

UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN

Facultad de Ingeniería

Escuela Profesional de Ingeniería en Informática y Sistemas

**Analítica de datos y performance de sensores de calidad
de aire en el laboratorio de la Escuela Profesional de
Ingeniería en Informática y Sistemas UNJBG, 2025**

TESIS

Presentada por:

Bach. Ruth de Jesús Ramirez Rejas

Para obtener el Título Profesional de:

INGENIERO EN INFORMÁTICA Y SISTEMAS

TACNA – PERÚ

2026

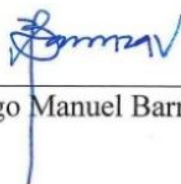
Universidad Nacional Jorge Basadre Grohmann
Facultad de Ingeniería
Escuela Profesional de Ingeniería en Informática y Sistemas

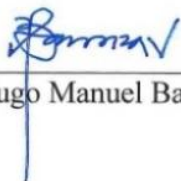
Analítica de datos y performance de sensores de calidad de aire en el laboratorio de la
Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG, 2025

Tesis presentada y aprobada el 17 de junio del 2026 estando el jurado calificador
integrado por:

Presidente : 
Dr. Edgar Aurelio Taya Acosta

Secretario : 
MSc. Arnold Christian Loaiza Fabian

Vocal : 
Dr. Hugo Manuel Barraza Vizcarra

Asesor : 
Dr. Hugo Manuel Barraza Vizcarra

CERTIFICADO DE SIMILITUD

Yo Hugo Manuel Barraza Vizcarra, en mi condición de asesor acreditado por la Resolución de Facultad N° 10427-2026-FAIN/UNJBG y Resolución de Facultad N° 9542-2025-FAIN/UNJBG de la tesis:

“Analítica de datos y performance de sensores de calidad de aire en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG, 2025”.

Presentado por la Bachiller Ruth de Jesús Ramirez Rejas para optar por el título profesional de Ingeniero en Informática y Sistemas.

Habiendo cumplido con lo establecido en el reglamento de originalidad y de similitud de trabajo de investigación y producción intelectual, considerando que según la revisión, evaluación y análisis realizado a través del **software de similitud textual** TURNITIN, cuenta con el **nivel de similitud permitido cuyo porcentaje es 8%**. Por lo que, **CERTIFICO LA SIMILITUD** de la tesis enunciado líneas arriba, la cual está expedita para continuar con los trámites para la obtención del Título Profesional según corresponda consiguientemente la publicación en el repositorio institucional.



Dr. Hugo Manuel Barraza Vizcarra
DNI: 46053783
ASESOR



Bach. Ruth de Jesús Ramirez Rejas
DNI: 72787626
TESISTA



Dedicatoria

*Dedico esta tesis a Dios, por brindarme
fortaleza, sabiduría y perseverancia para
alcanzar esta importante meta.*

*A mi madre, por su amor incondicional,
sacrificio, apoyo constante y por
acompañarme en cada etapa de mi vida y
formación profesional. Gracias por creer
siempre en mí y ser el pilar fundamental para
alcanzar este logro.*

*A mi padre, por su interés constante, sus
palabras de aliento y por acompañarme en
distintos momentos de mi formación
académica.*

*A mi familia y a cada una de las personas que,
de una u otra manera, estuvieron presentes
durante este proceso, brindándome ánimo,
confianza y motivación para alcanzar esta
meta.*

Agradecimientos

Mi especial agradecimiento al Dr. Hugo Manuel Barraza Vizcarra, asesor de tesis, por brindarme parte de su tiempo, su orientación y valiosos aportes durante el desarrollo de la investigación.

Asimismo, agradezco a los docentes de la Escuela Profesional de Ingeniería en Informática y Sistemas por facilitarme la construcción de mis conocimientos y habilidades durante mi formación profesional.

De manera especial, agradezco al proyecto de investigación “Determinación del impacto de la minería de datos, social media, internet de las cosas (IoT) y realidad aumentada en la educación superior universitaria” aprobada con Resolución Rectoral N°13626-2024-UNJBG, por el apoyo brindado para el desarrollo de esta tesis; así como al Dr. Edgar Aurelio Taya Acosta, investigador principal del proyecto, por su orientación, respaldo y valiosas recomendaciones durante la ejecución de esta investigación.

Índice temático

DEDICATORIA	II
AGRADECIMIENTOS	III
ÍNDICE TEMÁTICO	IV
ÍNDICE DE TABLAS	VIII
ÍNDICE DE FIGURAS	IX
RESUMEN	XI
ABSTRACT	XIII
INTRODUCCIÓN	1
CAPÍTULO I PLANTEAMIENTO DEL PROBLEMA	4
1.1 Antecedentes del problema	4
1.2 Descripción del problema	6
1.3 Formulación del problema	9
1.3.1 <i>Problema general</i>	9
1.3.2 <i>Problemas específicos</i>	9
1.4 Objetivos de la investigación	10
1.4.1 <i>Objetivo general</i>	10
1.4.2 <i>Objetivos específicos</i>	11
1.5 Justificación e importancia de la investigación	11
1.6 Limitaciones	13
1.7 Viabilidad del estudio	15
1.8 Formulación de la hipótesis	15

1.9	Variables	17
1.9.1	<i>Definición conceptual y operacional de las variables</i>	17
1.10	Operacionalización de variables	18
CAPÍTULO II MARCO TEÓRICO		19
2.1	Antecedentes del trabajo de investigación	19
2.2	Bases teóricas sobre el trabajo de investigación	24
2.2.1	<i>Performance de calidad de aire</i>	24
2.2.1.1	<i>Contaminación</i>	24
2.2.1.2	<i>Calidad de aire</i>	25
2.2.1.3	<i>Dióxido de carbono</i>	25
2.2.1.4	<i>Sensor CO₂</i>	26
2.2.1.5	<i>ESP32</i>	26
2.2.1.6	<i>Sensor K30</i>	27
2.2.1.7	<i>Sensor MQ135</i>	28
2.2.1.8	<i>PPM</i>	29
2.2.2	<i>Analítica de datos</i>	31
2.2.2.1	<i>Definición de analítica de datos</i>	31
2.2.2.2	<i>Series de tiempo</i>	31
2.2.2.3	<i>Red neuronal LSTM</i>	32
2.2.2.4	<i>Modelo ARIMAX</i>	33
2.3	Definiciones conceptuales	34
2.3.1	<i>Internet de las cosas (IoT)</i>	34
2.3.2	<i>ThingSpeak</i>	35
2.3.3	<i>Nube</i>	36
CAPÍTULO III MARCO METODOLÓGICO		37

3.1	Planteamiento metodológico	37
3.1.1	<i>Tipo de investigación</i>	37
3.1.2	<i>Nivel de la investigación</i>	37
3.1.3	<i>Diseño de la investigación</i>	38
3.2	Población y muestra	39
3.3	Equipos y materiales	39
3.4	Procedimiento de las pruebas experimentales	40
3.5	Técnicas e instrumentos de recolección de datos	42
3.6	Técnicas para el procesamiento y análisis de datos	43
CAPÍTULO IV RESULTADOS		46
4.1	Descripción de las pruebas experimentales	46
4.2	Presentación y análisis de los resultados	47
4.2.1	<i>Resultados para el indicador rendimiento de captura de datos</i>	47
4.2.2	<i>Resultados para el indicador sensibilidad</i>	51
4.2.3	<i>Resultados para el indicador estabilidad</i>	53
4.2.4	<i>Resultados para el indicador precisión de la predicción</i>	59
4.2.4.1	<i>Resultados del modelo ARIMAX</i>	60
4.2.4.2	<i>Resultados del modelo LSTM</i>	65
4.2.5	<i>Resultados complementarios</i>	67
4.3	Contrastación de hipótesis	70
CAPÍTULO V DISCUSIÓN		71
5.1	Pruebas de validación del modelo experimental	71
5.2	Aplicación de la tecnología encontrada	72
5.3	Contraste con trabajos de investigación similares	73

CONCLUSIONES	75
RECOMENDACIONES	78
REFERENCIAS BIBLIOGRÁFICAS	81
ANEXOS	87
ANEXO 1 MATRIZ DE CONSISTENCIA	87
ANEXO 2 AUTORIZACIÓN DE LA DIRECTORA DE LA ESCUELA PROFESIONAL DE INGENIERÍA EN INFORMÁTICA Y SISTEMAS	88
ANEXO 3 CORREO DE RESPUESTA CON LA INFORMACIÓN DE CANTIDAD DE ESTUDIANTES	89
ANEXO 4 DOCUMENTO DE SOLICITUD DE INFORMACIÓN DE POBLACIÓN ESTUDIANTEL	90
ANEXO 5 CÓDIGO DE GOOGLE COLAB	91

Índice de tablas

<i>Tabla N° 1: Operacionalización de variables.....</i>	<i>18</i>
<i>Tabla N° 2: Tipo de Equipo</i>	<i>40</i>
<i>Tabla N° 3: Indicadores de rendimiento de captura de datos de los sensores</i>	<i>48</i>
<i>Tabla N° 4: Correlación entre sensores y cantidad de alumnos.....</i>	<i>51</i>
<i>Tabla N° 5: Indicadores de estabilidad de los sensores</i>	<i>53</i>
<i>Tabla N° 6: Métricas del modelo ARIMAX aplicado al sensor K30.....</i>	<i>60</i>
<i>Tabla N° 7: Métricas del modelo ARIMAX aplicado al sensor MQ135</i>	<i>62</i>
<i>Tabla N° 8: Métricas del modelo LSTM aplicado a los sensores</i>	<i>65</i>

Índice de figuras

<i>Figura N° 1: Pabellón de la Escuela Profesional de Ingeniería Metalúrgica</i>	<i>14</i>
<i>Figura N° 2: Laboratorio Computación Aplicada</i>	<i>14</i>
<i>Figura N° 3: ESP32.....</i>	<i>27</i>
<i>Figura N° 4: Sensor K30.</i>	<i>28</i>
<i>Figura N° 5: MQ135.</i>	<i>29</i>
<i>Figura N° 6: Nivel PPM.</i>	<i>30</i>
<i>Figura N° 7: Diagrama de bloques ThingSpeak.</i>	<i>36</i>
<i>Figura N° 8: Diagrama de bloques de la investigación.</i>	<i>38</i>
<i>Figura N° 9: Prototipo de sensor K30 y MQ135.....</i>	<i>41</i>
<i>Figura N° 10: Ubicación del prototipo de sensor K30 y MQ135.....</i>	<i>42</i>
<i>Figura N° 11: Comparación de sensores K30 y MQ135 durante el periodo agosto – diciembre 2025.</i>	<i>49</i>
<i>Figura N° 12: Comparación de sensores K30 y MQ135 con la cantidad de Alumnos durante el periodo agosto – diciembre 2025.....</i>	<i>50</i>
<i>Figura N° 13: Sensores y cantidad de alumnos.</i>	<i>52</i>
<i>Figura N° 14: Boxplot de estabilidad de sensores.</i>	<i>54</i>

<i>Figura N° 15: Histograma del sensor K30.....</i>	<i>56</i>
<i>Figura N° 16: Histograma del sensor MQ135.</i>	<i>57</i>
<i>Figura N° 17: Detección de outliers del sensor K30 mediante KNN.</i>	<i>58</i>
<i>Figura N° 18: Detección de outliers del sensor MQ135 mediante KNN.....</i>	<i>59</i>
<i>Figura N° 19: Predicción del modelo ARIMAX aplicado al sensor K30.</i>	<i>61</i>
<i>Figura N° 20: Predicción del modelo ARIMAX aplicado al sensor MQ135.....</i>	<i>63</i>
<i>Figura N° 21: Predicción del modelo LSTM aplicado al gráfico real.</i>	<i>66</i>
<i>Figura N° 22: Promedio de sensores por hora del día.</i>	<i>68</i>
<i>Figura N° 23: Promedio de sensores por día de la semana.....</i>	<i>69</i>

RESUMEN

El presente trabajo de investigación se enfoca en comparar la performance de sensores de calidad de aire en la medición de dióxido de carbono (CO_2) en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas de la Universidad Nacional Jorge Basadre Grohmann. Se contrastó el desempeño de dos arquitecturas distintas de sensores: el K30, basado en tecnología NDIR y de mayor precisión, y el MQ135, de bajo costo y amplia disponibilidad. La metodología es de tipo aplicada, con enfoque descriptivo-comparativo y diseño no experimental. Los datos fueron recolectados durante el periodo del 18 de agosto al 31 de diciembre del 2025 mediante la placa de desarrollo ESP32, generando 464450 registros de datos que fueron enviados a un servicio en la nube gratuito ThingSpeak de MathWorks, que luego del preprocesamiento y limpieza se redujeron a 248172 de cada sensor. Se aplican técnicas de analítica de datos, incluyendo estadística descriptiva y modelos predictivos como LSTM y ARIMAX. Los resultados muestran que el K30 ofreció un mejor desempeño gracias a su alta sensibilidad y relación para capturar variaciones de CO_2 vinculadas al aforo de alumnos. En cambio, el MQ135 destacó por su estabilidad. Asimismo, para evaluar la precisión de la predicción de los niveles de CO_2 , los resultados mostraron que el modelo LSTM obtuvo un mejor desempeño predictivo,

alcanzando mayores valores de R^2 y menores errores MAE y RMSE en comparación con ARIMAX.

Palabras clave: Calidad de aire, performance de sensores, IoT, K30, MQ135, ESP32.

ABSTRACT

This research focuses on comparing the performance of air quality sensors in measuring carbon dioxide (CO₂) in the laboratory of the School of Computer and Systems Engineering at the Jorge Basadre Grohmann National University. The performance of two different sensor architectures was compared: the K30, based on NDIR technology and offering greater accuracy, and the MQ135, which is low-cost and widely available. The methodology is applied, with a descriptive-comparative approach and a non-experimental design. Data were collected from August 18 to December 31, 2025, using the ESP32 development board, generating 464,450 data records. These were sent to the free cloud service ThingSpeak by MathWorks, and after preprocessing and cleaning, were reduced to 248,172 records per sensor. Data analysis techniques were applied, including descriptive statistics and predictive models such as LSTM and ARIMAX. The results show that the K30 offered better performance thanks to its high sensitivity and ability to capture CO₂ variations linked to student occupancy. In contrast, the MQ135 stood out for its stability. Furthermore, to evaluate the accuracy of CO₂ level predictions, the results showed that the LSTM model achieved better predictive performance, reaching higher R² values and lower MAE and RMSE errors compared to ARIMAX.

Keywords: Air quality, sensor performance, IoT, K30, MQ135, ESP32.

INTRODUCCIÓN

La calidad del aire en ambientes cerrados se ha convertido en un tema de gran importancia debido a su influencia en la salud, el bienestar y el rendimiento académico de las personas. En espacios universitarios como aulas y laboratorios, la acumulación de dióxido de carbono (CO₂) puede incrementarse rápidamente debido a la concentración de estudiantes y la poca ventilación. Según Moreno et al. (2025), la exposición prolongada a niveles elevados de CO₂ en ambientes cerrados y con poca ventilación puede generar fatiga, dolores de cabeza, disminución de la concentración y afectar la capacidad de toma de decisiones, debido al deterioro de las funciones cognitivas relacionadas con la atención y la memoria.

Si en los próximos años no se realiza un monitoreo adecuado del CO₂ en ambientes cerrados, los problemas relacionados con la calidad del aire podrían volverse cada vez más frecuentes y difíciles de controlar. La acumulación constante del CO₂ en aulas, oficinas y laboratorios podría afectar de manera progresiva la salud y el rendimiento de las personas. Además, la ausencia de un monitoreo constante impediría identificar patrones de contaminación y tomar medidas preventivas a tiempo, lo que podría ocasionar ambientes poco saludables y una menor calidad de vida dentro de los espacios universitarios y laborales.

En los últimos años, se ha convertido en una gran alternativa para el monitoreo de CO₂ usar sensores de calidad de aire, ya que miden el ambiente en tiempo real. Entre los más comunes para medir el CO₂ son el K30 (que es muy preciso) y el MQ135 (que es de bajo costo y fácil de usar). Ambos se pueden conectar a un microcontrolador ESP32 para mandar y guardar todos los datos en la nube usando plataformas como ThingSpeak.

Asimismo, el crecimiento de la analítica de datos ha permitido aplicar técnicas avanzadas para el procesamiento y predicción de información ambiental. En esta investigación se emplearon modelos de series de tiempo como ARIMAX y redes neuronales LSTM con el propósito de analizar y predecir el comportamiento de los sensores de calidad de aire en función de las variaciones del ambiente y la cantidad de estudiantes presentes en el laboratorio.

La presente investigación tiene como objetivo comparar la performance de los sensores K30 y MQ135 en el monitoreo de calidad de aire dentro del laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas de la Universidad Nacional Jorge Basadre Grohmann, para determinar cuál ofrece un mejor desempeño en la medición de CO₂. Para ello, se analizaron indicadores como el rendimiento de captura de datos, sensibilidad y

estabilidad de los sensores; y para la analítica de datos el indicador será la precisión de sus predicciones.

La presente investigación se encuentra organizada en cinco capítulos. En el Capítulo I se desarrolla el planteamiento del problema, objetivos, hipótesis y justificación de la investigación. En el Capítulo II se presentan los antecedentes y bases teóricas relacionadas con sensores de calidad de aire, analítica de datos y modelos predictivos. En el Capítulo III se describe la metodología utilizada para la recolección y procesamiento de datos. En el Capítulo IV se muestran los resultados obtenidos y el análisis de los indicadores evaluados. Finalmente, en el Capítulo V se desarrolla la discusión de resultados, conclusiones y recomendaciones.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1 Antecedentes del problema

En los últimos años, la calidad del aire en espacios cerrados ha cobrado especial relevancia en el ámbito académico, dado que incide directamente en el desempeño académico, el bienestar general y la salud estudiantil. Fuentes-Ferragud et al. (2024) señalan que, en aulas universitarias españolas con ventilación limitada, los niveles de CO₂ tienden a incrementarse por encima de los valores recomendados debido a la concentración de personas y la falta de renovación adecuada del aire. La Organización Mundial de la Salud (2021) y American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) recomiendan mantener las concentraciones de CO₂ por debajo de 1,000 ppm en interiores, ya que niveles superiores pueden provocar fatiga, reducción de la atención y disminución del desempeño académico. Según Allen et al. (2016), en condiciones interiores con niveles elevados de dióxido de carbono (CO₂) y menor ventilación, las puntuaciones de función cognitiva fueron significativamente menores que en condiciones de mejor ventilación y menor CO₂, lo que sugiere un impacto negativo del CO₂ en la capacidad de desempeño en espacios cerrados.

Frente a esta problemática, se han desarrollado soluciones tecnológicas para monitorear el CO₂ en tiempo real. Los sensores de calidad del aire permiten realizar un monitoreo continuo de diferentes variables ambientales como dióxido de carbono (CO₂), partículas finas (PM2.5 y PM10), temperatura y humedad. Estos dispositivos proporcionan información en tiempo real sobre el estado del aire, lo que facilita la identificación de contaminantes y la implementación de medidas correctivas, como mejorar la ventilación o reducir fuentes de contaminación (Alhmiedat & Samara, 2017). Asimismo, el monitoreo constante contribuye a mantener niveles adecuados de confort y salud para los ocupantes de los espacios interiores.

Además, los sensores contribuyen a mejorar la eficiencia energética de los edificios. Por ejemplo, los sistemas de ventilación pueden ajustarse automáticamente en función de los niveles de CO₂ detectados en el ambiente. Este tipo de sistemas, conocidos como ventilación controlada por demanda, optimizan el consumo energético y mantienen condiciones adecuadas de aire interior para los ocupantes (Belimo, 2023).

Los sensores para medir CO₂ en investigaciones de bajo costo son el K30, que utiliza tecnología NDIR (infrarrojo no dispersivo), y el MQ135, un sensor de gas que responde a una variedad de gases contaminantes, incluyendo CO₂, amoníaco, benceno y monóxido de carbono. El sensor K30

es conocido por su alta precisión y estabilidad, aunque tiene un mayor costo; mientras que el MQ135 destaca por su bajo precio y disponibilidad, pero con limitaciones en precisión y sensibilidad específica al CO₂ (Spinelle et al., 2015; Hasenfratz et al., 2012).

