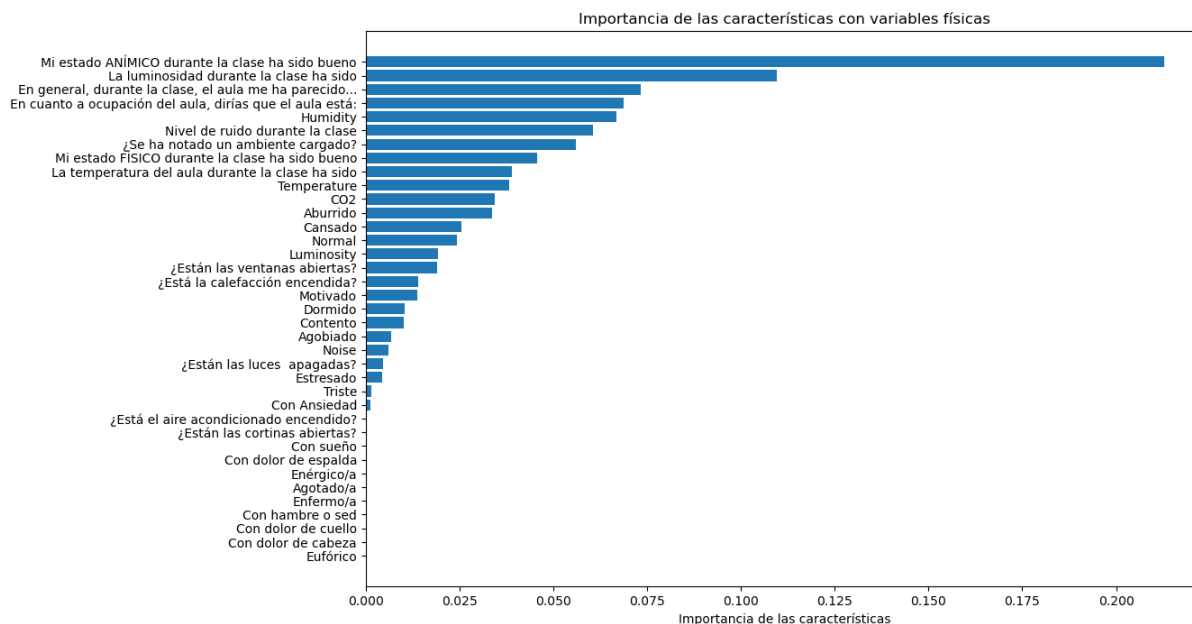


Figura 29: Características RandomForest Modelo 2 con estados Físicos

### 6.3.2 Modelo 2 sin estados físicos

A continuación, se evalúa la variable objetivo “La docencia impartida en la clase me ha parecido buena” excluyendo los estados físicos: Enérgico/a, Normal, Agotado/a, Enfermo/a, Con hambre o sed, Con dolor de cuello, Con dolor de cabeza, Con dolor de espalda y Con sueño. Ya que estos no tenían gran importancia en el modelo anterior.

#### Prueba 1

- **Algoritmo empleado:** RandomForest
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Hiperparámetros utilizados:** *RandomizedSearchCV()*
  - max\_depth: 20
  - max\_features: 1.0
  - min\_samples\_leaf: 5
  - min\_samples\_split: 6
  - n\_estimators: 203
- **Resultados obtenidos:**

Los resultados presentados en la Tabla 7 muestran que este modelo también funciona bastante bien. Tiene valores de  $R^2$  muy elevados y con poca diferencia entre los de prueba y entrenamiento. Los errores también son bajos y la media de validación cruzada es próxima al  $R^2$  de prueba, lo que indica que este modelo tiene buen rendimiento y generaliza bien sin tener signos evidentes de overfitting. En comparación con el anterior modelo, se puede decir que tienen un rendimiento similar, siendo ligeramente mejor el modelo que incluye los estados físicos. Aunque los estados físicos influían poco en el modelo, su inclusión no deteriora el rendimiento y puede contribuir a mejorar la robustez del modelo.

Métrica	Valor
R2 score (Train)	0.9898569992321855
R2 score (Test)	0.9815798124894358
Mean Absolute Error (Train)	0.06450385181293673
Mean Absolute Error (Test)	0.09483407829840772
Root Mean Squared Error (Train)	0.2062217496995262
Root Mean Squared Error (Test)	0.28301229203114264
Cross-validated R2 scores	[0.97085849, 0.97056391, 0.97623478, 0.97569301, 0.97368564]
Mean R2 score (Cross-validated)	0.9734071634429624

Tabla 7: Resultados Modelo 2 sin estados físicos

### Modelo Final

Ya que el modelo probado tiene muy buen rendimiento, se selecciona este para representar a la percepción de la docencia sin estados físicos.

### Importancia de Características SHAP

La importancia de características SHAP correspondiente a este modelo (Figura 30) es prácticamente igual que la del modelo 2 con estados físicos, por lo que excluirlos no ha cambiado la importancia de estas características.

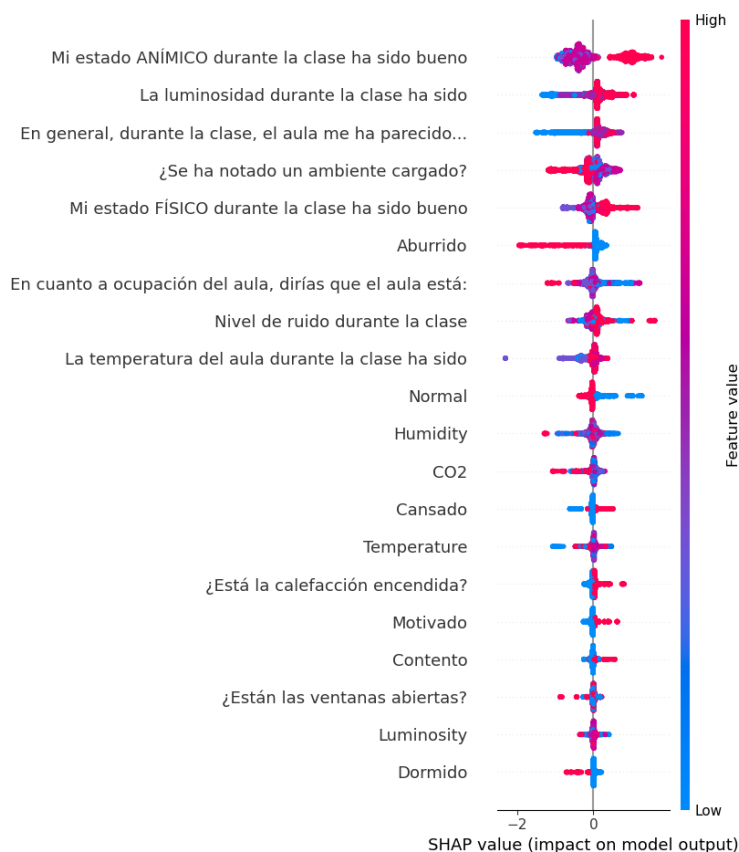


Figura 30: Gráfica Características SHAP modelo 2 sin estados físicos

## Importancia de Características RandomForest

A continuación, se ven representadas las características RandomForest (Figura 31) en orden de importancia, se puede observar que son muy similares al modelo con los estados físicos, sin embargo, en este caso, las variables ambientales percibidas por los sensores toman más importancia.

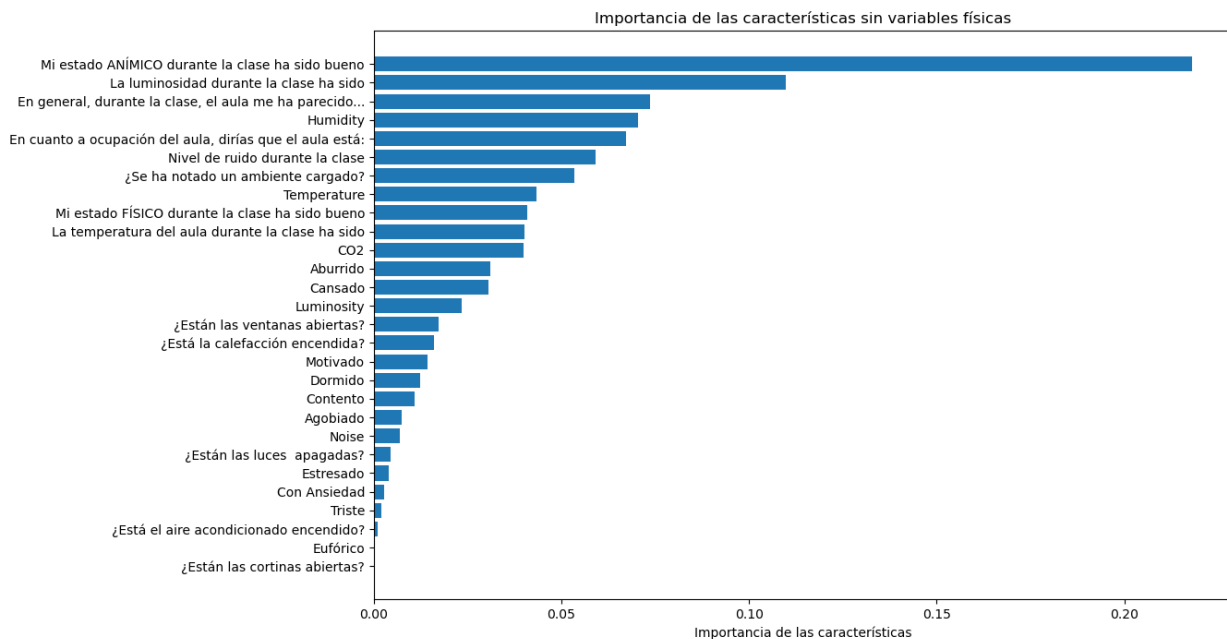


Figura 31: Características RandomForest Modelo 2 con estados Físicos

## 6.4 Modelo 3: Estados Anímicos Positivos

En este apartado se presentarán los modelos correspondientes a los estados anímicos positivos: Contento, Motivado y Eufórico, excluyendo los estados físicos, ya que estos no aportan información relevante para la investigación. Primero se compararán los resultados de rendimiento obtenidos, ya que se han evaluado con los mismos parámetros y finalmente se detallarán los resultados gráficos individualmente.

### Comparación resultados

- **Algoritmo empleado:** RandomForest
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Hiperparámetros utilizados:** *RandomizedSearchCV()*
  - max\_depth: 30
  - max\_features: log2
  - min\_samples\_leaf: 2
  - min\_samples\_split: 8
  - n\_estimators: 195

- **Resultados obtenidos:**

En la tabla 8 se pueden ver los resultados obtenidos tras realizar las predicciones, sin embargo, para evaluar más sencillamente los modelos se ha diseñado una gráfica comparativa (Figura 32)

Los modelos parecen tener un buen rendimiento, tienen valores altos de  $R^2$  próximos a la media de validación cruzada y errores relativamente bajos, lo que indica que generalizan bien los datos y pueden seguir siendo estudiados.

En la gráfica se puede ver que el modelo “Eufórico” tiene los menores valores de  $R^2$  en el conjunto de entrenamiento, de prueba y de validación cruzada, lo que indica que explica menos la variabilidad de los datos. Aunque sus errores son menores, el hecho de tener un coeficiente de determinación peor hace que sea descartado y no se analizará en detalle posteriormente.

Métrica	Contento	Motivado	Eufórico
R2 score (Train)	0.9657681335989	0.94888340389925	0.9336034783350
R2 score (Test)	0.9143963141652	0.87177907876160	0.8669358865894
MAE (Train)	0.03677540026225	0.04080614542813	0.01042746120275
MAE (Test)	0.05762631667527	0.06481576794626	0.01411724734580
RMSE (Train)	0.07501282279696	0.09110521249866	0.04508274532617
RMSE (Test)	0.11646914287140	0.14014433490111	0.06185839038974
Cross-validated R2 scores	[0.893410, 0.912513, 0.912903, 0.88857, 0.90997]	[0.863228, 0.890107, 0.891232, 0.847783, 0.858256]	[0.799973, 0.820698, 0.824700, 0.821492, 0.846534]
Mean R2 score (Cross-validated)	0.9034743786350	0.8701215323529	0.8226794636331

**Tabla 8: Resultados Modelo 3**

## Gráfica Comparativa Modelos

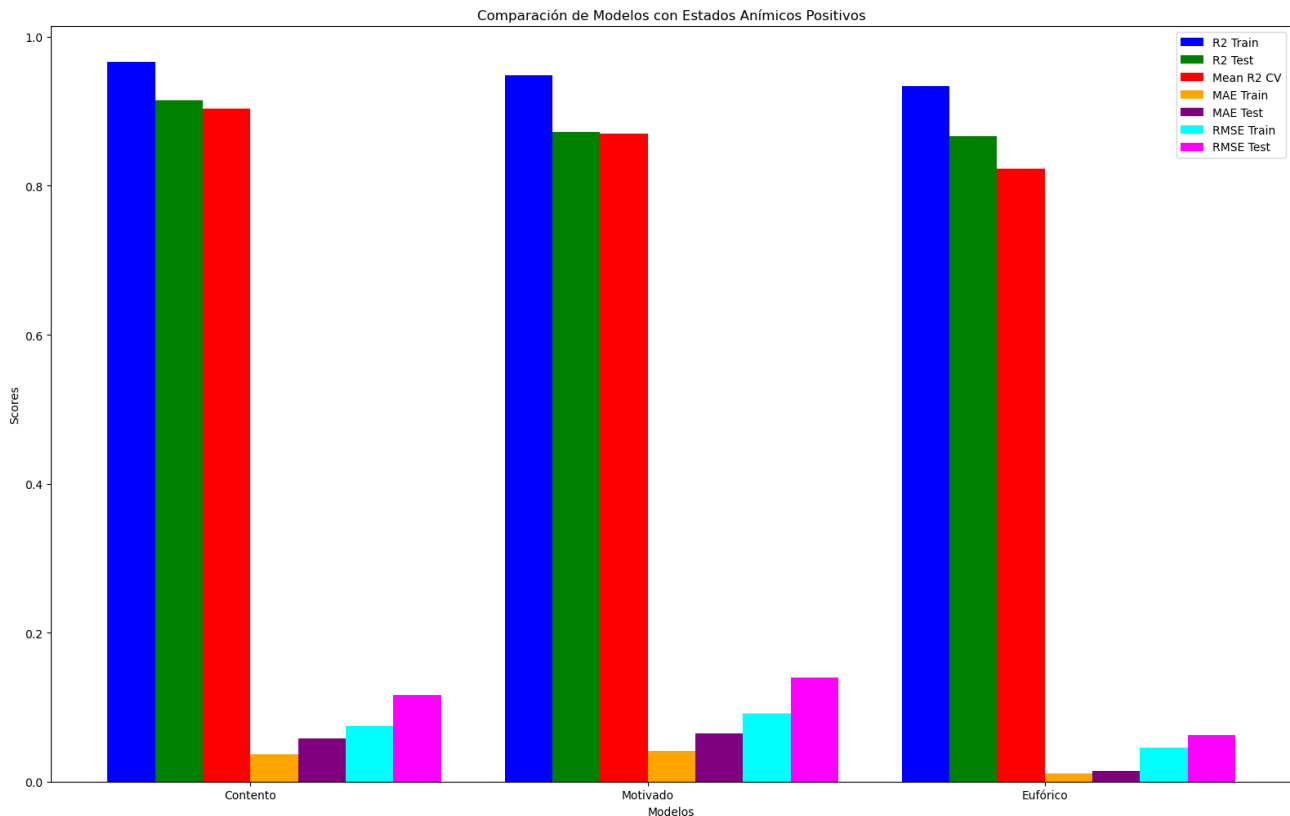


Figura 32: Gráfica Comparativa Modelo 3

### 6.4.1 Contenido

En este apartado se analizarán la importancia y el impacto de las características correspondientes al modelo que predice la variable objetivo “Contento” (Figura 33).

#### Importancia de Características SHAP

Las características más importantes del modelo son:

- Mi estado ANÍMICO durante la clase ha sido bueno: Los alumnos que se sienten bien anímicamente durante la clase están contentos.
- La docencia impartida en la clase me ha parecido buena: Una percepción positiva de la docencia se relaciona con que los alumnos estén más contentos.
- Humidity: Por primera vez en todo lo que se lleva de estudio aparecen variables recogidas por los sensores entre las cinco características más importante. En este caso niveles altos de humedad tienen un impacto positivo en la predicción de sentirse contento, mientras que, valores bajos de humedad tienen un impacto negativo.
- CO2: Al igual que la humedad, valores altos de CO2 afectan positivamente en sentirse contento.



- Temperature y Luminosity: Estas variables también influyen en la predicción. Sin embargo, a diferencia de la humedad y el CO2, valores bajos de temperatura y luminosidad se asocian con una mayor felicidad de los alumnos. Es decir, cuando la temperatura o luminosidad no son tan altas, los alumnos se sienten más contentos.
- Como dato curioso, el ruido capturado por el sensor o percibido por los alumnos no afecta prácticamente nada en la predicción.