En el ámbito nacional, aún son escasas las investigaciones que comparen directamente sensores de calidad de aire que midan el CO₂ como el K30 y MQ135 en entornos universitarios utilizando microcontroladores como el ESP32. Esta carencia evidencia la necesidad de desarrollar estudios que analicen su desempeño comparativo en condiciones reales de operación, con el fin de identificar soluciones precisas para el monitoreo ambiental en universidades públicas como la Universidad Nacional Jorge Basadre Grohmann.

1.2 Descripción del problema

En este tiempo la contaminación ambiental viene siendo una problemática muy crítica en nuestra sociedad, es un problema grave que crece con el tiempo, contribuyendo, por ejemplo, al efecto invernadero y al calentamiento global. Esta se origina tanto por gases contaminantes naturales como por la actividad humana, lo que afecta directamente al planeta. Existen diferentes tipos de contaminaciones con respecto al aire, pero uno de los más hablados es el dióxido de carbono (CO₂).

La contaminación por dióxido de carbono (CO₂) ha sido reconocida como un desafío ambiental de gran importancia debido a sus implicaciones tanto en la salud humana como en el cambio climático global. Si bien gran parte de la atención sobre las emisiones de CO₂ se ha centrado en su liberación a la atmósfera desde fuentes industriales y el transporte, la contaminación por CO₂ también es un problema en ambientes cerrados, particularmente en espacios como oficinas, aulas, hogares y otros entornos cerrados. En estos lugares, la acumulación de CO₂ puede generar efectos negativos en la salud de las personas y afectar su bienestar general.

El CO₂, un gas inodoro e invisible, puede alcanzar niveles elevados en espacios cerrados, lo que podría generar diversas molestias y efectos adversos en la salud, tales como fatiga, dolores de cabeza, mareos y dificultad para concentrarse. En casos más extremos, niveles elevados de CO₂ pueden causar asfixia o daño cerebral, aunque esto generalmente ocurre en concentraciones mucho mayores que las típicas en ambientes cerrados. Sin embargo, la exposición prolongada a niveles moderadamente elevados de CO₂ se asocia con una disminución en el rendimiento cognitivo, afectando la productividad y el aprendizaje, especialmente en entornos como las aulas universitarias y oficinas.

En la Universidad Nacional Jorge Basadre Grohmann (UNJBG) de Tacna, se ha identificado una problemática creciente relacionada con la contaminación del aire en ambientes cerrados, particularmente en lo que respecta a los niveles elevados de dióxido de carbono (CO₂) en sus aulas y otros espacios comunes. Esta situación afecta no solo la calidad del ambiente en el que los estudiantes, docentes y personal administrativo desarrollan sus actividades, sino también la salud y el rendimiento académico de la comunidad universitaria.

Uno de los principales factores que contribuyen a esta problemática es la ventilación inadecuada en varios edificios de la universidad. Muchos de estos espacios no cuentan con sistemas de ventilación eficientes, lo que impide la correcta circulación del aire y la expulsión de los contaminantes generados por la respiración de las personas. Como resultado, los niveles de CO₂ tienden a acumularse rápidamente, especialmente en aquellas aulas donde hay una alta concentración de estudiantes.

Además, la alta ocupación de los espacios en ciertos horarios de clases, como durante las horas de mayor demanda académica, genera un incremento en la emisión de CO₂ debido a la respiración de los estudiantes y docentes. Esta situación, unida a la falta de control en la cantidad de personas por aula, agrava aún más la calidad del aire interior.

Por otro lado, la falta de monitoreo y control efectivo de la calidad del aire en tiempo real agrava la situación, ya que no se cuenta con herramientas adecuadas para medir y gestionar los niveles de CO₂. Sin un sistema de medición confiable, es difícil tomar decisiones informadas sobre cómo mejorar las condiciones del ambiente interior.

En este contexto, resulta necesario abordar esta problemática mediante la comparación de sensores de calidad de aire en la medición de CO₂ en ambientes cerrados, con el fin de identificar cuál de ellos permite un mejor control de la calidad del aire en dichos espacios.

1.3 Formulación del problema

1.3.1 Problema general

¿Cuál es la performance de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG utilizando analítica de datos?

1.3.2 Problemas específicos

Para poder responder el problema general, se formularon los siguientes problemas específicos:

- ¿Cuál es el rendimiento de captura de datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG utilizando analítica de datos?
- ¿Cuál es la sensibilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG utilizando analítica de datos?
- ¿Cuál es la estabilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG utilizando analítica de datos?
- ¿Cuál es la precisión de la predicción de los niveles de CO₂ con los datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG utilizando analítica de datos?

1.4 Objetivos de la investigación

1.4.1 Objetivo general

Comparar la performance de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos

1.4.2 Objetivos específicos

- Determinar el rendimiento de captura de datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.
- Determinar la sensibilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos
- Determinar la estabilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos
- Determinar la precisión de la predicción de los niveles de CO₂ con los datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos

1.5 Justificación e importancia de la investigación

El estudio de este problema es relevante por varias razones:

Académica: Esta investigación contribuye al conocimiento científico sobre el uso y eficiencia de sensores de calidad de aire que midan el dióxido de carbono (CO₂), específicamente el MQ135 y el K30, en entornos educativos universitarios. A través del análisis comparativo y el uso de analítica de datos, se busca ofrecer evidencia empírica sobre el comportamiento y rendimiento de estos sensores en condiciones reales, aportando a futuras investigaciones relacionadas con la calidad del aire interior y su impacto en el aprendizaje.

Tecnológica: El estudio permite evaluar la viabilidad del uso de sensores de bajo y alto costo integrados con microcontroladores como el ESP32, promoviendo soluciones innovadoras para el monitoreo ambiental. Esta información resulta clave para el desarrollo de sistemas inteligentes y sostenibles de control de calidad del aire en espacios cerrados.

Práctica: Ofrece datos valiosos para directores de infraestructura, autoridades universitarias y responsables de salud ambiental, facilitando la toma de decisiones sobre qué tipo de sensor implementar según su costo y facilidad de integración. Esto puede llevar a una mejora directa de las condiciones ambientales en aulas, impactando positivamente en la salud y rendimiento cognitivo de los estudiantes.

Social: Promueve una mayor conciencia sobre la importancia de la ventilación adecuada y el control de la calidad del aire en espacios académicos, especialmente ante el aumento de enfermedades respiratorias y preocupaciones por ambientes cerrados post-pandemia. Implementar sensores eficientes podría contribuir a entornos de aprendizaje más seguros, saludables y sostenibles.

La presente investigación es importante porque permite establecer criterios objetivos para la selección de sensores de calidad de aire en la medición de CO₂ adecuados para espacios educativos, en función de su rendimiento de captura de datos, sensibilidad y estabilidad. Asimismo, aporta evidencia útil para la implementación de sistemas de monitoreo ambiental accesibles y eficientes en instituciones públicas, donde los recursos pueden ser limitados. Al identificar qué sensores ofrecen mejores resultados en condiciones reales, se favorece la adopción de tecnologías que mejoren la calidad del aire en las aulas, lo cual tiene implicancias directas en la salud, el bienestar y el rendimiento académico de los estudiantes universitarios.

1.6 Limitaciones

La investigación se realizó únicamente en el laboratorio de Computación aplicada de la Escuela Profesional de Ingeniería en Informática y Sistemas de la Universidad Nacional Jorge Basadre Grohmann, ubicado en

el tercer piso del pabellón de la Escuela Profesional de Ingeniería Metalúrgica, lo que limita la generalización de los resultados.



Figura N° 1: Pabellón de la Escuela Profesional de Ingeniería Metalúrgica
Fuente: Propia



Figura N° 2: Laboratorio de Computación Aplicada
Fuente: Propia

Solo se evaluaron los sensores MQ135 y K30 en condiciones específicas, sin considerar otros factores externos como variaciones climáticas o arquitectónicas.

1.7 Viabilidad del estudio

Se contó con el respaldo de las autoridades académicas, para realizar la instalación temporal de los sensores. Además, se contó con el equipo necesario (ESP32, sensores MQ135 y K30), equipos que forman parte del proyecto “Determinación del impacto de la minería de datos, social media, internet de las cosas (IoT) y realidad aumentada en la educación superior universitaria”, aprobado con Resolución Rectoral N° 13626-2024-UNJBG de fecha 20 de agosto del 2024, al cual me alineé como tesista. También cabe indicar que se dispone del conocimiento técnico para el desarrollo y análisis de los datos, lo que garantiza la factibilidad del estudio.

1.8 Formulación de la hipótesis

- H1: Existe diferencia en la performance de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.

- H0: No existe una diferencia en la performance de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.

Hipótesis Específica 1

- Existe diferencia en el rendimiento de captura de datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.

Hipótesis Específica 2

- Existe diferencia en la sensibilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.

Hipótesis Específica 3

- Existe diferencia en la estabilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.

Hipótesis Específica 4

- Existe diferencia en la precisión de la predicción de los niveles de CO₂ con los datos de los sensores MQ135 y K30 en el laboratorio de la

Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.

1.9 Variables

1.9.1 Definición conceptual y operacional de las variables

Variable principal: Performance de los sensores de calidad de aire

Definición conceptual

Según García y Rico (2019), los sensores de la calidad del aire son dispositivos electrónicos diseñados para detectar y medir la presencia y concentración de contaminantes atmosféricos, permitiendo obtener datos ambientales que facilitan el monitoreo y análisis de la contaminación del aire.

Definición operacional

La performance de los sensores puede evaluarse mediante parámetros como precisión, sensibilidad, estabilidad y tiempo de respuesta, los cuales permiten determinar la confiabilidad y calidad de las mediciones obtenidas por el dispositivo (Fraden, 2016).

Variable interviniente: Analítica de datos

Según Thomas H. Davenport y Jeanne G. Harris (2017), la analítica de datos consiste en el uso de datos, herramientas estadísticas y modelos predictivos para comprender fenómenos, identificar patrones y apoyar la toma

de decisiones basadas en información. La analítica de datos se entiende como el proceso de recopilación, procesamiento y análisis de datos con el fin de obtener información útil que permita comprender fenómenos y apoyar la toma de decisiones. En el ámbito tecnológico, permite interpretar los datos generados por sistemas y sensores para evaluar su funcionamiento

1.10 Operacionalización de variables

Tabla N° 1:
Operacionalización de variables

<i>Variables</i>	<i>Indicadores</i>
Variable principal:	Rendimiento de captura de datos
Performance de los sensores de calidad de aire	Estabilidad en el tiempo
	Sensibilidad
Variable Interviniente:	
Analítica de datos	Precisión de la predicción

Fuente: Elaboración propia

CAPÍTULO II

MARCO TEÓRICO

2.1 Antecedentes del trabajo de investigación

Diversos estudios recientes han demostrado un creciente interés por la evaluación de sensores de calidad de aire que midan el dióxido de carbono (CO₂) en entornos educativos, especialmente ante la preocupación por la calidad del aire en espacios cerrados. En este contexto, la microelectrónica de bajo costo ha permitido la incorporación de plataformas como el ESP32, un microcontrolador con conectividad Wi-Fi y Bluetooth que facilita la adquisición y transmisión de datos desde sensores hacia sistemas de visualización o almacenamiento en la nube (Hernández Arteaga et al. ,2025). Esta versatilidad ha hecho que el ESP32 sea ampliamente utilizado en diferentes aplicaciones de Internet de las Cosas (IoT) para el monitoreo mediante sensores, particularmente en entornos educativos.

Un estudio realizado por Martin et al. (2017) evaluó el desempeño de varios sensores SenseAir K30 para la medición de CO₂ en aire ambiente. Los autores compararon las mediciones del sensor con equipos de referencia de alta precisión y encontraron que, aunque inicialmente los sensores presentaban errores de medición, después de aplicar modelos de corrección ambiental la precisión mejoró significativamente, alcanzando errores entre

1.7 y 4.3 ppm, lo que demuestra que estos sensores pueden ser utilizados de manera confiable en sistemas de monitoreo ambiental.

De manera similar, Cai et al. (2024) analizaron el desempeño de sensores de dióxido de carbono de bajo costo para aplicaciones en ambientes interiores. Los resultados indicaron que sensores como el K30, después de un proceso de calibración, pueden proporcionar mediciones con un nivel de precisión adecuado para aplicaciones de monitoreo continuo en edificios inteligentes y sistemas de control de ventilación.

En un estudio desarrollado por Kumar et al. (2025) se diseñó un sistema de monitoreo de calidad del aire utilizando el sensor MQ135 conectado a un microcontrolador Arduino Uno. El sistema convierte las señales analógicas del sensor en valores de concentración expresados en partes por millón (ppm) y permite visualizar en tiempo real la calidad del aire mediante una pantalla LCD. Los resultados demostraron que el sistema puede detectar variaciones en la concentración de contaminantes y generar alertas cuando los niveles superan los valores establecidos.

Asimismo, Shrestha et al. (2022) desarrollaron un sistema de monitoreo de contaminación del aire utilizando el sensor MQ135 y un microcontrolador ATmega32A para evaluar la calidad del aire en una zona urbana de Katmandú. Los resultados mostraron que el sistema permite identificar

cambios en los niveles de contaminación atmosférica, lo que contribuye a la evaluación de la calidad del aire en diferentes áreas urbanas.

Un estudio realizado por Rochmania et al. (2021) desarrolló un sistema de monitoreo de dióxido de carbono basado en NodeMCU ESP8266 y sensor MQ135 para medir concentraciones de CO₂ en diferentes ubicaciones. Los resultados obtenidos mostraron concentraciones promedio cercanas a 549 ppm y 666 ppm, valores similares a los registrados por equipos de referencia, lo que demuestra que los sensores de bajo costo pueden utilizarse para monitoreo ambiental en sistemas IoT.

Un estudio realizado por Tyagi et al. (2026) comparó los modelos ARIMA y LSTM para la predicción de la calidad del aire en entornos urbanos, utilizando datos provenientes de sensores ambientales de contaminantes atmosféricos como CO₂ y material particulado. Los resultados evidenciaron que el modelo LSTM alcanzó un coeficiente de determinación (R^2) de aproximadamente 0.96, superando significativamente al modelo ARIMA, el cual presentó un menor nivel de ajuste y mayores errores de predicción. Esta diferencia en el desempeño se atribuye a la capacidad del modelo LSTM para capturar relaciones no lineales y dependencias temporales complejas en los datos. En contraste, el modelo ARIMA mostró limitaciones al asumir

relaciones lineales y condiciones de estacionariedad, lo que reduce su eficacia en escenarios con alta variabilidad.

En América Latina también se han desarrollado investigaciones orientadas al monitoreo de la calidad del aire mediante sensores de bajo costo en entornos universitarios. Pérez García (2024) desarrolló un sistema de monitoreo de contaminación del aire en el campus de la Universidad Técnica Estatal de Quevedo utilizando sensores de gases de la familia MQ integrados a una plataforma ESP8266. El sistema permitió registrar concentraciones de gases contaminantes en diferentes zonas del campus universitario, demostrando que los sensores de bajo costo pueden utilizarse para evaluar la calidad del aire en ambientes educativos y apoyar la gestión ambiental dentro de las instituciones académicas.

Estudios como el de Gutiérrez et al. (2021) evaluaron el desempeño de sensores MQ135 en ambientes educativos y concluyeron que, aunque útiles para detectar tendencias generales de calidad del aire, requieren calibración y ajustes para obtener resultados confiables.

En Perú, estudios como el de Mendoza Merma y Cordero Ochoa (2022), realizado en los ambientes interiores de la Facultad de Ingeniería de la Universidad Privada de Tacna, desarrollaron e implementaron un sistema IoT para el monitoreo remoto de la calidad del aire. Para la recolección de datos

se empleó el sensor BME680, el cual permite medir parámetros ambientales como la calidad del aire, temperatura, humedad y presión atmosférica. Los resultados del estudio permitieron registrar más de 54 000 datos almacenados en una base de datos, evidenciando que los valores de calidad del aire en los ambientes evaluados se encontraban dentro de los estándares de calidad ambiental establecidos, lo que demostró la viabilidad del sistema para el monitoreo continuo de ambientes interiores.

Así mismo, Luna, Talavera y Cano (2017), en un estudio realizado en el distrito de San Isidro, en la ciudad de Lima, investigadores del centro de investigación de la Universidad del Pacífico implementaron sensores electroquímicos de bajo costo para el monitoreo de la calidad del aire. El estudio permitió registrar concentraciones de contaminantes gaseosos en tiempo real, incluyendo CO₂, demostrando que este tipo de sensores puede ser utilizado como una alternativa viable para el monitoreo ambiental continuo y para la recolección de datos que faciliten el análisis de la calidad del aire en entornos urbanos.

Las investigaciones constituyen antecedentes relevantes que fundamentan el presente estudio, el cual se orienta a realizar un análisis comparativo entre los sensores MQ135 y K30 en aulas universitarias, haciendo uso del microcontrolador ESP32 como herramienta de integración

y procesamiento de datos. A diferencia de los estudios mencionados, esta investigación busca contextualizar el desempeño de ambos sensores en un mismo espacio físico bajo condiciones similares, lo que permitirá obtener conclusiones más precisas sobre su aplicabilidad técnica y económica en el monitoreo de la calidad del aire en la Universidad Nacional Jorge Basadre Grohmann.

2.2 Bases teóricas sobre el trabajo de investigación

2.2.1 Performance de calidad de aire

2.2.1.1 Contaminación

La contaminación se refiere a la presencia y acumulación de sustancias en el ambiente que afectan negativamente a los seres vivos y al equilibrio natural. Aunque normalmente se asocia a espacios abiertos, en la actualidad la contaminación en ambientes cerrados ha cobrado gran relevancia debido al aumento de actividades realizadas en interiores, como oficinas, escuelas y universidades. Esta contaminación puede originarse por fuentes internas, como la exhalación humana, materiales de construcción, productos de limpieza y deficiencias en la ventilación. La Organización Mundial de la Salud (2021) advierte que la contaminación del aire interior representa un riesgo significativo para la salud pública, contribuyendo al desarrollo de

enfermedades respiratorias, cardiovasculares y disminuyendo la calidad de vida en general.

2.2.1.2 Calidad de aire

La calidad del aire interior es un aspecto crítico que incide directamente en el confort, salud y productividad de las personas que permanecen largos periodos en espacios cerrados. Esta calidad depende de diversos factores, como la ventilación, la cantidad de personas en un lugar, el uso de materiales contaminantes y la existencia de fuentes de emisión de gases. En aulas universitarias, donde la concentración de estudiantes es elevada, el monitoreo del aire se vuelve fundamental para garantizar un ambiente adecuado para el aprendizaje. Según Sundell (2004), una buena calidad del aire reduce el ausentismo escolar, mejora el estado de ánimo y previene enfermedades respiratorias, siendo un elemento clave en entornos educativos eficaces.

2.2.1.3 Dióxido de carbono

El dióxido de carbono (CO₂) es un gas esencial para el equilibrio natural. Es emitido principalmente a través de la respiración humana y actividades como la combustión o el uso de equipos electrónicos. Aunque no es tóxico en bajas concentraciones, niveles superiores a 1000 ppm pueden causar somnolencia, pérdida de concentración y malestar general. Estudios han demostrado que una alta concentración de CO₂ en aulas afecta

directamente la atención y el rendimiento de los estudiantes (Li et al., 2020). En ese sentido, controlar la presencia de este gas en entornos educativos se vuelve una necesidad urgente para preservar el bienestar y el aprendizaje.

2.2.1.4 Sensor CO₂

Los sensores de CO₂ son dispositivos que permiten medir de forma directa la cantidad de dióxido de carbono presente en un ambiente determinado. Existen distintos tipos de sensores, siendo los más comunes los sensores electroquímicos y los de tecnología NDIR (infrarrojo no dispersivo). El sensor MQ135 es ampliamente utilizado por su bajo costo y versatilidad, aunque su precisión puede verse afectada por la presencia de otros gases. En contraste, el sensor K30, basado en tecnología NDIR, ofrece mediciones más estables y confiables, aunque a un mayor costo. La elección de un sensor adecuado dependerá del objetivo del monitoreo y de las condiciones del entorno. Li et al. (2020) concluyen que, si bien los sensores económicos pueden ser útiles en investigaciones preliminares, para aplicaciones críticas es recomendable optar por sensores con mejor precisión y menor margen de error.