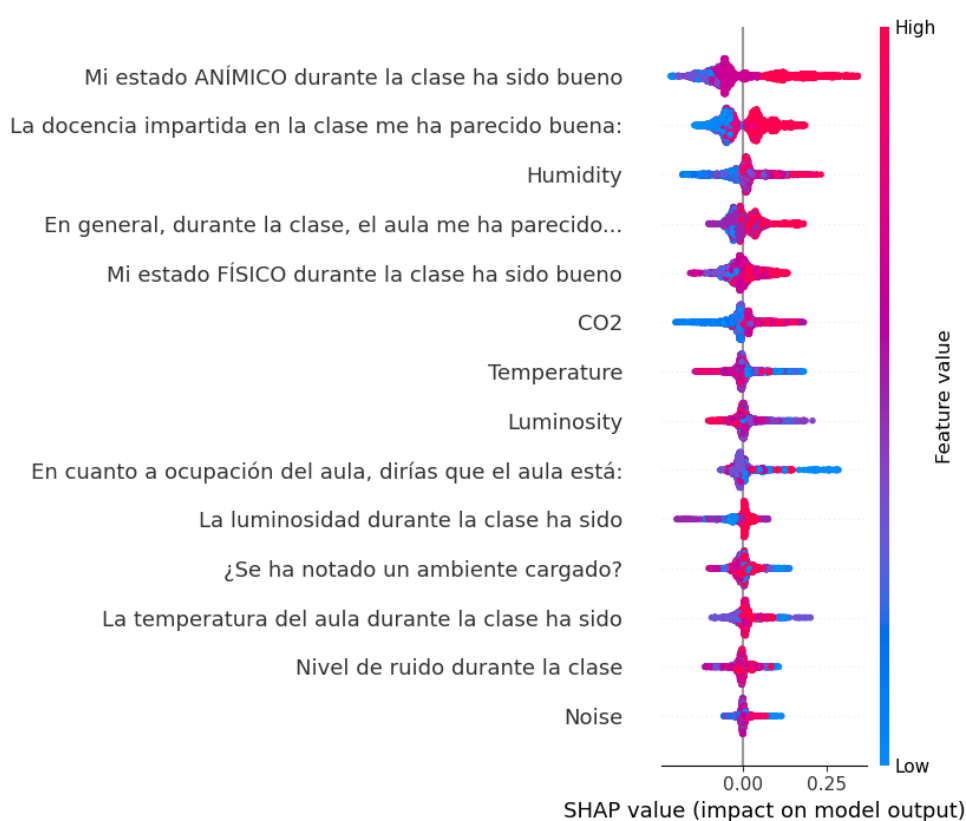


Figura 33: Gráfica Características SHAP estado Contento

#### 6.4.2 Motivado

En este apartado se analizarán la importancia y el impacto de las características correspondientes al modelo que predice la variable objetivo "Motivado" (Figura 34).

##### Importancia de Características SHAP

- Los resultados de esta predicción son similares a las del modelo anterior. Ambos modelos ("Contento" y "Motivado") comparten varias características importantes, como el estado anímico, la percepción de la docencia, y la humedad.
- Las variables Temperature y Luminosity tienen un impacto inverso que en "Contento", los alumnos que se sienten motivados prefieren una temperatura más alta y una luminosidad más baja.

- La percepción de ruido toma más importancia en esta predicción, lo que indica que los alumnos que están motivados prefieren que el ruido de la clase sea bajo. Sin embargo, el ruido percibido por el sensor es despreciable.

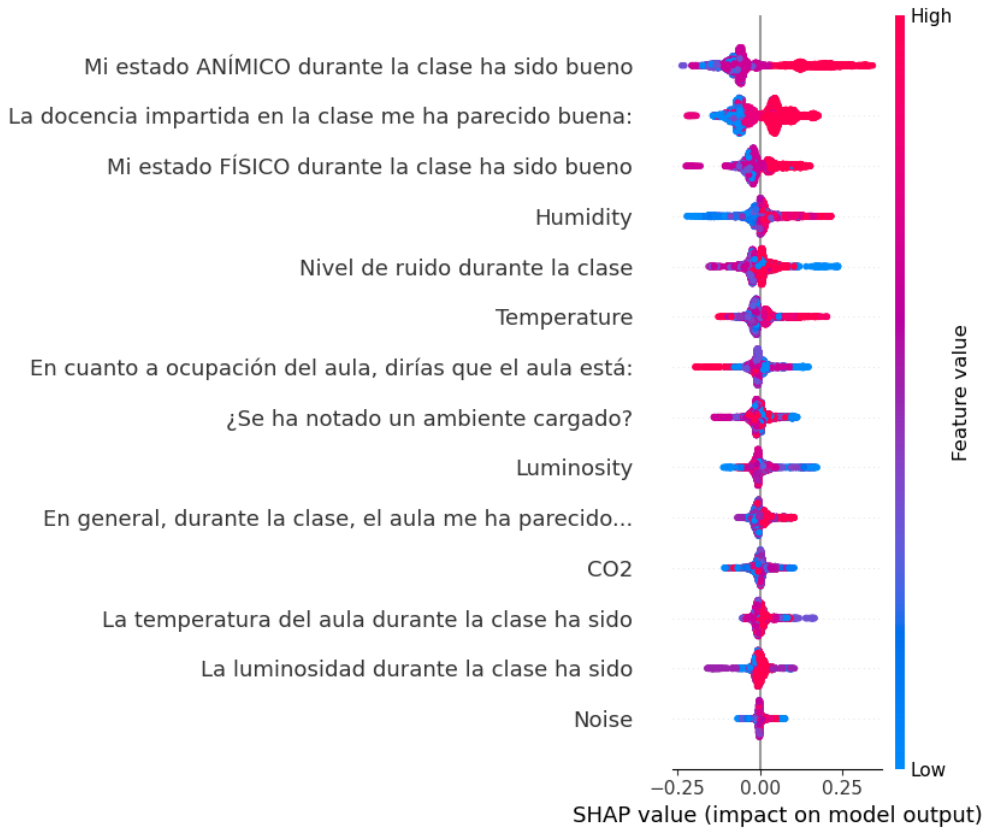


Figura 34: Gráfica Características SHAP estado Motivado

### 6.4.3 Pruebas adicionales

A los tres modelos comentados inicialmente “Contento”, “Motivado” y “eufórico” se les sometió a tres pruebas adicionales.

1. Se evaluaron los modelos con los estados físicos: Enérgico/a, Normal, Agotado/a, Enfermo/a, Con hambre o sed, Con dolor de cuello, Con dolor de cabeza, Con dolor de espalda y Con sueño.
2. Se evaluaron los modelos eliminando las características menos importantes de cada uno de ellos para ver si mejoraba su rendimiento.
3. Se ajustaron los hiperparámetros para ver si se conseguían mejores rendimientos.

En estas pruebas se obtuvieron resultados peores que los expuestos anteriormente, por lo que se descartaron.

---

## 6.5 Modelo 4: Estados Anímicos Negativos

En este apartado se presentarán los modelos correspondientes a los estados anímicos positivos: Agobiado, Estresado y Con Ansiedad. Primero se compararán los resultados de rendimiento obtenidos y finalmente se detallarán los resultados gráficos individualmente.

### Comparación resultados

- **Algoritmo empleado:** RandomForest
- **Conjunto de datos de entrenamiento y test (%):** 70%-30%
- **Hiperparámetros utilizados:** *RandomizedSearchCV()*

En este caso, se han utilizado hiperparámetros diferentes para cada estado, se mostrarán más adelante junto con los resultados.

- **Resultados obtenidos:**

En la tabla 9 se pueden ver los resultados obtenidos tras realizar las predicciones, sin embargo, para evaluar más sencillamente los modelos se ha diseñado una gráfica comparativa (Figura 35).

Los modelos parecen tener un buen rendimiento, tienen valores altos de  $R^2$  próximos a la media de validación cruzada y errores relativamente bajos, lo que indica que generalizan bien los datos y pueden seguir siendo estudiados.

En la gráfica se puede ver que el modelo “Con Ansiedad” tiene los menores valores de  $R^2$  en el conjunto de entrenamiento, de prueba y de validación cruzada, lo que indica que explica menos la variabilidad de los datos. Aunque sus errores son menores, el hecho de tener un coeficiente de determinación peor hace que sea descartado y no se analizará en detalle posteriormente. Además, se considera complicado evaluar ese estado, ya que no siempre viene relacionado con el estado de ánimo que tiene los alumnos en clase.

El modelo que mejores resultados muestra es el correspondiente al estado “Agobiado” ya que tiene valores más altos de  $R^2$ , por lo que tiene mayor capacidad para explicar la variabilidad de los datos.

Métrica	Agobiado	Estresado	Con Ansiedad
Best RandomForest parameters	{'max_depth': 20, 'max_features': 1.0, 'min_samples_leaf': 6, 'min_samples_split': 13, 'n_estimators': 191}	{'max_depth': 10, 'max_features': 1.0, 'min_samples_leaf': 5, 'min_samples_split': 10, 'n_estimators': 101}	{'max_depth': 10, 'max_features': 1.0, 'min_samples_leaf': 3, 'min_samples_split': 11, 'n_estimators': 241}
R2 score (Train)	0.9635000729849	0.9619891764654	0.9643433503865
R2 score (Test)	0.9333216765273	0.9597791244871	0.8939433458654
MAE (Train)	0.01404941977459	0.008512604543728	0.005308361457054
MAE (Test)	0.01790361774457	0.00969539718608	0.007769141099798
RMSE (Train)	0.05498489105523	0.046662295796299	0.028858524502408
RMSE (Test)	0.07057989399064	0.047226662715240	0.047459649206478
Cross-validated R2 scores	[0.88470, 0.91574, 0.89973, 0.90559, 0.90915]	[0.92159, 0.94241, 0.915046, 0.91956, 0.8783]	[0.88943, 0.90930, 0.86422, 0.88593, 0.87430]
Mean R2 score (Cross-validated)	0.90298544499376	0.91540056947549	0.8846391950515