2.2.1.5 ESP32

El ESP32 es un microcontrolador de bajo costo y alto rendimiento desarrollado por Espressif Systems. Está diseñado para aplicaciones del

Internet de las Cosas (IoT), ya que cuenta con conectividad Wi-Fi y Bluetooth integrada, así como múltiples pines de entrada/salida y capacidades para la adquisición de datos. Su arquitectura permite el procesamiento rápido de datos y el control eficiente de sensores, lo que lo convierte en una plataforma ideal para proyectos de monitoreo ambiental. Gracias a su flexibilidad, bajo consumo energético y compatibilidad con plataformas de desarrollo como Arduino o MicroPython, el ESP32 es ampliamente utilizado en entornos académicos y de investigación (Aggarwal & Kumar, 2020).



Figura N° 3: ESP32
Fuente: Saravia B. (2019)

2.2.1.6 Sensor K30

El sensor K30 es un dispositivo de medición de dióxido de carbono (CO_2) basado en la tecnología de infrarrojo no dispersivo (NDIR). Este sensor destaca por su alta precisión, estabilidad a largo plazo y confiabilidad en entornos donde se requiere una medición constante de la calidad del aire. Está

diseñado para operar en una amplia gama de aplicaciones, incluyendo laboratorios, aulas, invernaderos y sistemas de ventilación. Una de sus ventajas más notables es su capacidad de proporcionar datos en tiempo real con bajo margen de error, lo cual lo convierte en una excelente opción para investigaciones científicas. Según Li et al. (2020), el K30 presenta una desviación mínima en la lectura de CO₂, lo que lo posiciona como una herramienta confiable para el monitoreo ambiental en interiores.



*Figura N° 4: Sensor K30.
Fuente: CO2Meter.com*

2.2.1.7 Sensor MQ135

El sensor MQ135 es un dispositivo de bajo costo diseñado para detectar gases como amoníaco, óxidos de nitrógeno, alcohol, benceno, humo y dióxido de carbono (CO₂), aunque su precisión es limitada para lecturas específicas

de CO₂ debido a su sensibilidad a múltiples gases. Es ampliamente utilizado en proyectos escolares o de prototipado donde se requiere un monitoreo básico de la calidad del aire. Su bajo precio y fácil integración con microcontroladores como el ESP32 lo hacen atractivo, pero su fiabilidad puede verse afectada por factores como la temperatura, humedad o interferencia de otros compuestos (Huang et al., 2019). Por ello, se recomienda calibrarlo adecuadamente y utilizarlo como una herramienta de referencia más que de medición exacta.



*Figura N° 5: MQ135.
Fuente: CO2Meter.com*

2.2.1.8 PPM

El término “partes por millón” (ppm) es una unidad de medida que indica la concentración de una sustancia en otra. En el contexto de la calidad

del aire, las ppm se utilizan comúnmente para expresar la cantidad de dióxido de carbono (CO₂) presente en el aire. Por ejemplo, una concentración de 1,000 ppm de CO₂ significa que hay 1,000 moléculas de CO₂ por cada millón de moléculas de aire. Esta unidad es crucial para evaluar el nivel de ventilación y los posibles efectos en la salud de los ocupantes de un espacio cerrado. Según la Organización Mundial de la Salud (2021), niveles superiores a 1,000 ppm de CO₂ pueden causar somnolencia, malestar y disminución del rendimiento cognitivo.

CO2 [ppm]	Air Quality
2100	MALO Ambiente altamente contaminado. Se requiere ventilación.
2000	
1900	
1800	
1700	
1600	
1500	MEDIOCRE Ambiente contaminado. Se recomienda ventilación.
1400	
1300	
1200	
1100	
1000	NORMAL
900	
800	BUENO
700	
600	EXCELENTE
500	
400	

Figura N° 6: Nivel PPM.
Fuente: Gil García y Alemán Ramos (2019).

2.2.2 Analítica de datos

2.2.2.1 Definición de analítica de datos

La analítica de datos se refiere al proceso de recopilación, procesamiento y análisis de grandes volúmenes de datos con el objetivo de obtener información útil para la toma de decisiones y la comprensión de determinados fenómenos. Este proceso implica la aplicación de métodos estadísticos, herramientas informáticas y técnicas de análisis que permiten identificar patrones, tendencias y relaciones dentro de los datos recopilados (Kelleher & Tierney, 2018).

2.2.2.2 Series de tiempo

Las series de tiempo se definen como un conjunto de observaciones registradas de forma secuencial a lo largo del tiempo. Estas observaciones mantienen una dependencia temporal, lo que significa que los valores actuales pueden estar relacionados con valores anteriores. El análisis de series de tiempo permite identificar patrones, tendencias o comportamientos repetitivos con el fin de comprender el fenómeno estudiado y realizar predicciones futuras (Jason Brownlee, 2017).

2.2.2.3 Red neuronal LSTM

Las redes neuronales LSTM (Long Short-Term Memory) son un tipo de red neuronal recurrente (RNN) diseñadas para aprender dependencias a largo plazo en datos secuenciales, como series temporales, señales de sensores o datos ambientales. Este modelo introduce una celda de memoria que permite almacenar información relevante durante largos periodos de tiempo mediante mecanismos llamados puertas (gates), que regulan el flujo de información dentro de la red. Estas puertas controlan qué información debe olvidarse, qué información nueva debe almacenarse y qué información debe producir la salida del modelo (Sepp Hochreiter & Jürgen Schmidhuber, 1997).

Matemáticamente, el funcionamiento de una celda LSTM se describe mediante las siguientes ecuaciones:

Puerta de olvido

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

Puerta de entrada

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

Candidato al estado de memoria

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

Actualización del estado de la celda

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

Puerta de salida

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

Estado oculto de salida

$$h_t = o_t \cdot \tanh(C_t)$$

Donde x_t representa la entrada en el tiempo t , h_t el estado oculto, C_t el estado de la memoria de la celda, W los pesos del modelo, b los sesgos y σ la función sigmoide (Hochreiter & Schmidhuber, 1997).

2.2.2.4 Modelo ARIMAX

El modelo ARIMAX es una extensión del modelo ARIMA que permite incorporar variables exógenas, es decir, variables externas que pueden influir en el comportamiento de una serie temporal. Este modelo combina componentes autorregresivos (AR), diferenciación para lograr estacionariedad (I) y media móvil (MA), junto con variables adicionales que

ayudan a mejorar la capacidad de predicción del modelo en el análisis de series de tiempo (George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel & Greta M. Ljung, 2016).

Matemáticamente, el modelo ARIMAX(p,d,q) puede representarse como:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{k=1}^m \beta_k x_{k,t} + \varepsilon_t$$

donde y_t es el valor de la serie en el tiempo t , $X_{k,t}$ representa las variables exógenas, ϕ_i y θ_j son los parámetros del modelo, y ε_t corresponde al término de error.

2.3 Definiciones conceptuales

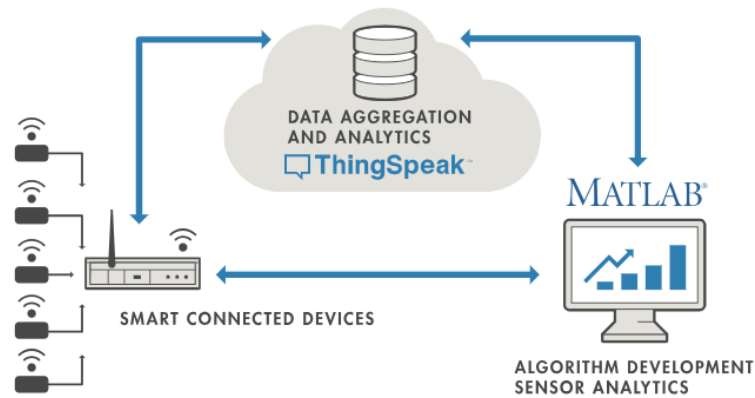
2.3.1 Internet de las cosas (IoT)

El Internet de las cosas (IoT) es una red de dispositivos interconectados que recopilan e intercambian datos a través de Internet. Esta tecnología ha revolucionado múltiples áreas, incluyendo la salud, la agricultura, la industria y la educación. En el contexto de la calidad del aire, el IoT permite la implementación de sistemas inteligentes de monitoreo, donde sensores recolectan información sobre la concentración de gases como el CO₂ y la envían a plataformas digitales para su análisis. Esto facilita la toma de

decisiones en tiempo real y permite implementar mejoras inmediatas en ventilación o control ambiental. Gubbi et al. (2013) destacan que el IoT es una herramienta esencial para crear entornos más sostenibles y saludables gracias a su capacidad de automatizar procesos y reducir la intervención humana.

2.3.2 ThingSpeak

ThingSpeak es una plataforma de código abierto para la recolección, almacenamiento y visualización de datos en tiempo real a través de Internet. Es especialmente utilizada en proyectos de IoT, ya que permite a dispositivos como el ESP32 enviar datos mediante el protocolo HTTP o MQTT hacia la nube. La plataforma proporciona herramientas para el análisis de datos y la generación de gráficos que permiten interpretar fácilmente las mediciones recogidas por sensores. Gracias a su interfaz intuitiva y su compatibilidad con entornos como MATLAB, ThingSpeak es una opción eficaz para investigadores y estudiantes que desean desarrollar soluciones de monitoreo en tiempo real (MathWorks, 2022).



*Figura N° 7: Diagrama de bloques ThingSpeak.
Fuente: ThingSpeak*

2.3.3 Nube

La computación en la nube consiste en el uso de servidores remotos para almacenar, procesar y gestionar información, sin depender de infraestructura física local. Esta tecnología permite acceder a datos desde cualquier lugar del mundo con conexión a Internet, lo que la convierte en una herramienta poderosa para proyectos de monitoreo ambiental. En investigaciones como la presente, el uso de la nube facilita el seguimiento en tiempo real de la calidad del aire, el almacenamiento de grandes volúmenes de información y la generación de alertas automáticas ante niveles peligrosos de CO₂. Armbrust et al. (2010) destacan que la nube proporciona flexibilidad, escalabilidad y accesibilidad, elementos clave para cualquier sistema IoT orientado a la recolección y análisis de datos ambientales.

CAPÍTULO III

MARCO METODOLÓGICO

3.1 Planteamiento metodológico

3.1.1 Tipo de investigación

La investigación es de tipo aplicada porque busca resolver un problema real relacionado con el monitoreo del CO₂ en el laboratorio de Computación Aplicada de la Escuela Profesional de Ingeniería en Informática y Sistemas, de acuerdo con Hernández Sampieri, Fernández Collado y Baptista Lucio (2014), la investigación aplicada se orienta a la solución de problemas prácticos y específicos, ya que busca utilizar los conocimientos científicos para intervenir en la realidad y generar mejoras concretas en contextos determinados.

3.1.2 Nivel de la investigación

El estudio corresponde al nivel descriptivo comparativo, debido a que tiene como finalidad describir y comparar el comportamiento de dos sensores de calidad de aire en la medición del CO₂ (MQ135 y K30) en un ambiente universitario. A través de la recopilación de datos en tiempo real, se analizarán y compararán indicadores como el rendimiento de captura de datos, sensibilidad, estabilidad. Según Moisés Ríos et al. (2019), el nivel descriptivo comparativo “interroga sobre el comportamiento de los valores de

la variable en grupos, espacios y periodos de tiempo diferentes”, permitiendo identificar similitudes y diferencias entre los elementos estudiados.

3.1.3 Diseño de la investigación

El diseño de la presente investigación es no experimental, longitudinal y comparativo, ya que no se manipulan las variables, sino que se observan y analizan en su contexto natural en el tiempo de agosto a diciembre del 2025. El estudio se centra en evaluar la performance de los sensores de calidad de aire en la medición de CO₂, comparando el comportamiento de los sensores MQ135 y K30 a partir de los datos recolectados en el laboratorio, los cuales serán analizados mediante técnicas de analítica de datos (Hernández, Fernández y Baptista, 2014).

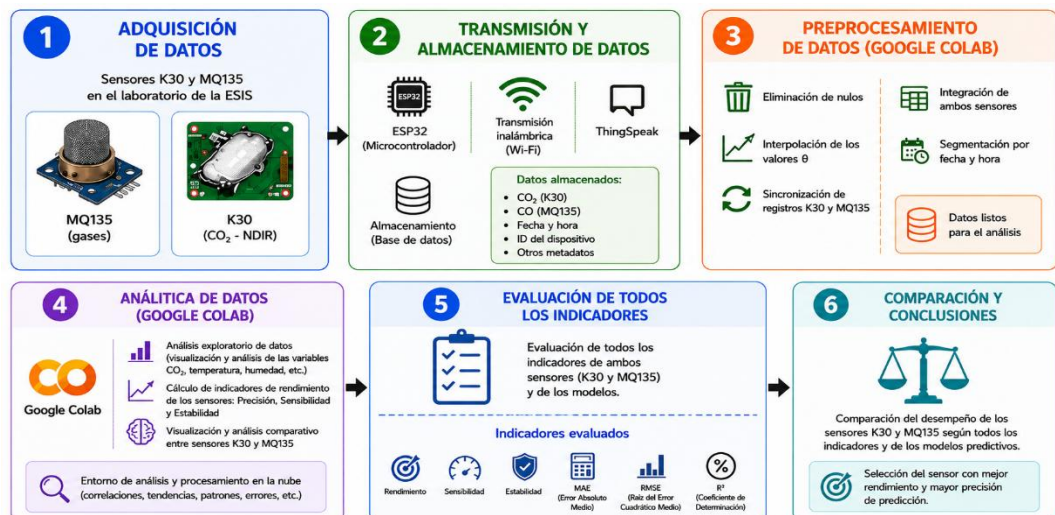


Figura N° 8: Diagrama de bloques de la investigación.
Fuente: Propia

3.2 Población y muestra

Según Hernández, Fernández y Baptista (2014), la población se entiende como el conjunto de todos los elementos que cumplen con ciertas características para ser objeto de estudio. En el presente estudio, la población está constituida por 464450 registros de los sensores de calidad de aire disponibles para evaluar su performance en aulas universitarias.

Dado que el objetivo es realizar un análisis comparativo, la muestra esta conformada por 248172 registros de datos válidos de los sensores K30 y MQ135, obtenidos mediante muestreo no probabilístico. Estos sensores fueron seleccionados por su disponibilidad y su uso frecuente en investigaciones sobre monitoreo de calidad del aire. Estudios previos han comparado ambos sensores, destacando la fácil disponibilidad de ellos y el bajo costo del MQ135 (Luca Spinelle et al., 2015; David Hasenfratz et al., 2012), lo que facilita su implementación en entornos académicos y proyectos experimentales. Además, representan diferentes tecnologías de medición, lo que permite analizar y comparar su desempeño en la detección de contaminantes.

3.3 Equipos y materiales

Para el desarrollo de la presente investigación se emplearon diversos equipos tecnológicos y herramientas digitales orientadas a la captura, transmisión, almacenamiento y análisis de datos relacionados con la calidad

del aire. Entre los principales recursos utilizados se encuentran sensores de medición de CO₂, un microcontrolador para la adquisición de datos y plataformas en la nube para el almacenamiento y procesamiento de la información.

Tabla N° 2:
Tipo de Equipo

<i>Tipo de equipo</i>	<i>Descripción</i>
Sensores de calidad de aire	K30
	MQ135
Microcontrolador	ESP32
Servicio Web	ThingSpeak
	Google Colab

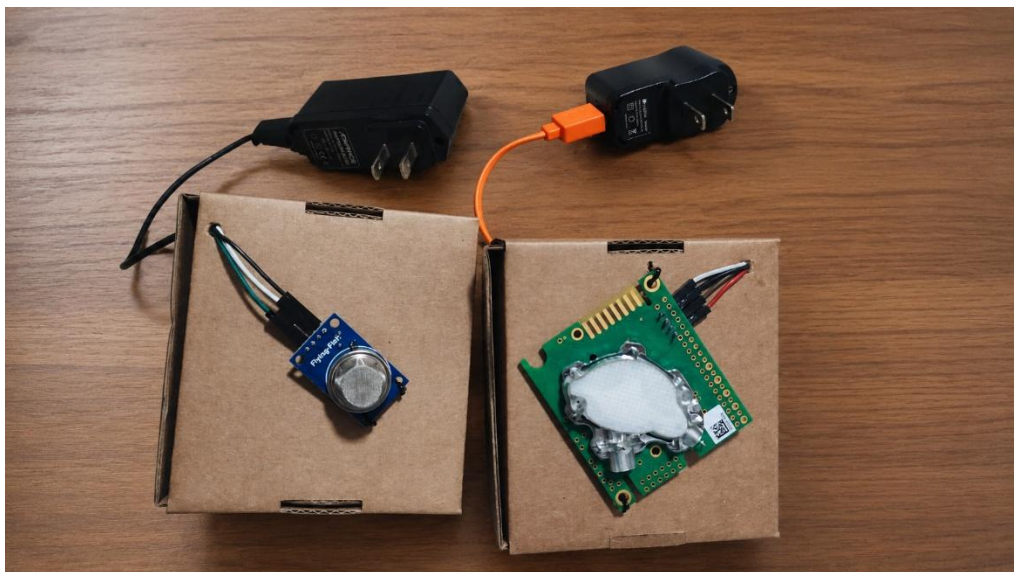
Fuente: Elaboración propia

3.4 Procedimiento de las pruebas experimentales

La presente investigación corresponde a un estudio no experimental; sin embargo, fue necesario desarrollar un procedimiento técnico para la obtención y monitoreo de los datos de calidad del aire. Para ello, se implementó un prototipo compuesto por los sensores K30 y MQ135, conectados a un microcontrolador ESP32, los sensores y el microcontrolador fueron colocados dentro de pequeñas carcasas elaboradas para el proyecto. Posteriormente, el sistema fue instalado en la parte superior del Laboratorio de Computación Aplicada, ubicado en el tercer piso del pabellón de la Escuela Profesional de Ingeniería Metalúrgica. La ubicación en la parte del techo

permitió obtener mediciones más estables y representativas del ambiente general del laboratorio.

Asimismo, el sistema fue alimentado mediante corriente eléctrica continua utilizando adaptadores de energía. Para garantizar la conectividad a internet y el envío de datos hacia la plataforma ThingSpeak, se utilizó un dispositivo móvil que compartía conexión Wi-Fi a través de su chip telefónico. Finalmente, los datos recolectados fueron almacenados en la nube para su posterior procesamiento y análisis en Google Colab.



*Figura N° 9: Prototipo de sensor K30 y MQ135.
Fuente: Propia*



*Figura N° 10: Ubicación del prototipo de sensor K30 y MQ135.
Fuente: Propia*

3.5 Técnicas e instrumentos de recolección de datos

Los instrumentos utilizados para la recolección de datos fueron los sensores K30 y MQ135. Ambos sensores generaron dos conjuntos de datos (datasets) independientes. Ambos dataset tienen 248172 registros para cada sensor.

Adicionalmente, se incorporó la variable "cantidad de alumnos" con el propósito de analizar la relación entre la ocupación del laboratorio y las concentraciones de CO₂ registradas por los sensores. Esta información fue obtenida mediante una solicitud realizada a la Dirección de Servicios y

Procesos Académicos de la Universidad Nacional Jorge Basadre Grohmann, en el marco del proyecto de investigación “Determinación del Impacto de la Minería de Datos, Social Media, Internet de las Cosas (IoT) y Realidad Aumentada en la Educación Superior Universitaria”. La información proporcionada incluyó los horarios académicos, asignaturas programadas y cantidad de estudiantes matriculados por sección y grupo durante el periodo del segundo semestre del 2025. Después los registros fueron organizados y sincronizados con las mediciones temporales generadas por los sensores K30 y MQ135, permitiendo su utilización en los análisis estadísticos y modelos predictivos desarrollados en la investigación.