Tabla 9: Resultados Modelo 4

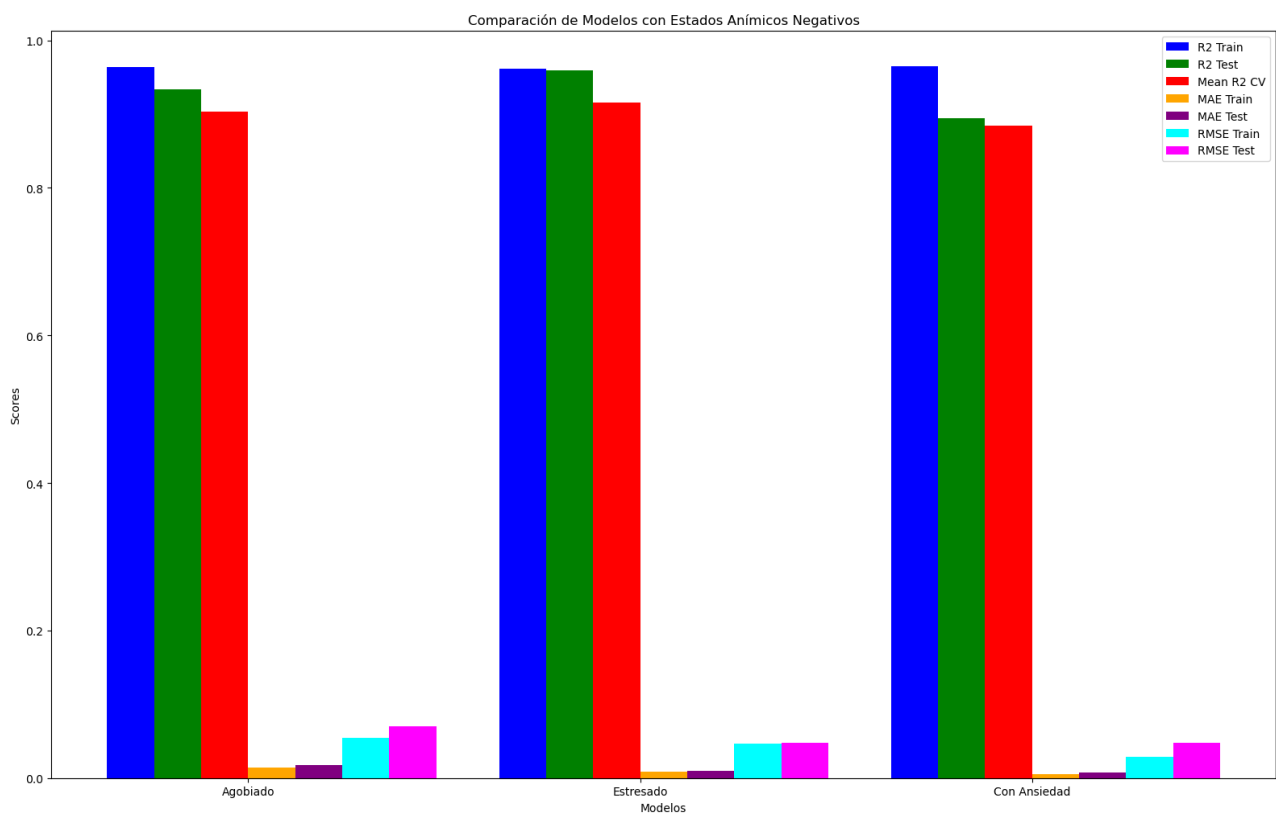


Figura 35: Gráfica Comparativa Modelo 4

### 6.5.1 Agobiado

En este apartado se analizarán la importancia y el impacto de las características correspondientes al modelo que predice la variable objetivo “Agobiado” (Figura 36).

#### Importancia de Características SHAP

Las características más importantes del modelo son:

- Mi estado ANÍMICO durante la clase ha sido bueno: Los alumnos que no tienen un buen estado anímico, están más agobiados.
- Percepción del ambiente cargado: Los alumnos se sienten más agobiados cuando notan un ambiente poco cargado.
- Humidity: En este caso niveles bajos de humedad tienen un impacto positivo en la predicción de sentirse Agobiado, por lo que a los alumnos en general les ayuda anímicamente tener niveles de humedad altos.
- Temperature: Esta variable indica que valores altos de temperatura se asocian con la sensación de sentirse agobiado.
- Una mayor ocupación del aula tiene un impacto positivo en el sentimiento de agobio.
- El ruido capturado por el sensor o percibido por los alumnos no afecta prácticamente nada en la predicción.

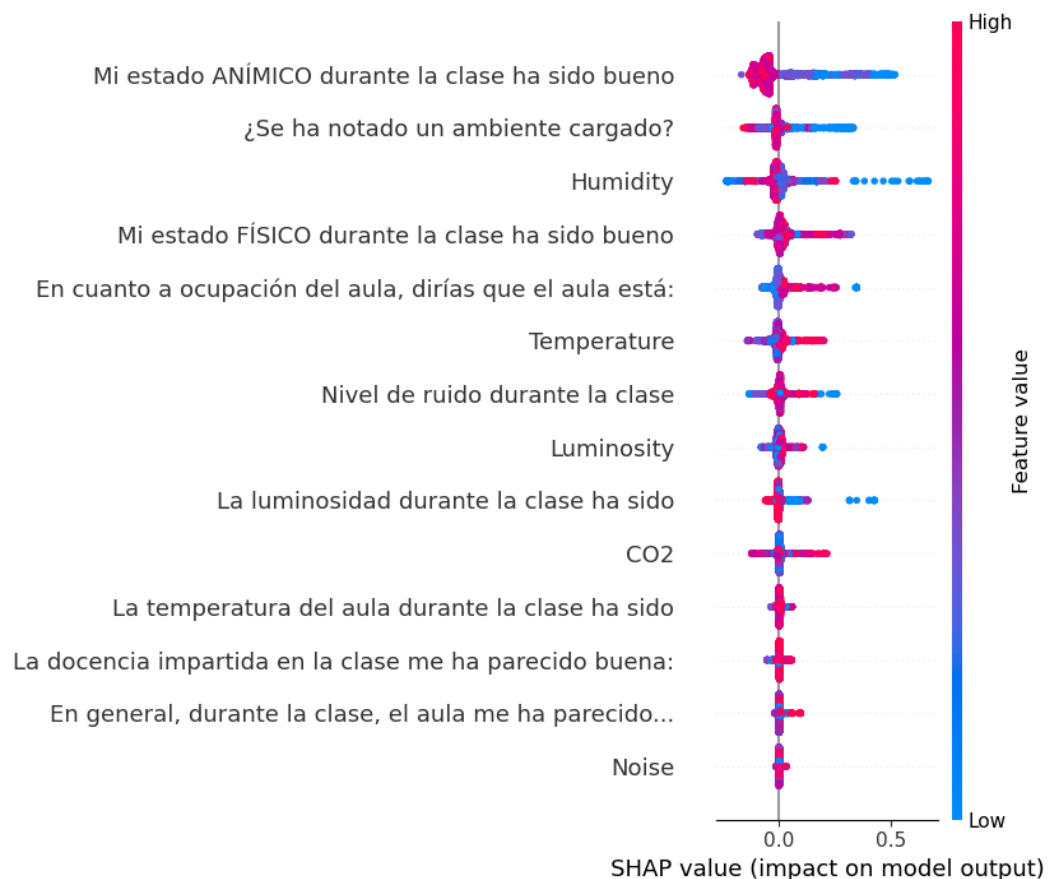


Figura 36: Gráfica Características SHAP estado Agobiado

### 6.5.2 Estresado

En este apartado se analizarán la importancia y el impacto de las características correspondientes al modelo que predice la variable objetivo “Estresado” (Figura 37).

#### Importancia de Características SHAP

- Los resultados de esta predicción son similares a las del modelo anterior. Ambos modelos (“Agobiado” y “Estresado”) comparten varias características importantes, como el estado anímico, físico, la percepción del ambiente cargado y la humedad.
- El CO2 entra entre las más influyentes. Los alumnos se sienten más estresados con valores altos de CO2, sin embargo, esta variable afecta poco cuando se sienten agobiados.
- En esta predicción la percepción de ruido toma más importancia, lo que indica que los alumnos que están motivados prefieren que el ruido de la clase sea bajo. Sin embargo, el ruido percibido por el sensor es despreciable.
- En estos dos modelos el estado físico bueno afecta a sentirse más agobiado y estresado, esto puede deberse a que los alumnos no entienden de la misma forma el concepto de “estado físico”.

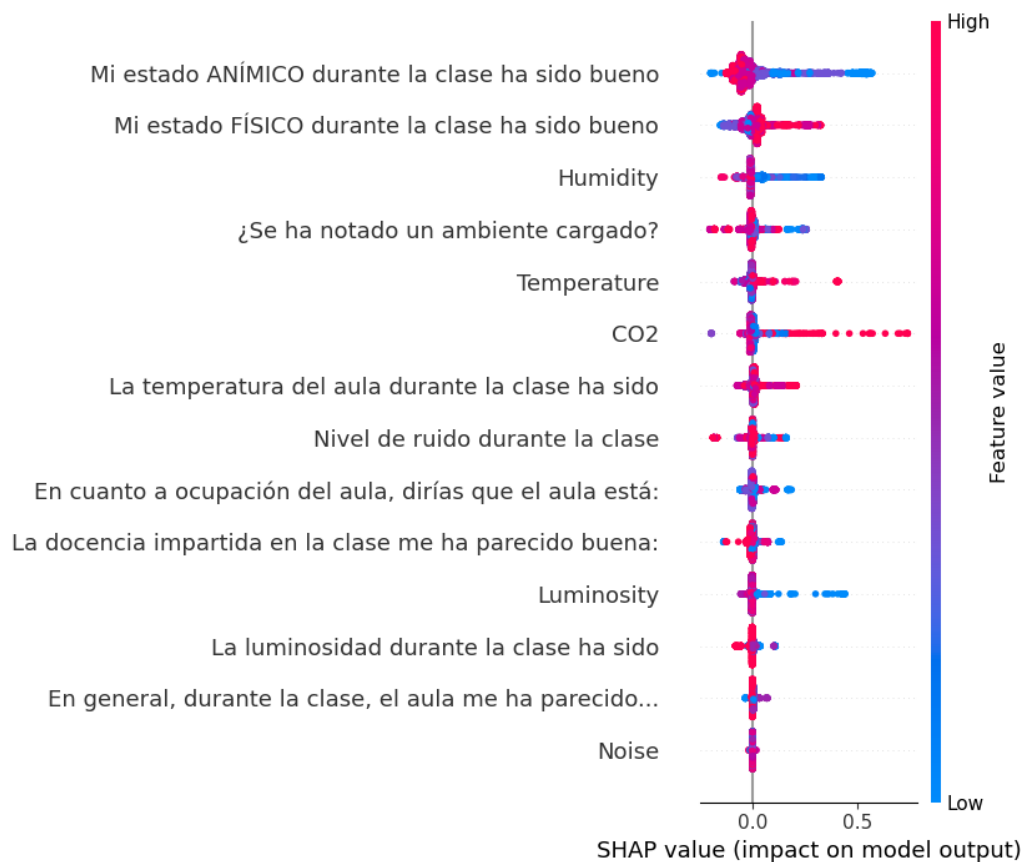


Figura 37: Gráfica Características SHAP estado Estresado

### 6.5.2 Pruebas Adicionales

A los tres modelos comentados inicialmente “Agobiado”, “Estresado” y “Con Ansiedad” se les sometió a tres pruebas adicionales.

1. Se evaluaron los modelos con los estados físicos: Enérgico/a, Normal, Agotado/a, Enfermo/a, Con hambre o sed, Con dolor de cuello, Con dolor de cabeza, Con dolor de espalda y Con sueño.
2. Se evaluaron los modelos eliminando las características menos importantes de cada uno de ellos para ver si mejoraba su rendimiento.
3. Se ajustaron los hiperparámetros para ver si se conseguían mejores rendimientos.

En estas pruebas se obtuvieron resultados peores que los expuestos anteriormente, por lo que se descartaron.

## 6.6 Implementación Cuadro de Mando en Power BI

Por último, se ha creado un cuadro de mando en Power BI, en el que el usuario puede ver los resultados obtenidos en los modelos mencionados anteriormente, exceptuando el modelo “Eufórico” y “Con Ansiedad”, ya que estos fueron descartados.

El cuadro de mando cuenta con 6 páginas diferentes que se describirán a continuación:

### 6.6.1 Panel de Control

La primera página del cuadro de mando es un panel de control (Figura 38) mediante el cual el usuario puede navegar por los diferentes modelos pulsando uno de los cinco botones implementados:



Figura 38: Panel de Control Power BI

---

### 6.6.2 Sensores

En esta página se pueden ver los datos recolectados por los sensores del aula, el promedio de cada variable medida durante los dos meses del estudio y si está fuera del rango establecido. De este modo, se puede ver fácilmente si las condiciones del aula cumplen los niveles ambientales necesarios para un aula universitaria. El usuario puede acercar el ratón a los valores de los sensores y podrá ver los criterios ambientales óptimos para el aula.

La información sobre las medidas ambientales óptimas se ha recogido de un informe del ministerio de trabajo y economía social que evalúa la Calidad del Ambiente Interior en el trabajo [51]. Se compara la calidad ambiental en un aula con la del trabajo, debido a que ambas son medidas en espacios cerrados donde los usuarios necesitan tener concentración y un buen estado anímico para llevar a cabo sus tareas.

Variable	Valor óptimo
Ruido	< 87 dB
Iluminación	750-1000 Lux
CO2	< 500 ppm
Humedad	30-70%
Temperatura	17-27°C

Tabla 10: Criterios Ambientales

Se ha considerado la iluminación en un espacio cerrado con exigencia visual alta debido a que el aula que está siendo estudiada es un laboratorio con un ordenador en cada puesto. Por ello se considera necesario tener buena iluminación para minimizar el esfuerzo visual y evitar el deslumbramiento en las pantallas.

A continuación, se muestra la implementación en Power BI (figura 39):

Se puede observar que tanto la temperatura como el CO2 y la luminosidad sobrepasan los límites recomendados.

- Cuando la concentración de CO2 supera los 1000 ppm puede causar sensación de incomodidad, e indica que es necesario ventilar la sala. Este valor es preocupante cuando supera los 3000 ppm, sin embargo, el valor obtenido en el aula estudiada (1400 ppm), no indica ningún peligro para la salud.
- La temperatura supera el valor recomendado (27 °C). En las predicciones de “Estresado” y “Agobiado” esta variable ambiental percibida por el sensor en bastante influyente, lo cual indica que valores altos de temperatura, hacen que los alumnos no estén bien anímicamente. Por lo que se puede proponer tener más control de la calefacción en invierno para mejorar el confort de los estudiantes. Sin embargo, una temperatura tan alta en el aula estudiada es lógica debido a que se trata de un aula de



laboratorio en la que hay muchos ordenadores funcionando al mismo tiempo y generando más calor. Hay que tener especial cuidado en las aulas de laboratorio y hacer un mejor uso de la calefacción.

- La temperatura recomendada es de 500 lux en espacios cerrados con exigencia visual alta. En el caso del aula estudiada tiene un promedio de 725.02 lux, lo que indica que sobrepasa bastante el límite recomendado. Sin embargo, esta variable no parece afectar a las percepciones ni a los estados anímicos de los estudiantes.