3.6 Técnicas para el procesamiento y análisis de datos

Para el procesamiento y análisis de los datos recolectados por los sensores K30 y MQ135, se emplearon diversas técnicas de preprocesamiento y analítica de datos con la finalidad de mejorar la calidad y confiabilidad de la información obtenida en tiempo real. En una primera etapa, se filtraron los datos correspondientes al II semestre del año 2025, del 18 de agosto al 31 de diciembre, para esta investigación, luego se realizó la limpieza de datos, identificando registros nulos que podían generar inconsistencias durante el análisis. Debido a ello, se aplicaron técnicas de imputación de datos mediante interpolación, permitiendo estimar valores cero a partir de los registros cercanos y mantener la continuidad de las mediciones de CO₂ en el tiempo.

Posteriormente, se utilizaron técnicas de integración y sincronización de datos para unir la información generada por ambos sensores en un mismo intervalo temporal, facilitando la comparación directa de su comportamiento bajo las mismas condiciones ambientales. Asimismo, se aplicaron técnicas de normalización y organización temporal de los registros, con el propósito de preparar adecuadamente los datos para las etapas de análisis y modelado predictivo.

Como parte de la analítica de datos, se desarrolló un análisis exploratorio de datos (EDA), utilizando gráficos y medidas estadísticas descriptivas, calculando medidas como media, desviación estándar, valores mínimos y máximos, para identificar patrones, variaciones, tendencias y posibles cambios en la concentración de CO₂ dentro del laboratorio. Finalmente, se aplicaron técnicas de modelado predictivo mediante los algoritmos ARIMAX y LSTM, los cuales permitieron analizar la capacidad de predicción de los sensores. Para el entrenamiento y evaluación de los modelos se realizó una partición temporal de los datos, destinando el 80% de los registros al conjunto de entrenamiento y el 20% al conjunto de prueba. Según Aurélien Géron (2022), la partición de los datos en un 80% para entrenamiento y un 20% para prueba permite que el modelo aprenda patrones a partir de una cantidad suficiente de datos históricos y luego sea evaluado

con información no utilizada durante el entrenamiento, proporcionando una estimación más objetiva de su capacidad predictiva.

En el caso específico del modelo ARIMAX, se utilizó como variable exógena la cantidad de alumnos presentes en el laboratorio para cada intervalo temporal. Esta variable fue incorporada debido a su relación directa con la generación de dióxido de carbono en ambientes cerrados.

Para evaluar el desempeño de los modelos y la precisión de las predicciones, se utilizaron métricas como MAE, RMSE y R^2 , permitiendo comparar el comportamiento y desempeño de los sensores K30 y MQ135.

CAPÍTULO IV

RESULTADOS

4.1 Descripción de las pruebas experimentales

Las pruebas experimentales se realizaron en el laboratorio de Computación Aplicada de la Escuela Profesional de Ingeniería en Informática y Sistemas de la Universidad Nacional Jorge Basadre Grohmann durante el periodo del 18 de agosto al 31 de diciembre del 2025. Para el monitoreo de la calidad del aire se utilizaron los sensores K30 y MQ135 conectados a un ESP32, el cual permitió la captura y transmisión de datos hacia la plataforma ThingSpeak para su almacenamiento en la nube.

El sensor K30 fue utilizado para medir concentraciones de dióxido de carbono mediante tecnología NDIR, mientras que el sensor MQ135 permitió detectar gases asociados a la calidad del aire. Ambos sensores fueron programados para registrar datos diarios en el laboratorio, en intervalos de 15 segundos.

Durante el proceso de recolección se obtuvieron 464450 registros correspondientes a mediciones de CO₂, fecha y hora por cada sensor. Posteriormente, los datos fueron procesados utilizando Python en Google Colab, donde se aplicaron técnicas de preprocesamiento orientadas a mejorar la calidad y consistencia de la información obtenida por los sensores K30 y

MQ135. Entre las técnicas empleadas se realizaron la eliminación de registros nulos, la sincronización temporal de las mediciones, la imputación de valores faltantes mediante interpolación y la integración de datos por fecha y hora. Estas técnicas permitieron mantener la continuidad de los registros, reducir la pérdida de información y organizar adecuadamente los datos para su análisis estadístico y predictivo. Asimismo, se incorporó información relacionada con la cantidad de estudiantes presentes en el laboratorio durante cada horario académico. Esta información fue obtenida de los registros proporcionados por la Dirección de Servicios y Procesos Académicos de la Universidad Nacional Jorge Basadre Grohmann, con el objetivo de analizar la relación existente entre el aforo del laboratorio y las variaciones del CO₂. Para ello, se aplicaron técnicas de analítica de datos y modelos predictivos como ARIMAX y LSTM.

4.2 Presentación y análisis de los resultados

4.2.1 Resultados para el indicador rendimiento de captura de datos

Con el propósito de evaluar el rendimiento de captura de datos de los sensores K30 y MQ135, se calcularon medidas descriptivas relacionadas con el promedio de valores registrados, valores mínimos, máximos y cantidad de registros obtenidos durante el monitoreo ambiental.

La Tabla 3 muestra los principales indicadores estadísticos relacionados con el rendimiento de captura de datos de los sensores utilizados durante el monitoreo ambiental.

Tabla N° 3:
Indicadores de rendimiento de captura de datos de los sensores

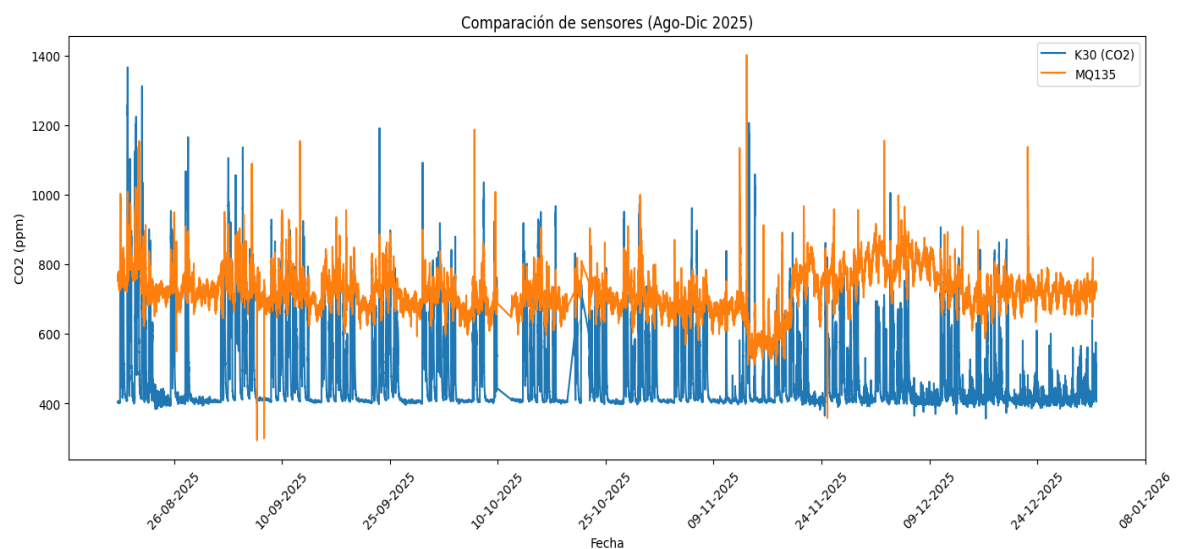
<i>Sensor</i>	<i>Promedio</i>	<i>Valor Mínimo</i>	<i>Valor Máximo</i>	<i>Cantidad</i>
K30	480.18	356	1366	248172
MQ135	721.57	293	1401	248172

Fuente: Elaboración propia

Se observa que el sensor K30 presentó un promedio de 480.18 ppm, mientras que el sensor MQ135 registró un promedio de 721.57 ppm, lo que sugiere que el MQ135 tiene una tendencia a sobreestimar las concentraciones de CO₂ o a responder a otros gases presentes en el ambiente. Por otro lado, aunque ambos sensores registran valores máximos similares (superiores a 1300 ppm), el K30 muestra valores mínimos más coherentes con condiciones ambientales normales, lo que indica que tiene mejor respuesta frente a variaciones asociadas al aforo o mayor relación con la ocupación del laboratorio.

La Figura 11 muestra el comportamiento temporal de los sensores K30 y MQ135 durante el periodo comprendido entre el 18 de agosto y 31 de diciembre del 2025. Se observa que ambos sensores presentan variaciones relacionadas con los cambios de CO₂ dentro del laboratorio. El sensor K30

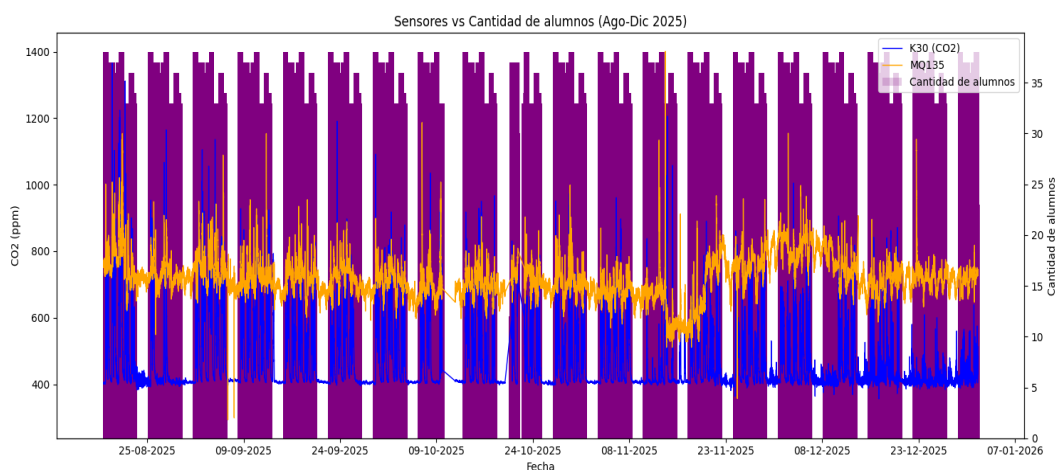
mostró picos superiores a 1200 ppm durante determinados horarios, especialmente en momentos de alta ocupación estudiantil. Por su parte, el sensor MQ135 presentó valores generalmente más elevados y fluctuaciones continuas a lo largo del periodo de monitoreo.



*Figura N° 11: Comparación de sensores K30 y MQ135 durante el periodo agosto – diciembre 2025.
Fuente: Propia*

Mirando también la figura 12 se muestra el comportamiento de los sensores K30 y MQ135 en relación con la cantidad de alumnos presentes en el laboratorio entre agosto y diciembre de 2025. Se observa que el sensor K30 mantiene valores relativamente estables alrededor de 400 ppm, aunque presenta algunos incrementos bruscos en determinados momentos. Por otro lado, el MQ135 registra una mayor variabilidad y valores más elevados, llegando en algunos casos por encima de los 1000 ppm, lo que indica una

mayor sensibilidad frente a cambios en el ambiente o presencia de otros gases. Asimismo, cuando aumenta la cantidad de alumnos, especialmente en ciertos periodos de noviembre y diciembre, ambos sensores tienden a mostrar incrementos en sus mediciones, evidenciando una posible relación entre la ocupación del laboratorio y la concentración de gases en el ambiente.



*Figura N° 12: Comparación de sensores K30 y MQ135 con la cantidad de alumnos durante el periodo agosto – diciembre 2025.
Fuente: Propia*

Los resultados evidencian que el sensor K30 presenta una mayor capacidad para detectar variaciones bruscas de dióxido de carbono dentro del ambiente, mientras que el MQ135 responde de manera más general frente a cambios ambientales. Esto permite inferir que el K30 posee un mejor rendimiento de captura de datos para el monitoreo específico de CO₂ en ambientes cerrados.

4.2.2 Resultados para el indicador sensibilidad

Con el objetivo de analizar la sensibilidad de los sensores frente a cambios relacionados con la ocupación del laboratorio, se calculó la correlación entre las mediciones de CO₂ y la cantidad de alumnos presentes durante el monitoreo.

La Tabla 4 presenta la correlación obtenida entre los sensores K30, MQ135 y la cantidad de alumnos presentes en el laboratorio. La variable cantidad de alumnos fue obtenida de los registros académicos oficiales proporcionados por la Dirección de Servicios y Procesos Académicos de la Universidad Nacional Jorge Basadre Grohmann y fue incorporada debido a que la ocupación de un ambiente cerrado influye directamente en la generación de dióxido de carbono producto de la respiración humana.

Tabla N° 4:
Correlación entre sensores y cantidad de alumnos

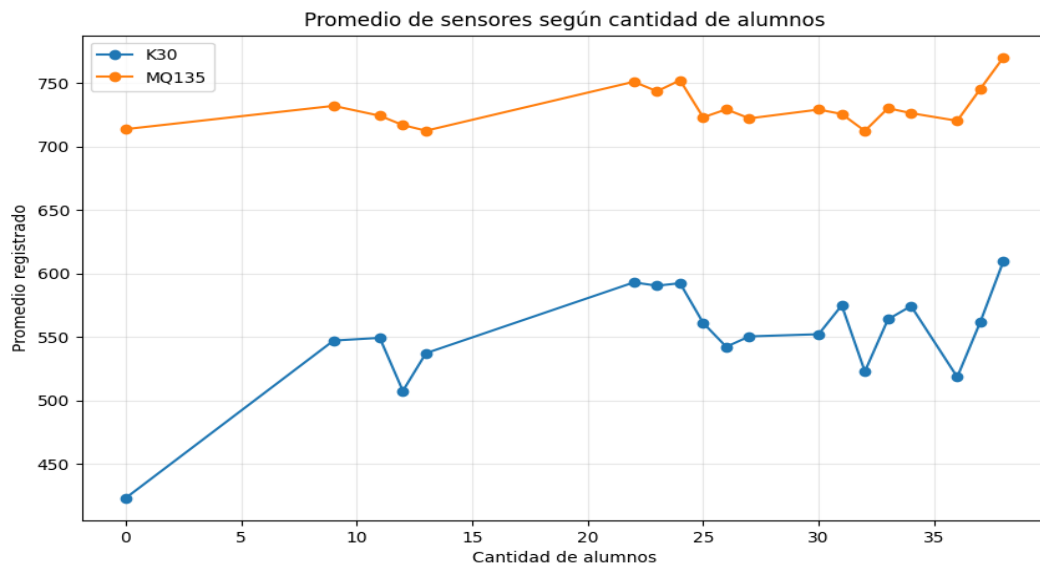
<i>Variable</i>	<i>K30</i>	<i>MQ135</i>	<i>Cantidad de Alumnos</i>
K30	1.000	0.372	0.567
MQ135	0.372	1.000	0.152
Cantidad de Alumnos	0.567	0.152	1.000

Fuente: Elaboración propia

Se observa que el sensor K30 presentó una correlación moderada positiva con la cantidad de alumnos (0.567), indicando que las

concentraciones de CO₂ aumentan conforme incrementa la ocupación del ambiente.

Por otro lado, el sensor MQ135 mostró una correlación menor respecto a la cantidad de estudiantes (0.152), evidenciando menor sensibilidad frente a cambios específicos relacionados con dióxido de carbono.



*Figura N° 13: Sensores y cantidad de alumnos.
Fuente: Propia*

La figura 13 muestra el comportamiento promedio de los sensores K30 y MQ135 según la cantidad de alumnos presentes en el laboratorio. Se observa que, a medida que aumenta la cantidad de estudiantes, los valores registrados por el sensor K30 también tienden a incrementarse, pasando aproximadamente de 420 ppm hasta valores cercanos a 600 ppm. Esto

evidencia una mayor sensibilidad del sensor frente a los cambios de ocupación del ambiente.

Por otro lado, el sensor MQ135 mantuvo valores promedio más estables, registrando concentraciones principalmente entre 710 y 770 ppm. Aunque también presentó ligeras variaciones conforme aumentó la cantidad de alumnos, sus cambios fueron menos notorios en comparación con el sensor K30. Se evidencia que el sensor MQ135 presenta menor variabilidad frente a cambios de ocupación del laboratorio

4.2.3 Resultados para el indicador estabilidad

Con la finalidad de evaluar la estabilidad de los sensores, se calcularon indicadores estadísticos relacionados con la dispersión y variabilidad de los datos registrados durante el periodo de monitoreo, considerando las fluctuaciones, picos y cambios bruscos presentes en las lecturas de los sensores K30 y MQ135 a lo largo del tiempo.

Tabla N° 5:
Indicadores de estabilidad de los sensores

<i>Sensor</i>	<i>Desviación estándar</i>	<i>Varianza</i>	<i>Coficiente de Varianza</i>
K30	115.12	13254.23	0.239
MQ135	63.66	4053.59	0.088

Fuente: Elaboración propia

La Tabla 5 presenta los indicadores estadísticos utilizados para evaluar la estabilidad de los sensores. Se observa que el sensor MQ135 presentó una menor desviación estándar de 63.66, menor varianza de 4053.59 y menor coeficiente de variación de 0.088 en comparación con el sensor K30, que registró una desviación estándar de 115.12, una varianza de 13254.23 y un coeficiente de variación de 0.239. Estos resultados muestran que el MQ135 mantuvo un comportamiento más uniforme y estable durante el monitoreo, mientras que el K30 presentó una mayor dispersión en sus lecturas debido a su sensibilidad frente a cambios bruscos en la concentración de dióxido de carbono dentro del laboratorio.

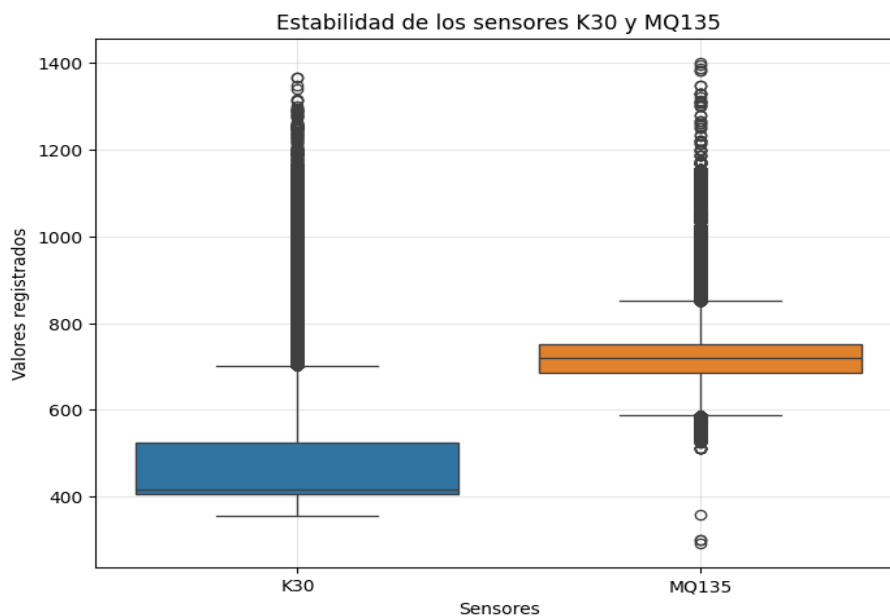
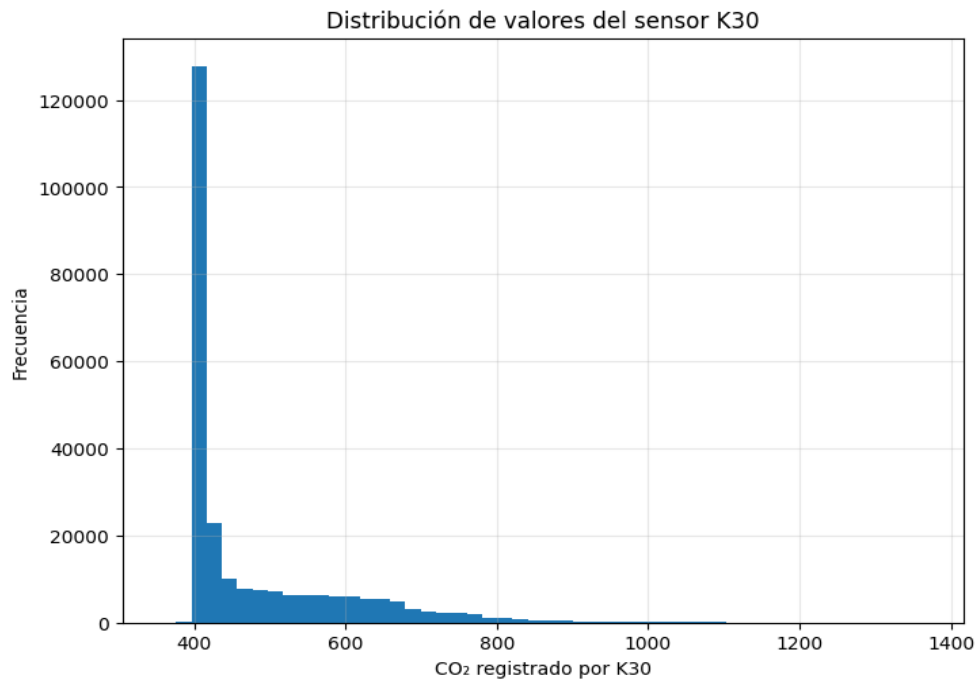


Figura N° 14: Boxplot de estabilidad de sensores.
Fuente: Propia

La figura 14 muestra la estabilidad de los sensores K30 y MQ135 mediante diagramas de caja. Se observa que el sensor K30 presentó valores concentrados principalmente alrededor de los 400 y 500 ppm, aunque con una mayor presencia de valores atípicos que alcanzan aproximadamente los 1400 ppm. Asimismo, el sensor MQ135 registró valores más concentrados entre los 680 y 750 ppm, mostrando una distribución más uniforme durante el monitoreo.

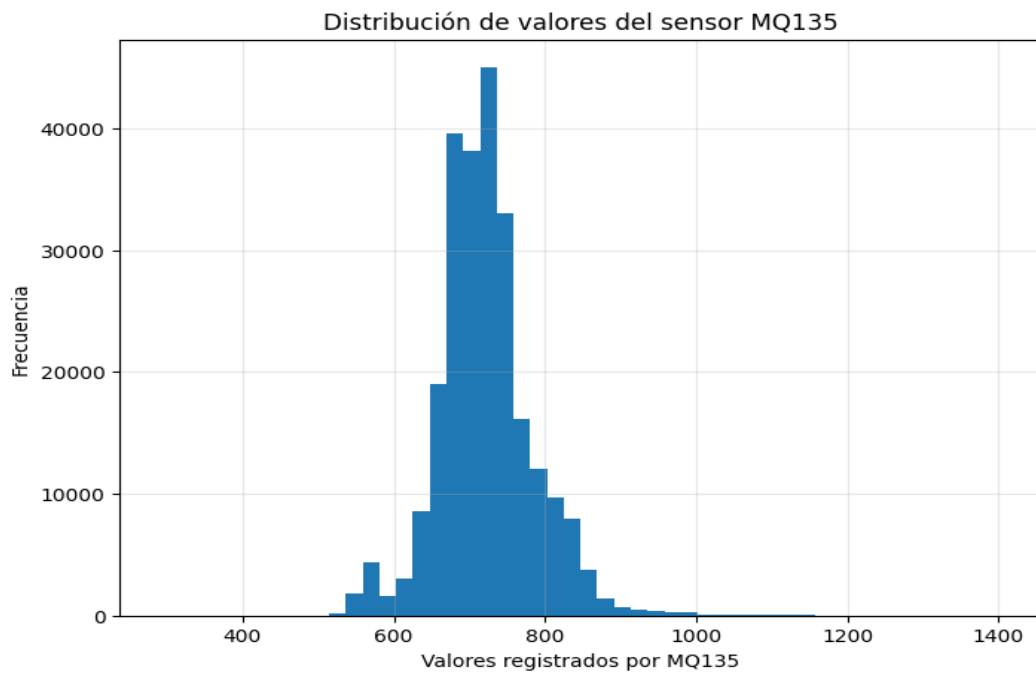
Los resultados coinciden con los indicadores estadísticos obtenidos, donde el MQ135 presentó menor desviación estándar (63.66) y menor coeficiente de variación (0.088) en comparación con el K30, evidenciando un comportamiento más estable durante las pruebas realizadas.

La figura 15 muestra la distribución de valores registrados por el sensor K30. Se observa que la mayor concentración de datos se encuentra alrededor de los 400 ppm, indicando que gran parte de las lecturas se mantuvieron en valores cercanos a ese rango. Asimismo, se presentan registros que alcanzan valores superiores a 800 y 1000 ppm, evidenciando momentos donde existieron incrementos importantes en la concentración de dióxido de carbono dentro del laboratorio. La distribución también muestra una mayor dispersión en comparación con otros rangos de valores, reflejando la sensibilidad del sensor frente a cambios en el ambiente.



*Figura N° 15: Histograma del sensor K30.
Fuente: Propia*

La figura 16 muestra la distribución de valores registrados por el sensor MQ135 durante el periodo de monitoreo. Se observa que la mayor concentración de datos se encuentra principalmente entre los 650 y 780 ppm, registrando su punto más alto de frecuencia alrededor de los 700 ppm. Asimismo, se presentan algunos valores superiores a 900 y 1000 ppm, aunque con menor frecuencia. La distribución evidencia un comportamiento más uniforme y concentrado en comparación con el sensor K30, reflejando una menor dispersión en las lecturas registradas por el MQ135.

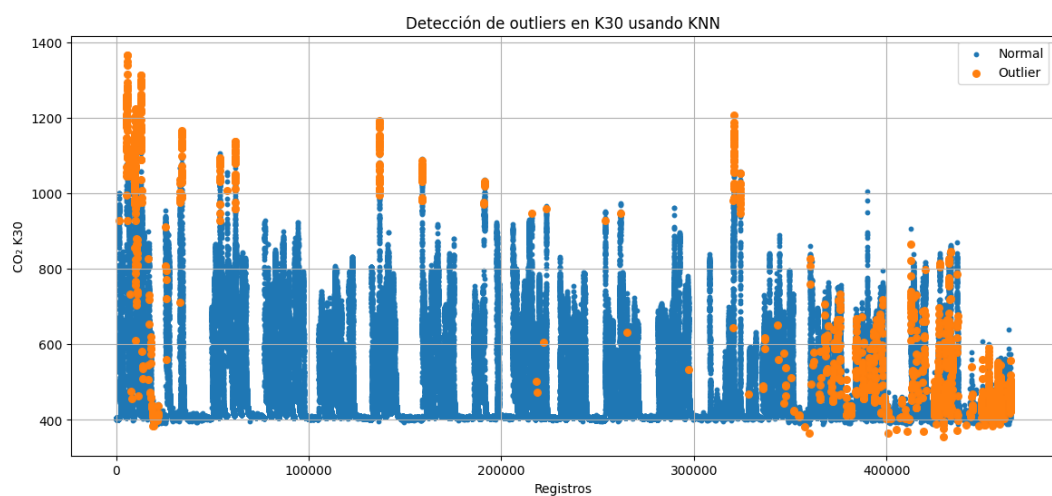


*Figura N° 16: Histograma del sensor MQ135.
Fuente: Propia*

Con el propósito de identificar valores atípicos relacionados con el comportamiento de los sensores, se aplicó la técnica de detección de anomalías mediante el algoritmo K-Nearest Neighbors (KNN).

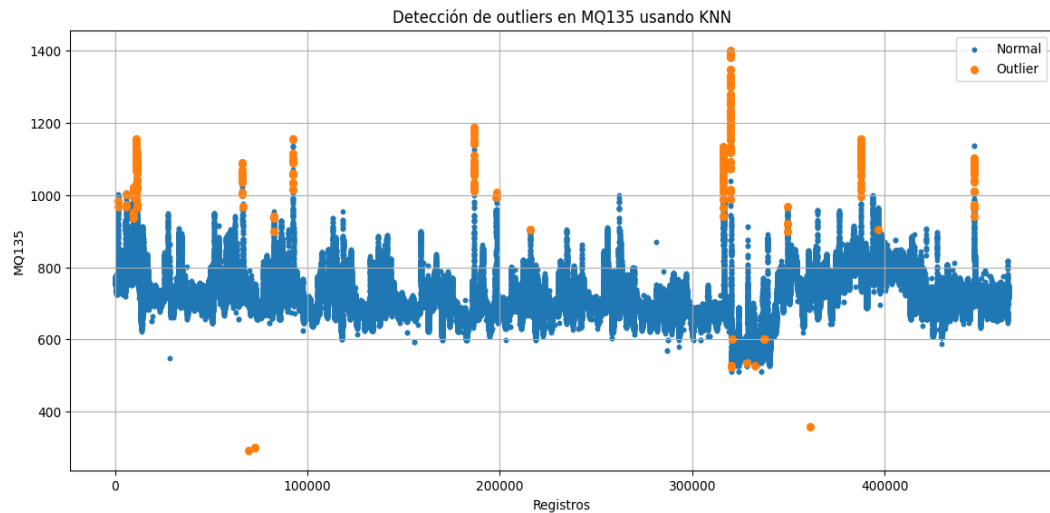
La figura 17 muestra la detección de outliers del sensor K30 utilizando el algoritmo KNN. Se observa que la mayor parte de los valores normales se concentraron alrededor de los 400 y 800 ppm, mientras que los valores identificados como outliers alcanzaron concentraciones superiores a 1000 y 1300 ppm. Asimismo, se presentan algunos valores atípicos por debajo de los 400 ppm durante determinados registros. Estos resultados evidencian que el

sensor K30 presentó variaciones bruscas y cambios importantes en la concentración de dióxido de carbono durante el monitoreo, reflejando una alta sensibilidad frente a cambios en el ambiente del laboratorio.



*Figura N° 17: Detección de outliers del sensor K30 mediante KNN.
Fuente: Propia*

La figura 18 muestra la detección de outliers del sensor MQ135 utilizando el algoritmo KNN. Se observa que la mayor parte de los valores normales se concentraron principalmente entre los 650 y 850 ppm durante el periodo de monitoreo. Asimismo, los valores identificados como outliers alcanzaron concentraciones superiores a 1000 y 1400 ppm, además de algunos valores inferiores a 400 ppm registrados en determinados momentos.



*Figura N° 18: Detección de outliers del sensor MQ135 mediante KNN.
Fuente: Propia*

4.2.4 Resultados para el indicador precisión de la predicción

Con la finalidad de evaluar la precisión de predicción de los modelos utilizados en la investigación, se aplicaron los modelos ARIMAX y LSTM para analizar el comportamiento temporal de las concentraciones de dióxido de carbono registradas por los sensores. Para el desarrollo de los modelos predictivos ARIMAX y LSTM se utilizó una partición temporal de los datos, donde el 80% de los registros fue empleado para el entrenamiento y el 20% para la fase de prueba.

Para evaluar el desempeño predictivo de ambos modelos se usaron las métricas de R^2 (coeficiente de determinación) que mide qué porcentaje de los cambios observados en los datos reales puede ser explicado por el modelo

predictivo, MAE (Mean Absolute Error) que representa el error absoluto promedio entre los valores reales y los valores predichos por el modelo y el RMSE (Root Mean Squared Error) mide la magnitud promedio de los errores de predicción, penalizando principalmente errores grandes dentro de la serie temporal.

4.2.4.1 Resultados del modelo ARIMAX

El modelo ARIMAX fue utilizado para analizar el comportamiento temporal de los sensores K30 y MQ135 considerando una variable exógena relacionada con la ocupación del laboratorio. La incorporación de esta variable permitió considerar el efecto de la ocupación del ambiente sobre la generación de dióxido de carbono, mejorando la capacidad explicativa del modelo.

a) Modelo ARIMAX aplicado al sensor K30

Tabla N° 6:

Métricas del modelo ARIMAX aplicado al sensor K30

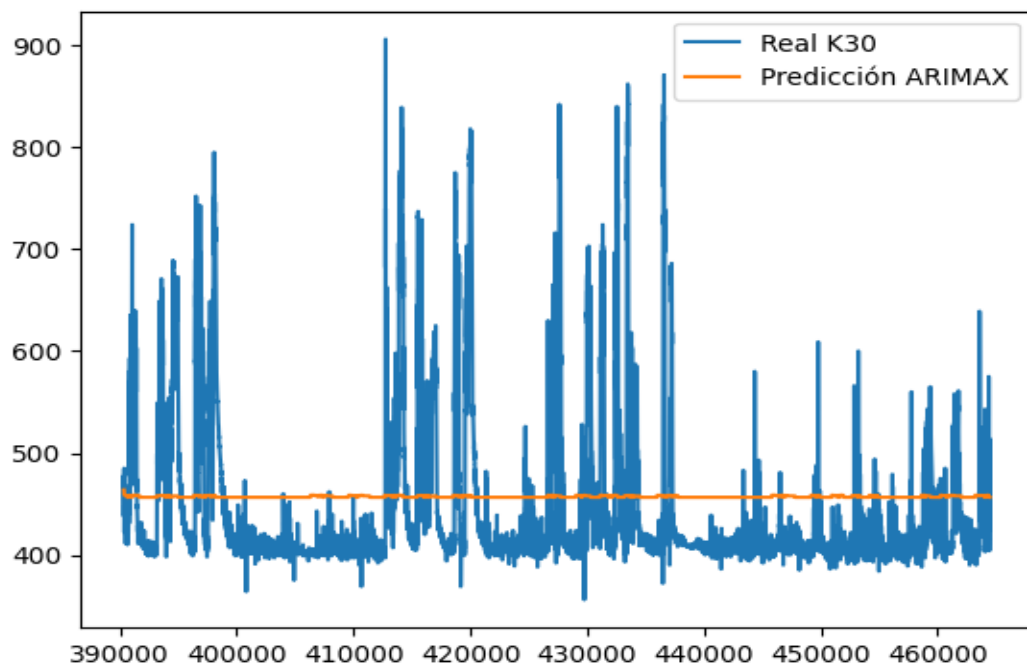
<i>Métrica</i>	<i>Valor</i>
R ²	-0.03
MAE	56.76
RMSE	73.94

Fuente: Elaboración propia

La Tabla 6 presenta las métricas obtenidas del modelo ARIMAX aplicado al sensor K30 para la predicción de concentraciones de CO₂. Se

observa que el modelo alcanzó un coeficiente de determinación R^2 de -0.03, lo que indica que el modelo no logra explicar adecuadamente la variabilidad de los datos y tiene un desempeño inferior a un modelo promedio. Asimismo, el valor MAE de 56.76 y RMSE de 73.94 evidencian errores relativamente altos en las predicciones. En conjunto, estos resultados muestran que el modelo ARIMAX no fue adecuado para predecir el comportamiento del sensor K30, presentando baja capacidad predictiva.

La figura 19 presenta la comparación entre los valores reales registrados por el sensor K30 y los valores predichos mediante el modelo ARIMAX.



*Figura N° 19: Predicción del modelo ARIMAX aplicado al sensor K30.
Fuente: Propia*

Se observa que las predicciones generadas por el modelo se mantienen próximas al valor promedio de la serie temporal, alrededor de 450 ppm. Sin embargo, el modelo presenta dificultades para reproducir las variaciones bruscas y los picos observados en los datos reales, los cuales alcanzan valores superiores a 800 ppm durante determinados periodos. Estos resultados evidencian que, aunque la variable exógena *cantidad de alumnos* aporta información relevante para la predicción, no logra explicar completamente la variabilidad de las concentraciones de CO₂ presentes en el ambiente monitoreado, bajo las condiciones que se establecieron para este estudio.

Los resultados obtenidos permiten concluir que el modelo ARIMAX presentó una baja capacidad predictiva sobre el comportamiento del sensor K30, evidenciada por un coeficiente de determinación negativo y errores elevados en las métricas de evaluación.

b) Modelo ARIMAX aplicado al sensor MQ135

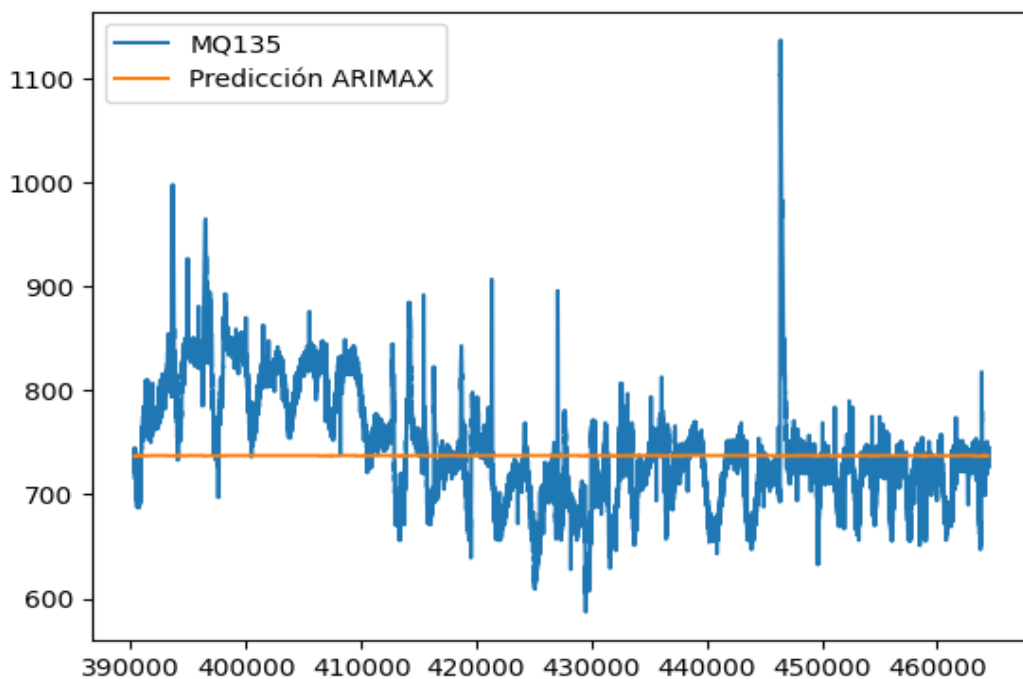
Tabla N° 7:

Métricas del modelo ARIMAX aplicado al sensor MQ135

<i>Métrica</i>	<i>Valor</i>
R ²	-0.004
MAE	40.26
RMSE	53.90

Fuente: Elaboración propia

La Tabla 7 presenta las métricas obtenidas del modelo ARIMAX aplicado al sensor MQ135 para la predicción de concentraciones de gases. Se observa que el modelo obtuvo un coeficiente de determinación R^2 de -0.004, lo que indica una capacidad explicativa baja. Además, los valores de MAE de 40.26 y RMSE de 53.90 muestran la presencia de errores considerables en las predicciones. Estos resultados evidencian que el modelo ARIMAX no logra capturar adecuadamente el comportamiento del sensor MQ135, presentando un bajo rendimiento predictivo.



*Figura N° 20: Predicción del modelo ARIMAX aplicado al sensor MQ135.
Fuente: Propia*

La figura 20 muestra la comparación entre los valores reales registrados por el sensor MQ135 y los valores predichos mediante el modelo ARIMAX.

Se muestra que el modelo ARIMAX no logró seguir de manera precisa el comportamiento del sensor MQ135, presentando diferencias notables entre los valores predichos y los datos reales, incluso en rangos donde las mediciones se mantienen entre 700 y 800 ppm. Asimismo, se observan picos superiores a 1000 ppm y caídas bruscas donde el modelo presenta mayores desviaciones, evidenciando dificultades para capturar adecuadamente las variaciones del sensor.

Se observa que el modelo no logra representar correctamente la tendencia general de los datos, mostrando limitaciones en la predicción de las fluctuaciones del sensor. Esto puede deberse a que el MQ135 posee sensibilidad frente a diferentes gases ambientales, lo que incrementa la complejidad de la serie temporal y dificulta su modelado mediante enfoques tradicionales como ARIMAX.

Se puede indicar que las métricas obtenidas evidencian que el modelo ARIMAX presenta una baja capacidad predictiva para ambos sensores. Si bien el sensor MQ135 obtuvo un valor de R^2 ligeramente superior (-0.004) en comparación con el K30 (-0.03), este resultado sigue indicando un desempeño

inadecuado, ya que el modelo no logra explicar la variabilidad de los datos ni representar correctamente su comportamiento temporal.