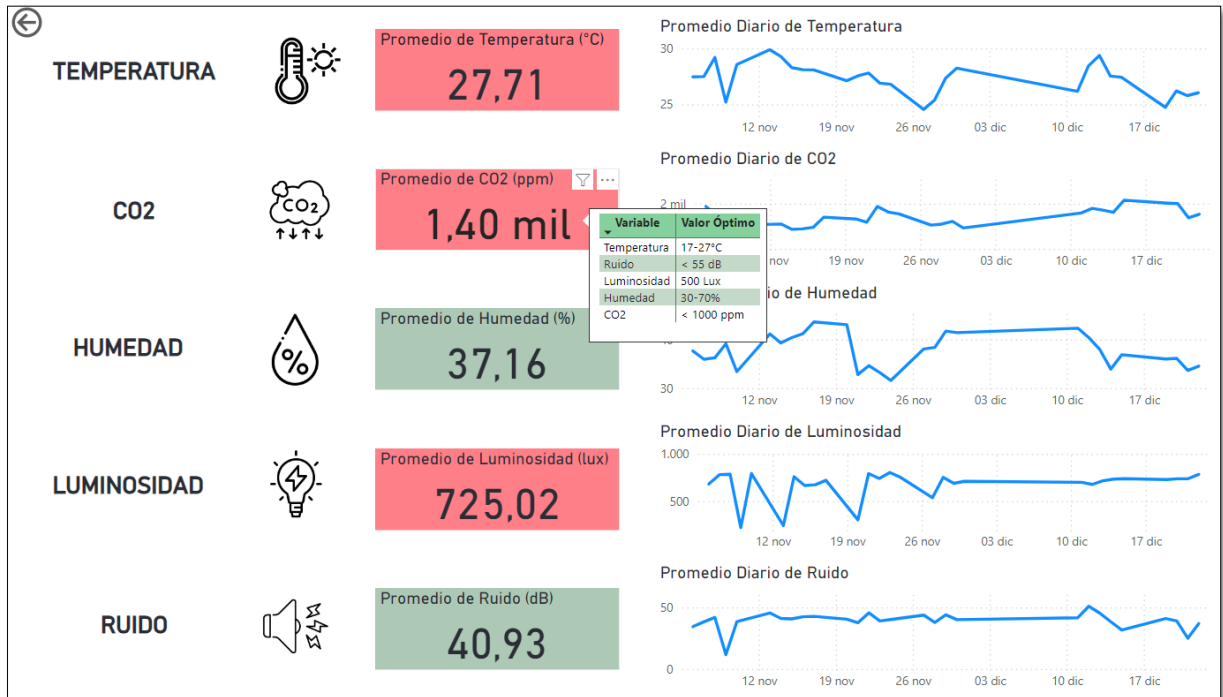


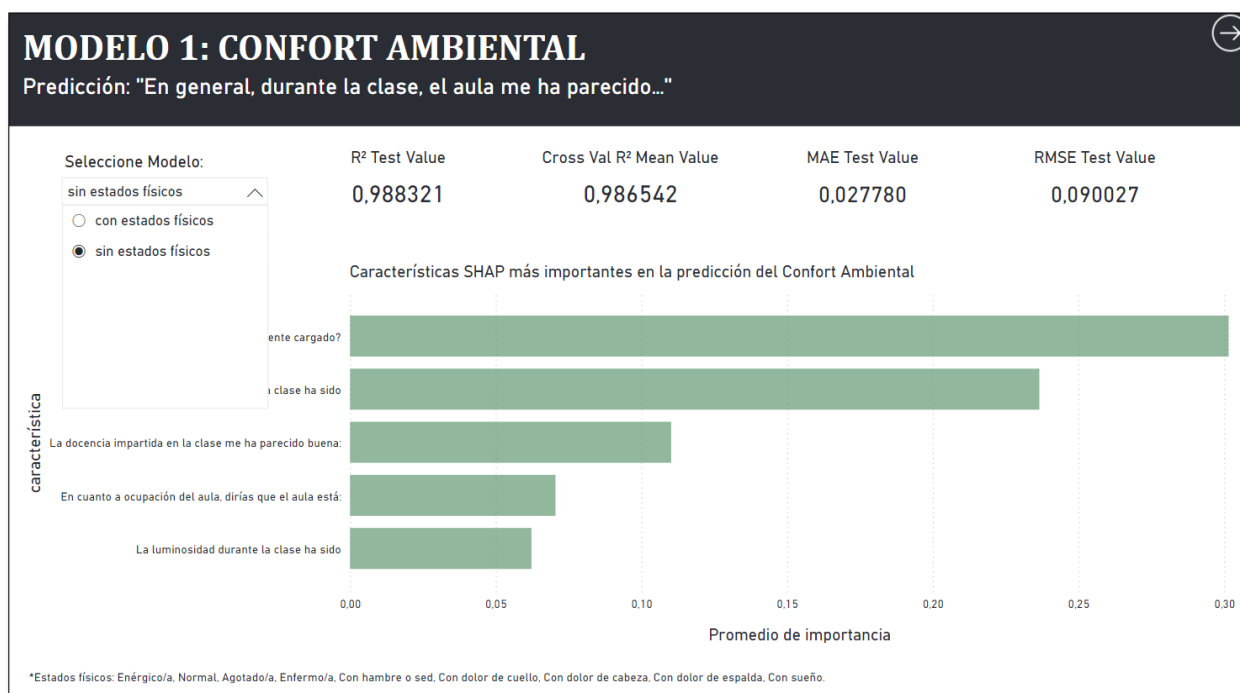
Figura 39: Resumen Sensores

### 6.6.3 Resultados de los modelos

En esta página (figura 40) se pueden ver los resultados explicados anteriormente del Modelo 1 y sus dos variantes.

La grafica de barras muestra la media del valor absoluto de las cinco características SHAP más importantes de cada modelo, de modo que facilita la visualización del impacto de cada característica en el mismo. También se pueden ver los resultados obtenidos en cada uno de ellos.

Se ha implementado un botón para pasar de página y ver el resto de los modelos. Los demás modelos se representan igual que el mostrado a continuación, cambiado las variantes de cada uno de ellos.



**Figura 40: Visualización Confort Ambiental**

#### 6.6.4 Pruebas Adicionales

A parte de los diseños enseñados para cada modelo, se intentó cargar a Power BI el modelo uno de ellos, en concreto el “modelo 2: Percepción de la Docencia”. El objetivo era permitir que el usuario pudiera rellenar la encuesta desde el propio cuadro de mando y obtener la predicción de la docencia de la clase impartida en una escala de Likert (1-5). De este modo, se podría evaluar si a un alumno le ha gustado la clase impartida sin sesgar su criterio al indicarle la evaluación.

No se pudo llevar a cabo ya que para cargar un modelo a Power Bi se necesitaba tener una licencia de Azure. Se intentó realizar el código desde el objeto visual de Python de la aplicación, pero el resultado no fue el esperado.

---

## 7. Presupuesto

En este apartado se detallará el presupuesto necesario para la elaboración de este proyecto.

Para calcular las horas empleadas, se han tenido en cuenta las diferentes tareas en las que se divide el proyecto, y las horas necesarias para elaborar cada una de ellas:

Tareas realizadas	Tiempo necesario
Investigación previa	75 horas
Preprocesamiento y limpieza de los datos	25 horas
Diseño de los modelos de aprendizaje automático	120 horas
Análisis de los resultados	30 horas
Diseño cuadro de mando	25 horas
Redacción de la memoria	75 horas
<b>TOTAL</b>	<b>350 horas</b>

Tabla 11: Evaluación Tiempo Utilizado

También se debe calcular el sueldo correspondiente a un ingeniero junior, siendo aproximadamente de 14€/hora. Por último, se deben evaluar:

- **Costes Directos:**
  - Mano de obra
  - Recursos materiales
  - Material Fungible
- **Costes Indirectos:** Se calcula sobre el 15% del Coste directo, corresponde al sueldo del jefe de proyecto, suministros generales etc.
- **Beneficio Industrial:** Se calcula como un 6% sobre la suma del coste directo y el indirecto, corresponde al margen de ganancia.

A continuación, se puede ver el presupuesto total desglosado:

COSTE DE MANO DE OBRA	Horas	Precio/hora	total
	350	14 €	4.900 €

COSTE DE RECURSOS MATERIALES	Precio de Venta	Tiempo de Uso	Amortización	total
Ordenador	1.200,00 €	10 meses	5 años	200,00 €

MATERIAL FUNGIBLE	Coste
Sensor SCD30	59,25 €
Sensor VEML7700	21,62 €
Sensor KECG2240PBJ	14,56 €
licencia Power BI Desktop	0 €
Jupyter Python	0 €

<b>TOTAL, COSTE DIRECTO</b>	<b>5.195 €</b>
-----------------------------	----------------

COSTE INDIRECTO	Base	total
15%	sobre CD	779,31 €

BENEFICIO INDUSTRIAL	Base	total
6%	sobre CD+CI	358,48 €

SUBTOTAL PRESUPUESTO		6.333,23 €
IVA APLICABLE	21%	1.329,98 €

<b>TOTAL, PRESUPUESTO</b>	<b>7.663,21 €</b>
---------------------------	-------------------

Tabla 12: Tabla de Presupuestos

---

## 8. Impacto del proyecto

Para poder desarrollar este capítulo, se va a dividir el impacto del proyecto en diferentes ámbitos:

### **Impacto de Salud y Seguridad**

Las mejoras de las condiciones ambientales de las aulas universitarias tienen un impacto directo en la salud y seguridad de los estudiantes y profesores. Regular adecuadamente la temperatura, la humedad y la ventilación de espacios cerrados ayuda a mejorar el entorno de aprendizaje y a prevenir problemas de salud como alergias, problemas respiratorios o estrés. Tomar las medidas indicadas ayuda a proteger el bienestar físico y mental de los usuarios de la universidad.

### **Impacto económico y energético:**

Los resultados observados en este estudio demuestran que es necesario bajar la temperatura de la calefacción en invierno, sobre todo en las aulas de laboratorio que cuentan con varios puestos con ordenadores, y ventilar más las aulas universitarias. A lo largo del tiempo, tomar estas medidas pueden generar un gran impacto económico y energético, disminuyendo la huella de carbono de la universidad y reduciendo los costos de la energía.

### **Impacto Tecnológico:**

En este proyecto se han utilizado sensores para capturar las diferentes variables ambientales dentro de un aula universitaria y algoritmos de aprendizaje automático para analizar los datos, lo cual supone un gran avance en el área tecnológica de investigación en entornos educativos. Gracias a este proyecto, la universidad podrá mejorar sus espacios para los alumnos y el profesorado, lo que puede inspirar a otras instituciones a utilizar tecnologías similares para mejorar el ambiente y la calidad de la universidad.

### **Aportación a los Objetivos de Desarrollo Sostenible (ODS)**

Este proyecto contribuye a varios Objetivos de Desarrollo Sostenible de las Naciones Unidas [52], fortaleciendo el compromiso de la Universidad Politécnica de Madrid con la sostenibilidad y el bienestar global.

A continuación, se detallan los ODS alineados con el estudio:

- **ODS 3 Salud y Bienestar:** Este ODS busca garantizar una vida sana y promover el bienestar para todos en todas las edades. En este caso, el proyecto pretende mejorar el bienestar de los estudiantes universitarios y las condiciones ambientales del aula para prevenir posibles enfermedades provocadas por pasar mucho tiempo en espacios cerrados mal acondicionados.

- 
- **ODS 4 Educación de Calidad:** Busca garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos. Se alinea con el proyecto, ya que proporcionar un ambiente de aprendizaje óptimo, mejora la experiencia educativa y el rendimiento académico de los estudiantes. Esto fomenta una educación más inclusiva y con mejor calidad.
  - **ODS 11 Ciudades y Comunidades Sostenibles:** Se busca lograr que las ciudades y los asentamientos urbanos sean inclusivos, seguros, resilientes y sostenibles. Entender y mejorar las condiciones ambientales en entornos educativos contribuye a la creación de comunidades más sostenibles.



Figura 41: Objetivos de desarrollo sostenible

---





---

## 9. Conclusiones

### 9.1 Conclusiones

En este Trabajo de Fin de Grado se ha llevado a cabo un estudio cuyo objetivo principal era el diseño e implementación de un cuadro de mando basado en inteligencia colectiva para potenciar la sostenibilidad y el bienestar en entornos educativos.