4.2.4.2 Resultados del modelo LSTM

El modelo LSTM fue implementado para analizar patrones temporales complejos presentes en las mediciones de los sensores K30 y MQ135. Para ello, los datos fueron normalizados mediante MinMaxScaler y organizados en secuencias temporales de 10 pasos, permitiendo que la red neuronal aprenda el comportamiento histórico de las concentraciones de CO₂. También se usaron las métricas de R², MAE y RMSE.

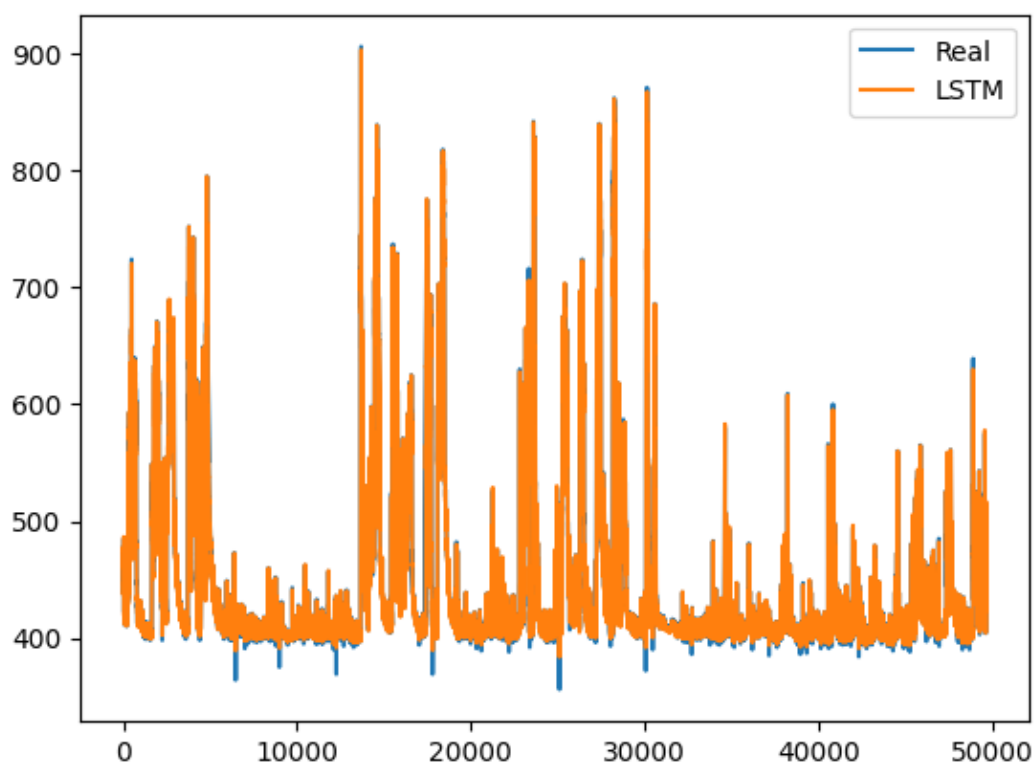
Tabla N° 8:
Métricas del modelo LSTM aplicado a los sensores

<i>Métrica</i>	<i>Valor</i>
R ²	0.98
MAE	4.496
RMSE	8.348

Fuente: Elaboración propia

La Tabla 8 presenta las métricas obtenidas del modelo LSTM aplicado a los datos de los sensores para la predicción de concentraciones de CO₂. Se observa que el modelo alcanzó un coeficiente de determinación R² de 0.98, lo que indica un alto nivel de ajuste entre los valores reales y los valores predichos. Asimismo, el valor de MAE fue de 4.496, evidenciando un bajo

error absoluto promedio en las predicciones realizadas por el modelo. Del mismo modo, el RMSE obtenido fue de 8.348, mostrando que las diferencias entre los valores reales y estimados fueron reducidas. En conjunto, estos resultados demuestran que el modelo LSTM presentó un desempeño eficiente y una alta precisión predictiva en el análisis de los datos obtenidos por los sensores.



*Figura N° 21: Predicción del modelo LSTM aplicado al gráfico real.
Fuente: Propia*

La figura 21 muestra que el modelo LSTM logró seguir el comportamiento general de los datos reales del sensor, manteniendo la mayor

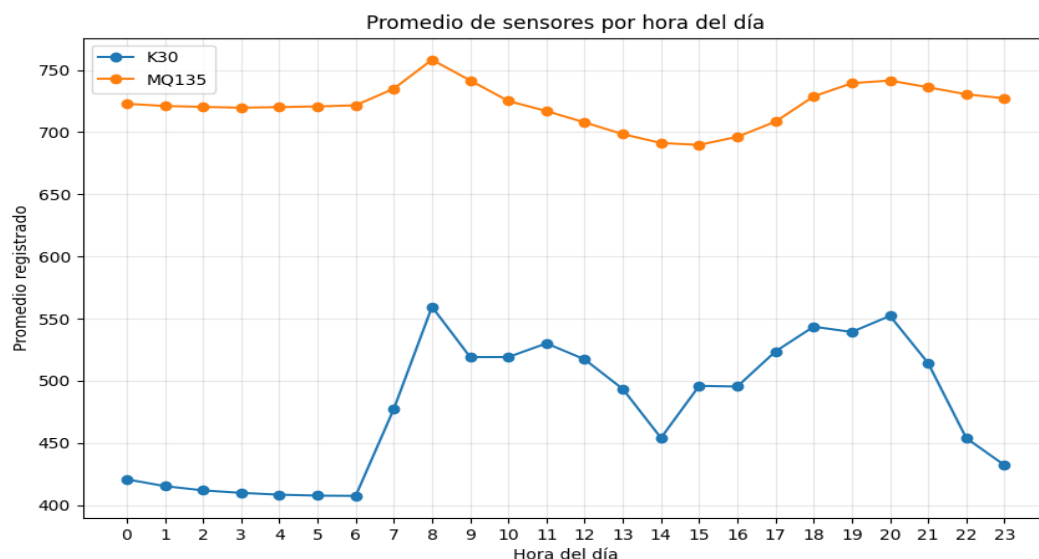
parte de las predicciones alrededor de los 400 ppm. Asimismo, se observan múltiples picos que superan los 800 y 900 ppm, indicando momentos donde existieron cambios bruscos en la concentración de dióxido de carbono. Estos resultados muestran que el modelo LSTM fue capaz de identificar patrones temporales y variaciones importantes dentro del comportamiento del sensor durante el monitoreo.

Los resultados muestran que la red neuronal logró aprender adecuadamente el comportamiento histórico de las concentraciones de CO₂, generando predicciones cercanas a los valores reales observados.

Además, el modelo permitió identificar tendencias y fluctuaciones relacionadas con horarios académicos y cantidad de estudiantes presentes en el laboratorio, demostrando la utilidad de las redes neuronales para el análisis predictivo de calidad del aire.

4.2.5 Resultados complementarios

La figura 22 muestra el comportamiento promedio de los sensores durante las diferentes horas del día.



*Figura N° 22: Promedio de sensores por hora del día.
Fuente: Propia*

Se observa que el sensor K30 presentó sus valores más bajos durante la madrugada, registrando aproximadamente entre 400 y 420 ppm entre las 00:00 y 06:00 horas. Posteriormente, los valores aumentaron desde las 07:00 horas, alcanzando picos cercanos a 560 ppm alrededor de las 08:00 y 20:00 horas, coincidiendo con horarios de mayor actividad dentro del laboratorio.

Por otro lado, el sensor MQ135 mantuvo valores promedio más estables durante el día, registrando concentraciones principalmente entre 690 y 760 ppm. Asimismo, sus valores más altos se observaron alrededor de las 08:00 horas y durante la noche entre las 18:00 y 21:00 horas, evidenciando variaciones relacionadas con los cambios de actividad y ocupación del ambiente.

Se observa que las mayores concentraciones de CO₂ se registraron durante horarios de actividad académica, especialmente en horas donde existe mayor ocupación del laboratorio.

La figura 23 muestra el promedio de valores registrados por los sensores K30 y MQ135 según el día de la semana. Se observa que el sensor K30 presentó sus valores más altos durante el miércoles, alcanzando aproximadamente 530 ppm, mientras que sus valores más bajos se registraron durante el sábado y domingo, con concentraciones cercanas a 410 ppm. Esto evidencia una disminución de dióxido de carbono durante los días con menor actividad académica.

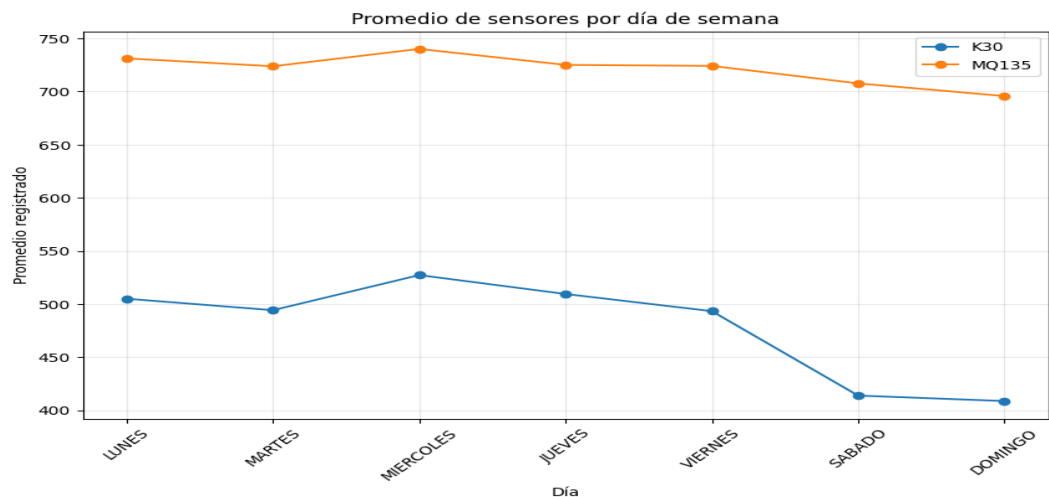


Figura N° 23: Promedio de sensores por día de la semana.
Fuente: Propia

Por otro lado, el sensor MQ135 mantuvo valores promedio más estables durante toda la semana, registrando concentraciones entre 695 y 740 ppm. Asimismo, los valores más altos se observaron durante el miércoles, mientras que los valores más bajos se registraron el domingo, coincidiendo con una menor presencia de estudiantes dentro del laboratorio.

4.3 Contrastación de hipótesis

Para la contrastación de hipótesis se compararon las mediciones obtenidas por los sensores K30 y MQ135 mediante análisis estadístico descriptivo y correlacional.

Los resultados obtenidos muestran diferencias en el comportamiento de ambos sensores durante el monitoreo de calidad del aire dentro del laboratorio.

El sensor K30 presentó mayor sensibilidad y una mejor respuesta para detectar cambios en las concentraciones de CO₂, mientras que el sensor MQ135 destacó por mantener un comportamiento más estable durante el periodo de estudio.

Por lo tanto, se acepta la hipótesis general de investigación, concluyendo que existe diferencia en la performance de los sensores K30 y MQ135 utilizando técnicas de analítica de datos.

CAPÍTULO V

DISCUSIÓN

5.1 Pruebas de validación del modelo experimental

Los resultados obtenidos durante la investigación permitieron validar el comportamiento de los sensores K30 y MQ135 en el monitoreo de calidad de aire dentro del laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas.

El sensor K30 presentó un mejor desempeño en relación con la detección de variaciones de dióxido de carbono, especialmente en horarios donde existía una mayor cantidad de alumnos dentro del ambiente. Esto puede deberse a que el sensor utiliza tecnología NDIR, especializada en la medición de CO₂, permitiendo obtener mediciones más específicas y sensibles frente a cambios en la concentración del gas.

Por otro lado, el sensor MQ135 presentó un comportamiento más estable durante largos periodos de monitoreo. Sin embargo, su sensibilidad específica al dióxido de carbono fue menor debido a que este sensor también responde a otros gases presentes en el ambiente.

Asimismo, los modelos predictivos ARIMAX y LSTM permitieron analizar el comportamiento temporal de los sensores; sin embargo, los resultados obtenidos mediante las métricas R^2 , MAE y RMSE evidenciaron diferencias observables en su desempeño. El modelo ARIMAX presentó valores negativos de R^2 y mayores errores, lo que indica una baja capacidad para representar el comportamiento de los datos. Por otro lado, el modelo LSTM demostró un alto nivel de ajuste, con valores elevados de R^2 y menores errores, logrando una mejor capacidad de representación y predicción de los datos recolectados durante el periodo de estudio.

5.2 Aplicación de la tecnología encontrada

La tecnología implementada en esta investigación puede ser utilizada para el monitoreo de la calidad del aire en aulas universitarias y otros entornos educativos, permitiendo relacionar las concentraciones de CO_2 con variables como fatiga, estrés, nivel de concentración y rendimiento académico de los estudiantes. De esta manera, el análisis de los datos obtenidos contribuiría a comprender cómo las condiciones ambientales influyen en el bienestar y desempeño dentro de espacios cerrados. El uso de sensores conectados mediante tecnologías IoT permite obtener información en tiempo real sobre los niveles de dióxido de carbono presentes en un ambiente.

También, la integración de sensores con plataformas en la nube y técnicas de analítica de datos facilita la implementación de sistemas

inteligentes de monitoreo ambiental en tiempo real, permitiendo identificar ambientes con poca ventilación y contribuir a la toma de decisiones relacionadas con la salud y comodidad de los estudiantes.

Además, el uso de sensores de bajo costo como el MQ135 representa una alternativa accesible para instituciones educativas que no cuentan con un alto presupuesto para implementar sistemas especializados de monitoreo ambiental.

5.3 Contraste con trabajos de investigación similares

Los resultados obtenidos en la presente investigación coinciden con lo mencionado por Spinelle et al. (2015), quienes señalan que los sensores basados en tecnología NDIR, como el K30, presentan un mayor rendimiento de captura y sensibilidad en la medición de dióxido de carbono en comparación con sensores de bajo costo como el MQ135. Durante el desarrollo de esta investigación se observó que el sensor K30 logró detectar de manera más eficiente las variaciones de CO₂ relacionadas con la cantidad de alumnos presentes en el laboratorio.

Asimismo, los resultados también guardan relación con lo mencionado por Hasenfratz et al. (2012) y Kumar et al. (2025), quienes indican que sensores como el MQ135 son capaces de detectar variaciones en la concentración de contaminantes y pueden presentar limitaciones debido a que

reaccionan frente a diferentes tipos de gases presentes en el ambiente, lo cual afecta la especificidad en la medición de dióxido de carbono. Esto coincide con los resultados obtenidos, donde el MQ135 presentó lecturas más variables y una menor correlación con la cantidad de estudiantes dentro del laboratorio.

De igual forma, la aplicación de modelos predictivos ARIMAX y LSTM presentó resultados comparables con la investigación de Tyagi et al. (2026), al comparar modelos ARIMA y LSTM para la predicción de la calidad del aire, se evidenció que el modelo LSTM alcanzó un coeficiente de determinación (R^2) cercano a 0.96, superando significativamente al modelo ARIMA debido a su capacidad para capturar patrones no lineales y dinámicos en los datos. En este sentido, los resultados de la presente investigación coinciden con dichos hallazgos, ya que el modelo LSTM demostró un mejor desempeño en la predicción del comportamiento temporal de los sensores de CO₂, mientras que el modelo ARIMAX presentó limitaciones en su capacidad predictiva, evidenciadas en valores negativos del coeficiente de determinación (R^2) y mayores errores de predicción.

A diferencia de otros trabajos relacionados con monitoreo ambiental, la presente investigación incorporó la relación entre la cantidad de alumnos y las variaciones de dióxido de carbono dentro de un laboratorio universitario, permitiendo analizar cómo la ocupación del ambiente influye directamente en el comportamiento de los sensores K30 y MQ135.

CONCLUSIONES

Primera conclusión

Se concluyó que el sensor K30 presentó una mejor performance en comparación con el sensor MQ135 para el monitoreo de calidad de aire dentro del laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas en la Universidad Nacional Jorge Basadre Grohmann. Esto se evidenció debido a su mayor capacidad para detectar variaciones de CO₂ relacionadas con la presencia de alumnos dentro del laboratorio.

Segunda conclusión

Con respecto al indicador de rendimiento de captura de datos, los resultados obtenidos permitieron observar que el sensor K30 mostró una mayor capacidad para detectar variaciones en la concentración de CO₂ durante los diferentes horarios. Asimismo, el sensor MQ135 presentó lecturas más variables debido a que este tipo de sensor también responde a otros gases presentes en el ambiente.

Tercera conclusión

Respecto al indicador sensibilidad, el sensor K30 presentó una mayor relación con la cantidad de alumnos presentes en el laboratorio, evidenciando una mejor capacidad para detectar variaciones de CO₂ dentro del ambiente.

Esto se observó principalmente en los horarios con mayor ocupación estudiantil, donde los valores registrados por el sensor aumentaban de manera más evidente.

Cuarta conclusión

En cuanto al indicador estabilidad, el sensor MQ135 mostró un comportamiento más constante durante largos periodos de monitoreo continuo. Sin embargo, a pesar de su estabilidad, presentó menor sensibilidad específica al CO₂ en comparación con el K30.

Quinta conclusión

En relación con el indicador precisión de la predicción, correspondiente a la variable interviniente analítica de datos, los resultados obtenidos mediante las métricas R², MAE y RMSE evidenciaron diferencias en el desempeño de los modelos predictivos ARIMAX y LSTM. Aunque ambos modelos permitieron analizar el comportamiento temporal de las concentraciones de CO₂, el modelo LSTM mostró una capacidad predictiva significativamente superior al modelo ARIMAX, constituyéndose como la alternativa más adecuada para la predicción de datos obtenidos mediante sensores de calidad de aire.

Sexta conclusión

La incorporación de técnicas de analítica de datos permitió transformar los datos capturados por los sensores en información útil para el análisis y la predicción del comportamiento de la calidad del aire dentro del laboratorio, demostrando el potencial de estas herramientas para apoyar la toma de decisiones en entornos universitarios.

Séptima conclusión

La integración de sensores de calidad de aire, tecnologías IoT y técnicas de analítica de datos permitió desarrollar un sistema de monitoreo capaz de analizar el comportamiento ambiental dentro de un laboratorio universitario. Asimismo, la investigación permitió demostrar la importancia del monitoreo continuo de CO₂ en ambientes cerrados para contribuir a mejores condiciones de ventilación y bienestar en la universidad.

RECOMENDACIONES

Primera recomendación

Se sugiere implementar sistemas de monitoreo de calidad de aire en aulas y laboratorios universitarios con la finalidad de controlar los niveles de CO₂ y mejorar las condiciones de ventilación dentro de ambientes cerrados.

Segunda recomendación

Es recomendable utilizar sensores K30 en investigaciones donde se requiera un mayor rendimiento con respecto a la captura de datos del CO₂, especialmente en espacios donde exista alta concentración de personas.

Tercera recomendación

Para futuras investigaciones se recomienda que incorporen otros factores ambientales como temperatura, humedad y ventilación, con el propósito de realizar análisis más completos y tener más seguridad con respecto a la sensibilidad de los datos del CO₂ en espacios universitarios.

Cuarta recomendación

Se recomienda ampliar el periodo de recolección de datos y realizar pruebas en diferentes ambientes, tales como que ingrese una cantidad de estudiantes con actividad física y otra cantidad sin actividad física, para

comparar el comportamiento y estabilidad de los sensores en distintos escenarios académicos.

Quinta recomendación

Se propone ampliar las variables empleadas en los modelos predictivos mediante la incorporación de factores ambientales como temperatura, humedad, ventilación y ocupación del ambiente, con el propósito de mejorar el desempeño predictivo de los modelos de CO₂. Del mismo modo, se recomienda profundizar la investigación en modelos basados en redes neuronales recurrentes, especialmente LSTM, dado que demostraron una mejor capacidad para representar el comportamiento temporal de los datos monitoreados.

Sexta recomendación

Para un mejor análisis de datos, se recomienda desarrollar futuros sistemas integrados con dashboards o plataformas de visualización en tiempo real, permitiendo que autoridades y docentes puedan monitorear de manera rápida el estado de la calidad de aire dentro de las aulas.

Séptima recomendación

Para una captura de datos real, se debe realizar la calibración de los sensores antes de iniciar el proceso de recolección de datos, debido a que los

valores obtenidos por defecto pueden no representar mediciones reales del ambiente. Asimismo, se recomienda mantener los sensores en funcionamiento durante un periodo previo de estabilización, aproximadamente una semana, con la finalidad de mejorar la estabilidad y confiabilidad de las lecturas obtenidas durante el monitoreo.

REFERENCIAS BIBLIOGRÁFICAS

- Aggarwal, S., & Kumar, N. (2020). A review on the applications of ESP32 microcontroller. *International Journal of Engineering Research & Technology*, 9(6), 478–482.
- Alhmiedat, T., & Samara, G. (2017). A low cost ZigBee sensor network architecture for indoor air quality monitoring. arXiv. <https://arxiv.org/abs/1712.04190>
- Allen, J. G., MacNaughton, P., Satish, U., Santanam, S., Vallarino, J., & Spengler, J. D. (2016). Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: A controlled exposure study of green and conventional office environments. *Environmental Health Perspectives*, 124(6), 805–812. <https://doi.org/10.1289/ehp.1510037>
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50–58. <https://doi.org/10.1145/1721654.1721672>
- Belimo. (2023). La importancia de monitorizar los niveles de CO₂ en interiores en las escuelas. <https://www.belimo.com>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

- Brownlee, J. (2017). Introduction to Time Series Forecasting with Python. Machine Learning Mastery. <https://machinelearningmastery.com/convert-time-series-supervised-learning-problem-python/>
- Cai, J., et al. (2024). Evaluation of low-cost CO₂ sensors using reference instruments and standard gases for indoor use. Sensors, 24. <https://doi.org/10.3390/s24092680>
- Davenport, T. H., & Harris, J. G. (2017). Competing on analytics: The new science of winning. Harvard Business Review Press.
- Fraden, J. (2016). Handbook of Modern Sensors: Physics, Designs and Applications (5th ed.). Springer.
- Fuentes-Ferragud, E., López, A., Piera, J. M., Yusà, V., Garrigues, S., de la Guardia, M., López Labrador, F. X., Camaró, M., Ibáñez, M., & Coscollà, C. (2024). Indoor air quality and bioaerosols in Spanish university classrooms. Toxics, 12(3), 227. <https://doi.org/10.3390/toxics12030227>
- García Navarrete, G., & Rico Soto, K. G. (2019). Sensores de bajo costo para el monitoreo de calidad del aire. Epistemus, 13(27), 30–37. <https://www.redalyc.org/pdf/7262/726281459005.pdf>
- Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.

Gil García, E. D., & Alemán Ramos, P. F. (2019). Estudio de las fuentes relevantes en la investigación educativa sobre ocio digital. Universidad de Las Palmas de Gran Canaria.

https://accedacris.ulpgc.es/bitstream/10553/119900/1/estudio_investigacion_sobre.pdf

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660. <https://doi.org/10.1016/j.future.2013.01.010>

Gutiérrez, M., Romero, C., & Vásquez, J. (2021). Evaluación de sensores de bajo costo para monitoreo de calidad de aire en ambientes educativos. *Revista Colombiana de Computación*, 22(1), 33–45.

<https://repositorio.unal.edu.co/handle/unal/77273>

Hasenfratz, D., Saukh, O., Sturzenegger, S., & Thiele, L. (2012). Participatory air pollution monitoring using smartphones. *Mobile Sensing*, 1–5.

Hernandez Arteaga, R., Noemi Tellez Vargas, L., Ramírez Zamora, J. D., & Domínguez Ramírez, O. A. (2025). Diseño e implementación de una estación de monitoreo ambiental basada en ESP32 para laboratorios académicos. *Ibero Ciencias - Revista Científica y Académica*, 4(4), 418–446. <https://doi.org/10.63371/ic.v4.n4.a389>

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (6.ª ed.). McGraw-Hill Education.

- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Huang, J., Lin, Z., & Zhang, Y. (2019). Evaluation of low-cost air quality sensors: Performance and applications. *Journal of Sensors*, 2019, 1–10. <https://doi.org/10.1155/2019/9219628>
- Kumar, M., Mishra, G., Sharma, A., Shaini, A., & Saxena, S. (2025). Air quality monitoring using MQ135 gas sensor and Arduino Uno. *International Journal of Latest Technology in Engineering, Management & Applied Science*. <https://doi.org/10.51583/IJLTEMAS.2025.140500119>
- Li, J., Wang, X., & Zhang, Y. (2020). A comparative study on the performance of low-cost CO₂ sensors. *Sensors*, 20(21), 6258. <https://doi.org/10.3390/s20216258>
- Luna, A., Talavera, Á., & Cano, L. (2017). Uso de sensores electroquímicos de bajo costo para el monitoreo de la calidad del aire en el distrito de San Isidro, Lima, Perú. Centro de Investigación de la Universidad del Pacífico.
- Martin, C. R., Zeng, N., Karion, A., Dickerson, R., Ren, X., Turpie, B., & Weber, K. (2017). Evaluation and environmental correction of ambient CO₂ measurements from a low-cost NDIR sensor. *Atmospheric Measurement Techniques*. <https://doi.org/10.5194/amt-10-2383-2017>

MathWorks. (2022). ThingSpeak Documentation.
<https://www.mathworks.com/help/thingspeak/>

Mendoza Merma, R. P., & Cordero Ochoa, V. P. A. (2022). Diseño de un sistema IoT para el monitoreo remoto de la calidad del aire en ambientes interiores de la Facultad de Ingeniería de la Universidad Privada de Tacna (Tesis de pregrado). Universidad Privada de Tacna.

Moisés Ríos, B. W., Ango Bedriñana, J. H., Palomino Vargas, V. A., & Feria Macizo, E. (2019). Diseño del proyecto de investigación científica (2.^a ed.). Editorial San Marcos.

Moreno, J. A. B., & Escobar, J. R. E. (2025). Calidad del aire, impacto en el rendimiento laboral y funciones cognitivas. Revista Veritas.
<https://revistaveritas.org/index.php/veritas/article/view/1052>

Organización Mundial de la Salud. (2021). Guidelines on indoor air quality: selected pollutants. OMS.

Organización Mundial de la Salud. (2021). Ambient (outdoor) air pollution.
[https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)

- Pérez García, W. M. (2024). Sistema de monitoreo de polución de aire en los predios de la Universidad Técnica Estatal de Quevedo (Tesis de grado). Universidad Técnica Estatal de Quevedo, Ecuador. <https://repositorio.uteq.edu.ec/handle/43000/7243>
- Rochmania, A., Sucahyo, I., & Yantidewi, M. (2021). Monitoring kandungan CO₂ berbasis IoT dengan NodeMCU ESP8266 dan sensor MQ135. Jurnal Sains dan Pendidikan Fisika. <https://doi.org/10.35580/jspf.v17i3.30634>
- Shrestha, R., Maharjan, M., & Sharma, M. (2022). Air pollution monitoring system using microcontroller ATmega32A and MQ135 gas sensor. Semiconductor Science and Information Devices. <https://doi.org/10.30564/ssid.v4i2.4884>
- Spinelle, L., Gerboles, M., Kok, G., Persijn, S., & Sauerwald, T. (2015). Review of portable and low-cost sensors for the ambient air monitoring of benzene and other volatile organic compounds. Sensors, 17(7), 1520. <https://doi.org/10.3390/s17071520>
- Sundell, J. (2004). On the history of indoor air quality and health. Indoor Air, 14(s7), 51–58. <https://doi.org/10.1111/j.1600-0668.2004.00273.x>
- Tyagi, P., Sharma, A., & Kumar, R. (2026). Comparative analysis of ARIMA and LSTM architectures for urban air quality prediction: A case study of Delhi-NCR. International Journal of Computational Engineering and Data Science.

Anexos

Anexo 1

Matriz de consistencia

ANALÍTICA DE DATOS Y PERFORMANCE DE SENSORES DE CALIDAD DE AIRE EN EL LABORATORIO DE LA ESCUELA PROFESIONAL DE INGENIERÍA EN INFORMÁTICA Y SISTEMAS UNJBG, 2025

Problemas	Objetivos	Hipótesis	Variables	Dimensiones	Indicadores	Diseño de la investigación
<u>Problema principal</u> ¿Cuál es la performance de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG utilizando analítica de datos?	<u>Objetivo general</u> Comparar la performance de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.	<u>Hipótesis general</u> Existe diferencia en la performance de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.	Variable principal: Performance de sensores de calidad de aire	Rendimiento	Rendimiento de captura de datos	<u>Población:</u> Datos obtenidos de los Sensores K30 y MQ135. <u>Muestra</u> Los registros de datos válidos de los sensores K30 y MQ135, obtenidos mediante muestreo no probabilístico. <u>Tipo de investigación</u> Aplicada. <u>Diseño de investigación</u> No experimental, longitudinal y comparativo <u>Instrumentos</u> Sensor K30 y Sensor MQ135
<u>Problema específico 1</u> ¿Cuál es el rendimiento de captura de datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG utilizando analítica de datos?	<u>Objetivo específico 1</u> Determinar el rendimiento de captura de datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas de la UNJBG usando analítica de datos.	<u>Hipótesis específica 1</u> Existe diferencia en el rendimiento de captura de datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.		Estabilidad	Desviación estándar, varianza y coeficiente de variación	
<u>Problema específico 2</u> ¿Cuál es la sensibilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas utilizando analítica de datos?	<u>Objetivo específico 2</u> Determinar la sensibilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas de la UNJBG usando analítica de datos.	<u>Hipótesis específica 2</u> Existe diferencia en la sensibilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.		Sensibilidad	Correlación entre CO ₂ y cantidad de alumnos	
<u>Problema específico 3</u> ¿Cuál es la estabilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas utilizando analítica de datos?	<u>Objetivo específico 3</u> Determinar la estabilidad de los sensores de MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas de la UNJBG usando analítica de datos.	<u>Hipótesis específica 3</u> Existe diferencia en la estabilidad de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.	Variable interviniente: Analítica de datos	Modelado predictivo	Precisión de la predicción (MAE, RMSE y R ²)	
<u>Problema específico 4</u> ¿Cuál es la precisión de la predicción de los niveles de CO ₂ con los datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG utilizando analítica de datos?	<u>Objetivo específico 4</u> Determinar la precisión de la predicción de los niveles de CO ₂ con los datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas de la UNJBG usando analítica de datos.	<u>Hipótesis específica 4</u> Existe diferencia en la precisión de la predicción de los niveles de CO ₂ con los datos de los sensores MQ135 y K30 en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG usando analítica de datos.				

Anexo 2

Autorización de la Directora de la Escuela Profesional de Ingeniería en Informática y Sistemas



Universidad Nacional Jorge Basadre Grohmann
Facultad de Ingeniería
Escuela Profesional de Ingeniería en Informática y Sistemas



"Año de la recuperación y consolidación de la economía peruana"

Tacna, 28 de abril del 2025

Señorita

Bach. Ruth de Jesus Ramirez Rojas

Egresada de la ESIS

Presente.-

Asunto: Autorización para instalaciones de datos y energía para la aplicación de trabajo de investigación para título profesional

Es grato dirigirme a usted para expresarle un cordial saludo y, en atención a su solicitud, comunicarle que esta Dirección autoriza la aplicación del trabajo de investigación titulado *"Análítica de datos y performance de sensores de calidad de aire en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG, 2025"* desarrollado en el marco del Proyecto de Investigación "Determinación del impacto de la minería de datos, social media, internet de las cosas (IoT) y realidad aumentada en la Educación Superior Universitaria" aprobado con R.R. N° 13626-2024-UNJBG, para la obtención del título profesional.

La presente autorización se emite considerando que los resultados obtenidos contribuirán al fortalecimiento académico de la comunidad universitaria.

Sin otro particular de momento, le reitero las muestras de mi consideración y estima.

Atentamente,




Dra. Karin Yanet Supo Gavancho
Directora(e) de la Escuela Profesional
de Ingeniería en Informática y Sistemas

c.c. 
KYSG

Anexo 3

Correo de respuesta con la información de cantidad de estudiantes

EN RESPUESTA AL OFICIO N° 0053-2026-DIESU-EATA-FAIN-UNJBG - ramirezr@unjbg.edu.pe - Correo de Universidad Nacion...

 Buscar en todas las conversaciones

99+ Redactar

Mail 1 Recibidos 1.014

Chat Destacados

Meet Pospuestos

Enviados

Borradores 1

Categorías

Más

Etiquetas

EN RESPUESTA AL OFICIO N° 0053-2026-DIESU-EATA-FAIN-UNJBG

 **Nelson Abrahán Pablo Mollo Condori** <n.mollo@unjbg.edu.pe>
para proyecto_jot, Alicia, Dirección, mí

 Parece que este mensaje está en inglés  [Traducir al español](#)

Estimada Alicia se remite la información de lo solicitado por el proyecto de acuerdo lo coordinado con la parte de usua
Adjunto en formato XLS

—

Ing. Nelson Mollo Condori
UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN
DASA - URC
Movistar: +51-52-990071060
Rpm #990071060
Claro: +51-52-952356821

Un archivo adjunto • Analizados por Gmail  Añadir a Drive



Responder Responder a todos Reenviar  Nuevo
Compartir en chat

Anexo 4
Documento de solicitud de información de población estudiantil

 PROYECTO DE INVESTIGACIÓN "DETERMINACIÓN DEL IMPACTO DE LA MINERÍA DE DATOS, SOCIAL MEDIA, INTERNET DE LAS COSAS (IoT) Y REALIDAD AUMENTADA EN LA EDUCACIÓN SUPERIOR UNIVERSITARIA" <i>"Año de la Esperanza y el Fortalecimiento de la Democracia"</i>	
OFICIO N°0053-2026-DIESU-EATA-FAIN-UNJBG Tacna, 14 de mayo del 2026	2626920 UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN DIRECCIÓN DE SERVICIOS Y PROCESOS ACADÉMICOS RECIBIDO 15 MAY 2026 Reg. N° Hora 9:00 Firma
SEÑOR: DR. LUIS ANDRÉS AMAYA CEDRÓN DIRECCIÓN DE SERVICIOS Y PROCESOS ACADÉMICOS Presente. -	
ASUNTO: SOLICITUD DE INFORMACIÓN DE POBLACIÓN ACADÉMICA ESTUDIANTIL <i>Proyecto de Investigación: "Determinación del Impacto de la Minería de Datos, Social Media, Internet de las Cosas (IoT) y Realidad Aumentada en la Educación Superior Universitaria"</i>	
REF: 1. RESOLUCIÓN DE FACULTAD N°10427-2026-FAIN/UNJBG 2. RESOLUCIÓN DE FACULTAD N°9542-2025-FAIN/UNJBG 3. RESOLUCIÓN RECTORAL N°13626-2024-UNJBG	

Es grato dirigirme a usted, para saludarlo cordialmente y por medio del presente solicitar a su despacho información respecto a la población académica estudiantil de la Universidad Nacional Jorge Basadre Grohmann, según el siguiente detalle:

- Total de estudiantes matriculados por grupos (G1, G2 y G3) del periodo 2025-II, según cada curso realizado en el laboratorio de Computación Aplicada de la Escuela Profesional de Ingeniería en Informática y Sistemas.
- Código de los estudiantes y sexo de los matriculados de cada curso y grupo en el periodo 2025-II.
- Notas del conocimiento, notas de desempeño y notas de producto de cada curso y grupo en el periodo 2025-II.

Elio con la finalidad de proporcionar información al proyecto de tesis presentado por el Bach. Ruth de Jesús Ramírez Rejas proyecto denominado: "Analítica de datos y performance de sensores de calidad de aire en el laboratorio de la Escuela Profesional de Ingeniería en Informática y Sistemas UNJBG, 2025", tesista que realizará la ejecución de su proyecto de tesis en el marco de nuestro proyecto de investigación.

Sin otro particular, agradezco su atención al presente, me suscribo de usted.

Atentamente,


"DETERMINACIÓN DEL IMPACTO DE LA MINERÍA DE DATOS, SOCIAL MEDIA, INTERNET DE LAS COSAS (IoT) Y REALIDAD AUMENTADA EN LA EDUCACIÓN SUPERIOR UNIVERSITARIA"
.....
DR. EDGAR AURELIO TAYA ACOSTA
DIRECTOR PRINCIPAL

Anexo 5

Código de Google Colab

```
!pip install pyod
```

```
Collecting pyod
  Downloading pyod-3.6.0-py3-none-any.whl.metadata (58 kB)
    58.2/58.2 kB 3.2 MB/s eta 0:00:00
Requirement already satisfied: joblib>=1.5 in /usr/local/lib/python3.12/dist-packages (from pyod) (1.5.3)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.12/dist-packages (from pyod) (3.10.0)
Requirement already satisfied: numpy>=1.19 in /usr/local/lib/python3.12/dist-packages (from pyod) (2.0.2)
Requirement already satisfied: numba>=0.51 in /usr/local/lib/python3.12/dist-packages (from pyod) (0.60.0)
Requirement already satisfied: scipy>=1.5.1 in /usr/local/lib/python3.12/dist-packages (from pyod) (1.16.3)
Requirement already satisfied: scikit-learn>=0.22.0 in /usr/local/lib/python3.12/dist-packages (from pyod) (1.6.1)
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in /usr/local/lib/python3.12/dist-packages (from numba>=0.51->pyod) (0.43.0)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.12/dist-packages (from scikit-learn>=0.22.0->pyod) (3.1.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib->pyod) (1.3.3)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.12/dist-packages (from matplotlib->pyod) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib->pyod) (4.63.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib->pyod) (1.5.0)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib->pyod) (26.2)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.12/dist-packages (from matplotlib->pyod) (11.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib->pyod) (3.3.2)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.12/dist-packages (from matplotlib->pyod) (2.9.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.7->matplotlib->pyod) (1.16.0)
Downloading pyod-3.6.0-py3-none-any.whl (411 kB)
    411.4/411.4 kB 23.3 MB/s eta 0:00:00
Installing collected packages: pyod
Successfully installed pyod-3.6.0
```

```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.dates as mdates
from statsmodels.tsa.statespace.sarimax import SARIMAX
from pyod.models.knn import KNN
```

```
# 1. Cargar archivos
df = pd.read_csv("feeds (1).csv")
horario = pd.read_excel("CANTIDAD LAB A COMPLETO.xlsx")

# 2. Fechas del II semestre
df["created_at"] = pd.to_datetime(df["created_at"])
df = df[(df["created_at"] >= "2025-08-18") &
        (df["created_at"] <= "2025-12-31 23:59:59")]

# 3. Separar sensores
df_k30 = df[["created_at", "field1"]].dropna().rename(columns={"field1": "K30"})
df_mq135 = df[["created_at", "field2"]].dropna().rename(columns={"field2": "MQ135"})

df_k30 = df_k30.sort_values("created_at")
df_mq135 = df_mq135.sort_values("created_at")

# 4. Juntar K30 y MQ135 por fecha/hora más cercana
df_final = pd.merge_asof(
    df_k30,
    df_mq135,
    on="created_at",
    direction="nearest",
    tolerance=pd.Timedelta("30s")
)

# 5. Crear columnas de fecha, hora y día
df_final["FECHA"] = df_final["created_at"].dt.date
df_final["HORA"] = df_final["created_at"].dt.time

dias = {
    "Monday": "LUNES",
    "Tuesday": "MARTES",
    "Wednesday": "MIERCOLES",
    "Thursday": "JUEVES",
    "Friday": "VIERNES",
    "Saturday": "SABADO",
    "Sunday": "DOMINGO"
}

df_final["DIA"] = df_final["created_at"].dt.day_name().map(dias)

# 6. Preparar horario
horario["DIA"] = horario["DIA"].str.upper().str.strip()
horario["HORA INICIO"] = pd.to_datetime(horario["HORA INICIO"].astype(str)).dt.time
horario["HORA FIN"] = pd.to_datetime(horario["HORA FIN"].astype(str)).dt.time
```

```

# 7. Función para buscar curso según día y hora
def buscar_curso(fila):
    cursos = horario[
        (horario["DIA"] == fila["DIA"]) &
        (horario["HORA INICIO"] <= fila["HORA"]) &
        (horario["HORA FIN"] >= fila["HORA"])
    ]

    if len(cursos) > 0:
        curso = cursos.iloc[0]
        return pd.Series({
            "CURSO": curso["CURSO"],
            "CANTIDAD_ALUMNOS": curso["CANTIDAD"],
            "CICLO": curso["CICLO"],
            "TURN0": curso["TURN0"]
        })
    else:
        return pd.Series({
            "CURSO": "SIN CURSO",
            "CANTIDAD_ALUMNOS": 0,
            "CICLO": "",
            "TURN0": ""
        })

# 8. Agregar curso y cantidad de alumnos
datos_curso = df_final.apply(buscar_curso, axis=1)
df_final = pd.concat([df_final, datos_curso], axis=1)

# 9. Ordenar columnas finales
df_final = df_final[
    [
        "created_at",
        "FECHA",
        "DIA",
        "HORA",
        "K30",
        "MQ135",
        "CURSO",
        "CANTIDAD_ALUMNOS",
        "CICLO",
        "TURN0"
    ]
]

print(df_final)