El entorno universitario es un espacio dinámico donde la calidad del ambiente desempeña un papel importante en el éxito de los estudiantes y del profesorado. Un ambiente de aprendizaje óptimo está influenciado en gran medida por una buena iluminación, una temperatura agradable, una ventilación adecuada, entre otros factores. Estos factores ambientales permiten mejorar el rendimiento académico y la concentración de los alumnos, al mismo tiempo que promueven la participación, el trabajo en equipo y reducen la abstinencia de los estudiantes en las clases. En este estudio se ha explorado cómo la interacción social, complementada con el uso estratégico de la tecnología, puede realzar la efectividad académica y optimizar la toma de decisiones en ambientes educativos específicos, siendo en nuestro caso la universidad.

En concreto, se ha estudiado como las condiciones ambientales y las percepciones de los estudiantes pueden afectar su experiencia en el aula. Para ello, se recolectaron datos de humedad, temperatura, luminosidad, ruido y CO<sub>2</sub> mediante tres sensores ubicados en un aula de laboratorio de la Universidad Politécnica de Madrid y se realizaron encuestas a los estudiantes sobre su percepción del aula, de su estado anímico y físico durante las clases.

En primer lugar, se limpiaron los datos de los sensores y las encuestas, para después diseñar ocho modelos de aprendizaje automático usando el algoritmo RandomForest para predecir diversas percepciones y estados de los estudiantes. Por último, se diseñó un cuadro de mando en Power BI en el cual se pueden visualizar los resultados más importantes de cada modelo y la evolución y el promedio de las condiciones ambientales del aula estudiada durante los meses de invierno.

Como resultados, se observó que la percepción de un ambiente cargado y una temperatura alta tiene un impacto positivo en la percepción general del aula. También se descubrió que niveles altos de CO<sub>2</sub> puede aumentar el nivel de estrés entre los estudiantes, por lo que es recomendable ventilar las aulas también durante el invierno.

El resultado más relevante de esta investigación es que niveles altos de humedad tienen un impacto positivo en la predicción de sentirse contento, lo que indica que durante el invierno una humedad alta puede contribuir a la felicidad de los alumnos. Por lo contrario, niveles bajos de humedad se asociaron con la sensación de agobio. Esto se resume en que la humedad influye directamente en el bienestar emocional de los alumnos en el aula.

Con todo esto, para mejorar el entorno educativo y la calidad ambiental en las aulas, se propone regular la calefacción, ventilar las clases y poner humidificadores. Con estas medidas

---

se espera crear un ambiente más confortable para el aprendizaje, con el fin de mejorar la experiencia universitaria a los estudiantes y su rendimiento académico. Además, estas acciones pueden tener un impacto económico y energético positivo a largo plazo, reduciendo la huella de carbono de la universidad y disminuyendo los costos asociados al consumo energético.

Desde una perspectiva tecnológica, el uso de sensores para monitorear variables ambientales y la implementación de algoritmos de aprendizaje automático para analizar estos datos representan grandes avances en la investigación de entornos educativos. Esta tecnología permite una gestión más eficiente y personalizada del ambiente en las aulas y sienta las bases para futuras innovaciones en la mejora continua de la calidad educativa y ambiental.

## **9.2 Trabajos futuros**

Tras la finalización de este estudio, se presentan varias oportunidades para futuras investigaciones y mejoras en este campo. A continuación, se detallan algunas sugerencias para trabajos futuros:

- Probar los modelos de predicción diseñados con nuevos datos para comprobar su correcto funcionamiento.
- Realizar estudios para evaluar cómo las mejoras implementadas en las condiciones ambientales de las aulas afectan el bienestar y el rendimiento académico de los estudiantes a lo largo de varios años.
- Realizar comparaciones entre diferentes instituciones para identificar las mejores prácticas en términos de gestión ambiental y su impacto en la experiencia educativa y el desempeño académico.
- Realizar un estudio similar en un aula de la universidad que no sea un laboratorio para comprobar las diferencias de las variables ambientales y percepciones.
- Crear modelos específicos para cada mes del año, para comparar los resultados en cada estación.

---



---

## 10. Referencias

- [1] «Inteligencia colectiva: Qué es, características y beneficios». Accedido: 07 de enero de 2024. [En línea]. Disponible en: <https://ceupe.com/blog/inteligencia-colectiva.html>
- [2] A. W. Woolley, I. Aggarwal, y T. W. Malone, «Collective Intelligence and Group Performance», *Curr Dir Psychol Sci*, vol. 24, n.º 6, pp. 420-424, dic. 2015, doi: 10.1177/0963721415599543.
- [3] A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, y T. W. Malone, «Evidence for a collective intelligence factor in the performance of human groups», *Science (1979)*, vol. 330, n.º 6004, pp. 686-688, oct. 2010, doi: 10.1126/SCIENCE.1193147.
- [4] D. Marzo de, «INTELIGENCIA COLECTIVA POR PIERRE LÉVY», 2004, Accedido: 07 de enero de 2024. [En línea]. Disponible en: <http://inteligenciacolectiva.bvsalud.org>
- [5] D. W. Johnson, R. T. Johnson, y K. A. Smith, «Co-operative learning: Improving university instruction by basing practice on validated theory», *J Excell Coll Teach*, vol. 25, n.º 4, pp. 85-118, 2014.
- [6] R. E. Slavin, «Aprendizaje cooperativo y rendimiento académico: ¿por qué funciona el trabajo grupal?», *Anales de Psicología*, vol. 30, n.º 3, pp. 785-791, 2014, doi: 10.6018/ANALES.30.3.201201.
- [7] D. W. Johnson y R. T. Johnson, «An educational psychology success story: Social interdependence theory and cooperative learning», *Educational Researcher*, vol. 38, n.º 5, pp. 365-379, jun. 2009, doi: 10.3102/0013189X09339057.
- [8] E. Wenger, «Community of Practice: a Brief Introduction», *Learning in doing*, vol. 15, n.º 4, pp. 1-7, 1998, Accedido: 07 de enero de 2024. [En línea]. Disponible en: [WORD: http://www.ewenger.com/theory/](http://www.ewenger.com/theory/)
- [9] «La importancia de la inteligencia colectiva en la empresa». Accedido: 10 de enero de 2024. [En línea]. Disponible en: <https://www.coachhub.com/es/blog/entender-la-importancia-de-la-inteligencia-colectiva-en-la-empresa/>
- [10] «Calidad ambiental: qué es y que factores influyen | Blog EnergyGO». Accedido: 10 de enero de 2024. [En línea]. Disponible en: <https://blog.energygo.es/glosario/definicion-calidad-ambiental/>
- [11] U. Haverinen-Shaughnessy, R. J. Shaughnessy, E. C. Cole, O. Toyinbo, y D. J. Moschandreas, «An assessment of indoor environmental quality in schools and its association with health and performance», *Build Environ*, vol. 93, n.º P1, pp. 35-40, nov. 2015, doi: 10.1016/J.BUILDENV.2015.03.006.

- 
- [12] D. Yang y C. M. Mak, «Relationships between indoor environmental quality and environmental factors in university classrooms», *Build Environ*, vol. 186, p. 107331, dic. 2020, doi: 10.1016/J.BUILDENV.2020.107331.
- [13] «What is a sensor? | Fierce Electronics». Accedido: 10 de enero de 2024. [En línea]. Disponible en: <https://www.fierceelectronics.com/sensors/what-a-sensor>
- [14] «Sensores: Tipos Y Aplicaciones En La Automatización Y El Control. - Electropreguntas». Accedido: 10 de enero de 2024. [En línea]. Disponible en: <https://electropreguntas.com/sensores-tipos-y-aplicaciones-en-la-automatizacion-y-el-control/>
- [15] N. Gao, M. Marschall, J. Burry, S. Watkins, y F. D. Salim, «Understanding occupants' behaviour, engagement, emotion, and comfort indoors with heterogeneous sensors and wearables», *Scientific Data* 2022 9:1, vol. 9, n.º 1, pp. 1-16, jun. 2022, doi: 10.1038/s41597-022-01347-w.
- [16] «El 33% del consumo energético mundial tiene lugar en los edificios - Infoconstrucción». Accedido: 13 de enero de 2024. [En línea]. Disponible en: <https://www.infoconstruccion.es/noticias/20190306/edificios-consumo-energetico>
- [17] «Visualización de datos: cómo presentar la información para facilitar la toma de decisiones | Liderazgo & Gestión TI | CIO». Accedido: 13 de enero de 2024. [En línea]. Disponible en: <https://www.ciospain.es/liderazgo--gestion-ti/visualizacion-de-datos-como-presentar-la-informacion-para-facilitar-la-toma-de-decisiones>
- [18] «Visualización de datos. Parte (I), historia y contexto - esPublico Gestiona». Accedido: 13 de enero de 2024. [En línea]. Disponible en: <https://espublicogestiona.com/visualizacion-de-datos-parte-i-historia-y-contexto/>
- [19] «Visualización de datos: beneficios, historia, tipos y software». Accedido: 13 de enero de 2024. [En línea]. Disponible en: <https://www.itmastersmag.com/noticias-analisis/la-importancia-de-aplicar-una-visualizacion-de-datos/>
- [20] «Tipos de análisis de datos - ATLAS.ti». Accedido: 13 de enero de 2024. [En línea]. Disponible en: <https://atlasti.com/es/research-hub/tipos-de-analisis-de-datos#herramientas-de-analisis-de-datos>
- [21] «Sci-Hub | Evaluation and Analysis of Business Intelligence Data Visualization Tools. 2019 14th Iberian Conference on Information Systems and Technologies (CISTI) | 10.23919/cisti.2019.8760677». Accedido: 20 de enero de 2024. [En línea]. Disponible en: <https://sci-hub.se/10.23919/cisti.2019.8760677>
- [22] «Tableau: ¿Qué es y cómo trabajar con él?» Accedido: 20 de enero de 2024. [En línea]. Disponible en: <https://immune.institute/blog/tableau-que-es/>

- 
- [23] «Google Data Studio: ¿Qué es y cómo utilizarlo?» Accedido: 20 de enero de 2024. [En línea]. Disponible en: <https://bloo.media/blog/tutorial-google-data-studio/>
- [24] «¿Qué es Power BI? - Power BI | Microsoft Learn». Accedido: 20 de enero de 2024. [En línea]. Disponible en: <https://learn.microsoft.com/es-es/power-bi/fundamentals/power-bi-overview>
- [25] «Extracción, transformación y carga de datos (ETL) - Azure Architecture Center | Microsoft Learn». Accedido: 20 de enero de 2024. [En línea]. Disponible en: <https://learn.microsoft.com/es-es/azure/architecture/data-guide/relational-data/etl>
- [26] «Power BI, la herramienta de moda en gestión de datos». Accedido: 20 de enero de 2024. [En línea]. Disponible en: [https://www.enaes.es/blog/power-bi-la-herramienta-de-moda-en-gestion-de-datos?\\_adin=02021864894#gref](https://www.enaes.es/blog/power-bi-la-herramienta-de-moda-en-gestion-de-datos?_adin=02021864894#gref)
- [27] «Intro to dashboards for Power BI designers - Power BI | Microsoft Learn». Accedido: 20 de enero de 2024 [En línea]. Disponible en: <https://learn.microsoft.com/en-us/power-bi/create-reports/service-dashboards>
- [28] «Handling and Assembly Guide for SCD30», 2018, Accedido: 15 de febrero de 2024. [En línea]. Disponible en: [www.sensirion.com](http://www.sensirion.com)
- [29] «High Accuracy Ambient Light Sensor With I<sup>2</sup>C Interface Pinning», Accedido: 15 de febrero de 2024. [En línea]. Disponible en: [www.vishay.com/doc?91000](http://www.vishay.com/doc?91000)
- [30] «KECG2740PBJ Kingstate, Microphone, Electret Condenser, Omnidirectional | Farnell Portugal». Accedido: 15 de febrero de 2024. [En línea]. Disponible en: <https://pt.farnell.com/en-PT/kingstate/kecg2740pbj/microphone-miniature-electret/dp/1299885>
- [31] «Ventilación y CO<sub>2</sub>: cómo influye en el rendimiento escolar». Accedido: 15 de febrero de 2024. [En línea]. Disponible en: <https://www.educaciontrespuntocero.com/opinion/ventilacion-y-co2/>
- [32] «Qué es la humedad relativa y cuál es el porcentaje saludable | Hydrotec». Accedido: 15 de febrero de 2024. [En línea]. Disponible en: <https://hydrotec.es/humedades/que-es-la-humedad-relativa/>
- [33] «Importancia De Iluminación, Ruido Y Temperatura En El Trabajo: Consejos Y Prevención | La Pantalla». Accedido: 15 de febrero de 2024. [En línea]. Disponible en: [https://www.la-pantalla.com.ar/iluminacion-ruido-y-temperatura-en-el-trabajo/?shared=false#google\\_vignette&shared=true](https://www.la-pantalla.com.ar/iluminacion-ruido-y-temperatura-en-el-trabajo/?shared=false#google_vignette&shared=true)
- [34] «¿Qué es un árbol de decisión? | IBM». Accedido: 05 de marzo de 2024. [En línea]. Disponible en: <https://www.ibm.com/es-es/topics/decision-trees>

- 
- [35] «¿Qué es un bosque aleatorio? | IBM». Accedido: 05 de marzo de 2024. [En línea]. Disponible en: <https://www.ibm.com/es-es/topics/random-forest>
- [36] «Algoritmo random forest: qué es y cómo funciona». Accedido: 05 de marzo de 2024. [En línea]. Disponible en: <https://www.educaopen.com/digital-lab/blog/inteligencia-artificial/algoritmo-random-forest>
- [37] «File:Random forest explain.png - Wikimedia Commons». Accedido: 05 de marzo de 2024. [En línea]. Disponible en: [https://commons.wikimedia.org/wiki/File:Random\\_forest\\_explain.png](https://commons.wikimedia.org/wiki/File:Random_forest_explain.png)
- [38] «Optimización de hiperparámetros de un clasificador de Machine Learning en Python - ». Accedido: 19 de marzo de 2024. [En línea]. Disponible en: [https://panamahitek.com/optimizacion-de-hiperparametros-machine-learning/?utm\\_content=cmp-true](https://panamahitek.com/optimizacion-de-hiperparametros-machine-learning/?utm_content=cmp-true)
- [39] «Difference between GridSearchCV and RandomizedSearchCV - 360DigiTMG». Accedido: 19 de marzo de 2024. [En línea]. Disponible en: <https://360digitmg.com/blog/comparison-gridsearchcv-and-randomsearchcv>
- [40] «Qué es overfitting y underfitting y cómo solucionarlo | Aprende Machine Learning». Accedido: 19 de marzo de 2024. [En línea]. Disponible en: <https://www.aprendemachinelearning.com/que-es-overfitting-y-underfitting-y-como-solucionarlo/>
- [41] «Estadísticas de R cuadrado - Documentación de IBM». Accedido: 05 de abril de 2024. [En línea]. Disponible en: <https://www.ibm.com/docs/es/spss-statistics/saas?topic=model-r-squared-statistics>
- [42] «Conoce las métricas de precisión más comunes para Modelos de Regresión». Accedido: 05 de abril de 2024. [En línea]. Disponible en: <https://www.growupcr.com/post/metricas-precision>
- [43] «3.1. Cross-validation: evaluating estimator performance — scikit-learn 1.5.0 documentation». Accedido: 22 de abril de 2024. [En línea]. Disponible en: [https://scikit-learn.org/stable/modules/cross\\_validation.html](https://scikit-learn.org/stable/modules/cross_validation.html)
- [44] «Cross-Validation : definición e importancia en Machine Learning». Accedido: 22 de abril de 2024. [En línea]. Disponible en: <https://datascientest.com/es/cross-validation-definicion-e-importancia>
- [45] «Gráficas de residuos para Ajustar modelo de regresión - Minitab». Accedido: 03 de mayo de 2024. [En línea]. Disponible en: <https://support.minitab.com/es-mx/minitab/help-and-how-to/statistical-modeling/regression/how-to/fit-regression-model/interpret-the-results/all-statistics-and-graphs/residual-plots/>



- 
- [46] «Algoritmo Calcula Disminución Promedio Impureza - FasterCapital». Accedido: 03 de mayo de 2024. [En línea]. Disponible en: <https://fastercapital.com/es/palabra-clave/algoritmo-calcula-disminuci%C3%B3n-promedio-impureza.html>
- [47] «Una introducción a los valores SHAP y a la interpretabilidad del machine learning | DataCamp». Accedido: 03 de mayo de 2024. [En línea]. Disponible en: <https://www.datacamp.com/es/tutorial/introduction-to-shap-values-machine-learning-interpretability>
- [48] L. Breiman, «Random forests», *Mach Learn*, vol. 45, n.º 1, pp. 5-32, oct. 2001, doi: 10.1023/A:1010933404324.
- [49] «La librería Pandas | Aprende con Alf». Accedido: 08 de mayo de 2024. [En línea]. Disponible en: <https://aprendeconalf.es/docencia/python/manual/pandas/>
- [50] «One Hot Encoding | Interactive Chaos». Accedido: 08 de mayo de 2024. [En línea]. Disponible en: <https://interactivechaos.com/en/node/1073>
- [51] Instituto Nacional de Seguridad y Salud en el Trabajo (INSST), O.A., M.P., «Calidad del Ambiente Interior en el trabajo», NIPO: 118-22-055-X, Madrid, España, octubre 2022.
- [52] «La agenda 2030 y los objetivos de desarrollo sostenible - mentorDay WikiTips». Accedido: 20 de junio de 2024. [En línea]. Disponible en: [https://mentorday.es/wikitips/agenda-2030-objetivos-desarrollo-sostenible/?gad\\_source=1&gclid=CjwKCAjwyo60BhBiEiwAHmVLJdqRGK6DkvrCDPAMW4kpkXVDkL6lwNfPgD9cKLZrtTtMR-zm1qdPOBoCVUcQAvD\\_BwE](https://mentorday.es/wikitips/agenda-2030-objetivos-desarrollo-sostenible/?gad_source=1&gclid=CjwKCAjwyo60BhBiEiwAHmVLJdqRGK6DkvrCDPAMW4kpkXVDkL6lwNfPgD9cKLZrtTtMR-zm1qdPOBoCVUcQAvD_BwE)



---