```

```

/tmp/ipykernel_3383/2798180171.py:44: UserWarning: Could not infer format, so each element will be parsed individually, falling back to 'infer_datetime_format'.
horario["HORA INICIO"] = pd.to_datetime(horario["HORA INICIO"].astype(str)).dt.time
/tmp/ipykernel_3383/2798180171.py:45: UserWarning: Could not infer format, so each element will be parsed individually, falling back to 'infer_datetime_format'.
horario["HORA FIN"] = pd.to_datetime(horario["HORA FIN"].astype(str)).dt.time

```

	created_at	FECHA	DIA	HORA	K30	\
0	2025-08-18 00:00:28-05:00	2025-08-18	LUNES	00:00:28	403.0	
1	2025-08-18 00:00:44-05:00	2025-08-18	LUNES	00:00:44	403.0	
2	2025-08-18 00:01:00-05:00	2025-08-18	LUNES	00:01:00	403.0	
3	2025-08-18 00:01:15-05:00	2025-08-18	LUNES	00:01:15	403.0	
4	2025-08-18 00:01:31-05:00	2025-08-18	LUNES	00:01:31	403.0	
...	...	...	...	...	...	
464445	2025-12-31 23:56:35-05:00	2025-12-31	MIERCOLES	23:56:35	0.0	
464446	2025-12-31 23:57:26-05:00	2025-12-31	MIERCOLES	23:57:26	0.0	
464447	2025-12-31 23:58:34-05:00	2025-12-31	MIERCOLES	23:58:34	0.0	
464448	2025-12-31 23:59:22-05:00	2025-12-31	MIERCOLES	23:59:22	0.0	
464449	2025-12-31 23:59:53-05:00	2025-12-31	MIERCOLES	23:59:53	431.0	

	MQ135	CURSO	CANTIDAD_ALUMNOS	CICLO	TURN0
0	770.0	SIN CURSO	0		
1	NaN	SIN CURSO	0		
2	NaN	SIN CURSO	0		
3	NaN	SIN CURSO	0		
4	NaN	SIN CURSO	0		
...	...	...	...	...	...
464445	739.0	SIN CURSO	0		
464446	746.0	SIN CURSO	0		
464447	738.0	SIN CURSO	0		
464448	739.0	SIN CURSO	0		
464449	731.0	SIN CURSO	0		

[464450 rows x 10 columns]

```

df_final = df_final.dropna()
print(df_final)

```

	created_at	FECHA	DIA	HORA	K30	\
0	2025-08-18 00:00:28-05:00	2025-08-18	LUNES	00:00:28	403.0	

```

6      2025-08-18 00:02:02-05:00 2025-08-18 LUNES 00:02:02 403.0
7      2025-08-18 00:02:33-05:00 2025-08-18 LUNES 00:02:33 403.0
10     2025-08-18 00:03:20-05:00 2025-08-18 LUNES 00:03:20 402.0
11     2025-08-18 00:03:52-05:00 2025-08-18 LUNES 00:03:52 403.0
...
464445 2025-12-31 23:56:35-05:00 2025-12-31 MIERCOLES 23:56:35 0.0
464446 2025-12-31 23:57:26-05:00 2025-12-31 MIERCOLES 23:57:26 0.0
464447 2025-12-31 23:58:34-05:00 2025-12-31 MIERCOLES 23:58:34 0.0
464448 2025-12-31 23:59:22-05:00 2025-12-31 MIERCOLES 23:59:22 0.0
464449 2025-12-31 23:59:53-05:00 2025-12-31 MIERCOLES 23:59:53 431.0

```

```

MQ135  CURSO  CANTIDAD_ALUMNOS  CICLO  TURNO
0      770.0  SIN CURSO           0
6      751.0  SIN CURSO           0
7      751.0  SIN CURSO           0
10     767.0  SIN CURSO           0
11     767.0  SIN CURSO           0
...
464445 739.0  SIN CURSO           0
464446 746.0  SIN CURSO           0
464447 738.0  SIN CURSO           0
464448 739.0  SIN CURSO           0
464449 731.0  SIN CURSO           0

```

[248172 rows x 10 columns]

```
df_final['K30'] = df_final['K30'].replace(0, np.nan)
```

```
df_final['K30'] = df_final['K30'].interpolate(method='linear')
```

```
df_final['MQ135'] = df_final['MQ135'].replace(0, np.nan)
```

```
df_final['MQ135'] = df_final['MQ135'].interpolate(method='linear')
```

RENDIMIENTO

```

# =====
# COPIA DE DATOS
# =====

df_ind = df_final.copy()

# Asegurar formato fecha
df_ind['created_at'] = pd.to_datetime(df_ind['created_at'])

# Eliminar datos nulos importantes
df_ind = df_ind.dropna(subset=['K30', 'MQ135'])

# Eliminar valores cero del K30 si son errores del sensor
df_ind = df_ind[df_ind['K30'] > 0]

# =====
# 1. RENDIMIENTO
# =====

rendimiento = pd.DataFrame({
    'Sensor': ['K30', 'MQ135'],
    'Promedio': [
        df_ind['K30'].mean(),
        df_ind['MQ135'].mean()
    ],
    'Valor mínimo': [
        df_ind['K30'].min(),
        df_ind['MQ135'].min()
    ],
    'Valor máximo': [
        df_ind['K30'].max(),
        df_ind['MQ135'].max()
    ],
    'Cantidad de registros': [
        df_ind['K30'].count(),
        df_ind['MQ135'].count()
    ]
})

print("=== RENDIMIENTO ===")
display(rendimiento)

```

=== RENDIMIENTO ===

	Sensor	Promedio	Valor mínimo	Valor máximo	Cantidad de registros
0	K30	480.186880	356.0	1366.0	248172
1	MQ135	721.577124	293.0	1401.0	248172

```
plt.figure(figsize=(14,6))

# FECHAS EN EL EJE X
plt.plot(df_final['created_at'], df_final['K30'], label='K30 (CO2)')
plt.plot(df_final['created_at'], df_final['MQ135'], label='MQ135')

plt.title('Comparación de sensores (Ago-Dic 2025)')

# nombres de ejes
plt.xlabel('Fecha')
plt.ylabel('CO2 (ppm)')

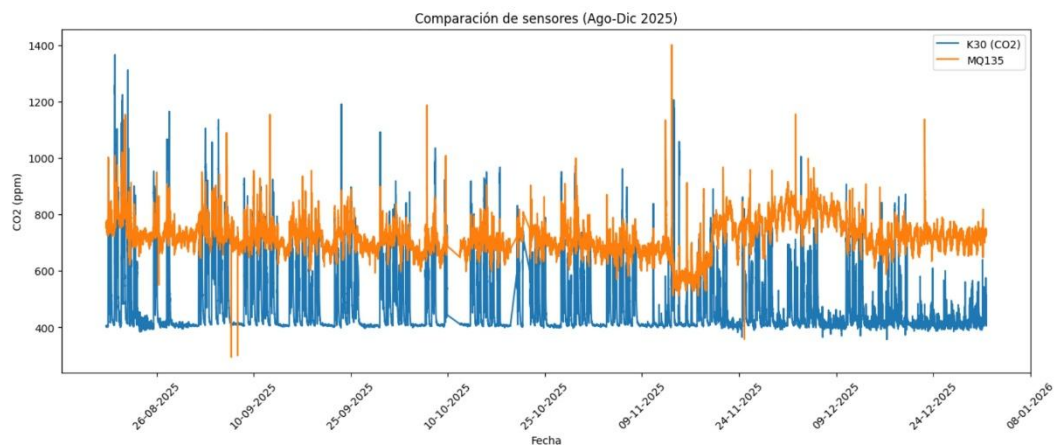
# mostrar fechas cada 15 días
plt.gca().xaxis.set_major_locator(mdates.DayLocator(interval=15))

# formato de fecha
plt.gca().xaxis.set_major_formatter(
    mdates.DateFormatter('%d-%m-%Y')
)

plt.legend()

# girar fechas
plt.xticks(rotation=45)

plt.tight_layout()
plt.show()
```



```
# =====
# CREAR FIGURA
# =====

fig, ax1 = plt.subplots(figsize=(15,6))

# =====
# GRAFICOS DE SENSORES
# =====

ax1.plot(
    df_final['created_at'],
    df_final['K30'],
    color='blue',
    linewidth=1,
```

```

        label='K30 (CO2)'
    )

    ax1.plot(
        df_final['created_at'],
        df_final['MQ135'],
        color='orange',
        linewidth=1,
        label='MQ135'
    )

    # etiquetas eje principal
    ax1.set_xlabel('Fecha')
    ax1.set_ylabel('CO2 (ppm)')

    # =====
    # SEGUNDO EJE (ALUMNOS)
    # =====

    ax2 = ax1.twinx()

    ax2.bar(
        df_final['created_at'],
        df_final['CANTIDAD_ALUMNOS'],
        color='purple',      # color barras
        alpha=0.20,          # transparencia
        width=0.8,
        label='Cantidad de alumnos'
    )

    ax2.set_ylabel('Cantidad de alumnos')

    # =====
    # PONER BARRAS DETRAS
    # =====

    ax2.set_zorder(0)
    ax1.set_zorder(1)
    ax1.patch.set_visible(False)

    # =====
    # FECHAS EN EJE X
    # =====

    ax1.xaxis.set_major_locator(
        mdates.DayLocator(interval=15)
    )

    ax1.xaxis.set_major_formatter(
        mdates.DateFormatter('%d-%m-%Y')
    )

    plt.xticks(rotation=45)

    # =====
    # TITULO
    # =====

    plt.title(
        'Sensores vs Cantidad de alumnos (Ago-Dic 2025)'
    )

    # =====
    # LEYENDA
    # =====

    lines1, labels1 = ax1.get_legend_handles_labels()
    lines2, labels2 = ax2.get_legend_handles_labels()

    ax1.legend(
        lines1 + lines2,
        labels1 + labels2,
        loc='upper right'
    )

    # =====
    # AJUSTAR GRAFICO
    # =====

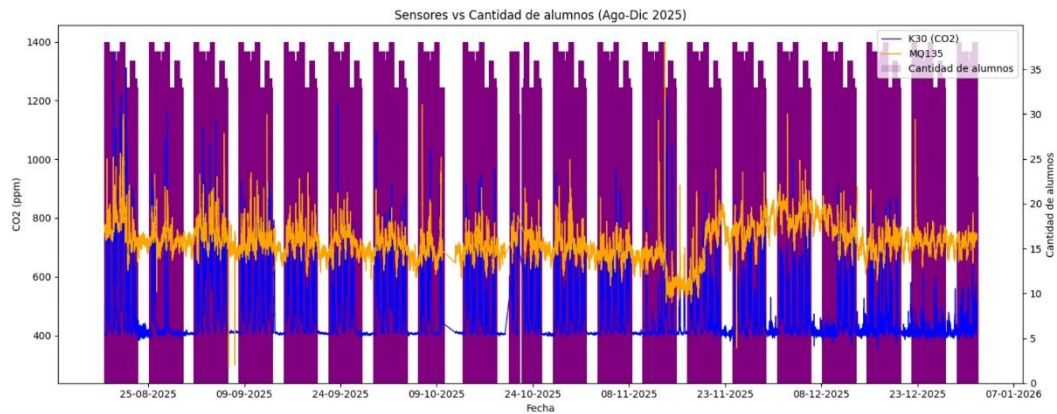
    plt.tight_layout()

    # =====

```

```
# MOSTRAR
# =====

plt.show()
```



✓ SENSIBILIDAD

```
# =====
# SENSIBILIDAD
# relación con cantidad de alumnos
# =====

sensibilidad = df_ind[
    ['K30', 'MQ135', 'CANTIDAD_ALUMNOS']
].corr()

print("=== SENSIBILIDAD / CORRELACIÓN ===")
display(sensibilidad)
```

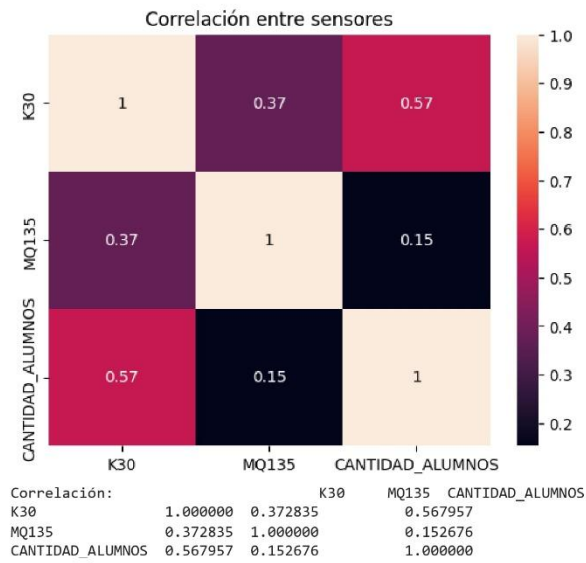
```
=== SENSIBILIDAD / CORRELACIÓN ===
```

	K30	MQ135	CANTIDAD_ALUMNOS
K30	1.000000	0.372835	0.567957
MQ135	0.372835	1.000000	0.152676
CANTIDAD_ALUMNOS	0.567957	0.152676	1.000000

```
corr_valor1 = df_final[['K30', 'MQ135', 'CANTIDAD_ALUMNOS']].corr()

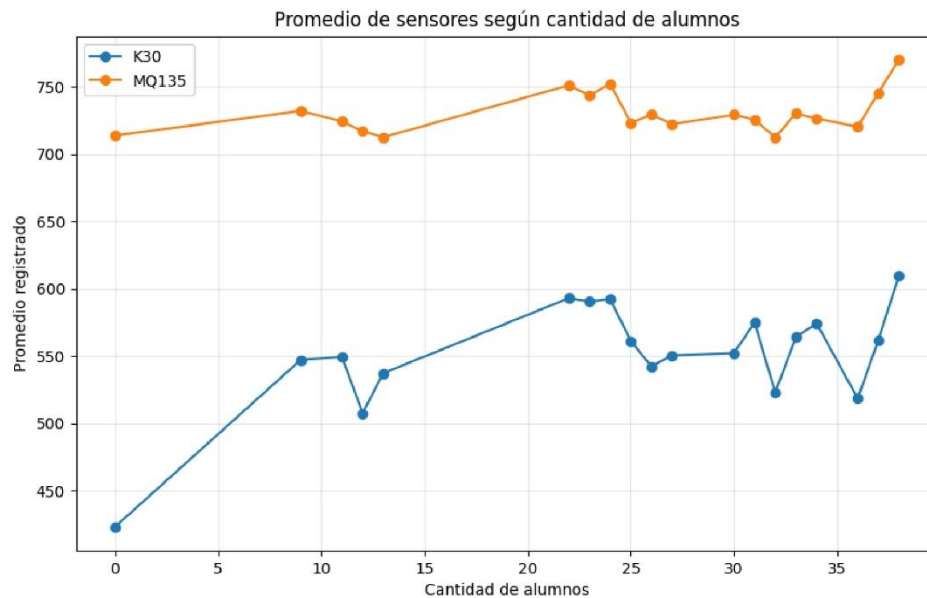
sns.heatmap(corr_valor1, annot=True)
plt.title("Correlación entre sensores")
plt.show()

print("Correlación:", corr_valor1)
```



```
#PROMEDIO DE CO2 VS CANTIDAD DE ALUMNOS
promedio_alumnos = df_final.groupby('CANTIDAD_ALUMNOS')[['K30', 'MQ135']].mean().reset_index()

plt.figure(figsize=(10,6))
plt.plot(promedio_alumnos['CANTIDAD_ALUMNOS'], promedio_alumnos['K30'], marker='o', label='K30')
plt.plot(promedio_alumnos['CANTIDAD_ALUMNOS'], promedio_alumnos['MQ135'], marker='o', label='MQ135')
plt.title('Promedio de sensores según cantidad de alumnos')
plt.xlabel('Cantidad de alumnos')
plt.ylabel('Promedio registrado')
plt.legend()
plt.grid(True, alpha=0.3)
plt.show()
```



ESTABILIDAD

```
# =====
# 2. ESTABILIDAD
# =====

estabilidad = pd.DataFrame({
    'Sensor': ['K30', 'MQ135'],
```

```

'Desviación estándar': [
    df_ind['K30'].std(),
    df_ind['MQ135'].std()
],
'Varianza': [
    df_ind['K30'].var(),
    df_ind['MQ135'].var()
],
'Coeficiente de variación': [
    df_ind['K30'].std() / df_ind['K30'].mean(),
    df_ind['MQ135'].std() / df_ind['MQ135'].mean()
]
])
print("=== ESTABILIDAD ===")
display(estabilidad)

# KNN (TRABAJA CON DISTANCIA)
#
# TECNICAS DE DETECCION DE OUTLIERS

```

```

=== ESTABILIDAD ===

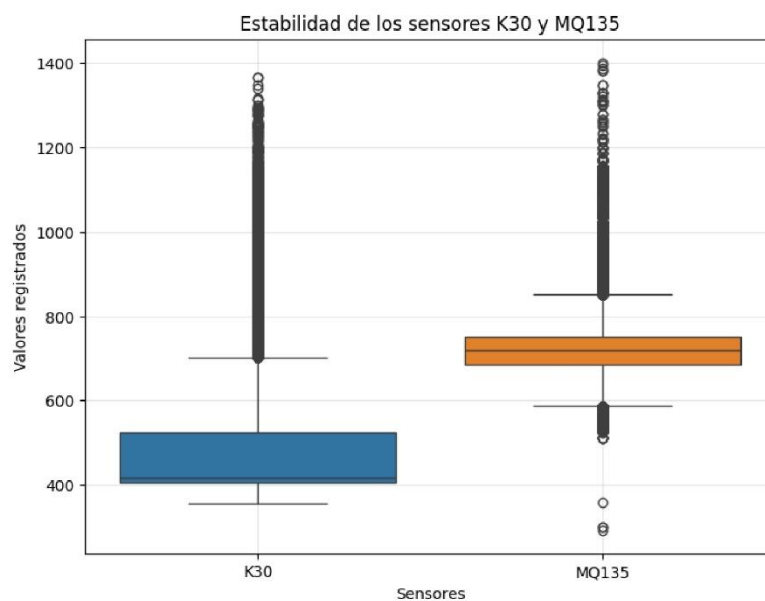
```

	Sensor	Desviación estándar	Varianza	Coeficiente de variación
0	K30	115.127029	13254.232718	0.239755
1	MQ135	63.667869	4053.597499	0.088234

```

#BOXPLOT
plt.figure(figsize=(8,6))
sns.boxplot(data=df_final[['K30', 'MQ135']])
plt.title('Estabilidad de los sensores K30 y MQ135')
plt.ylabel('Valores registrados')
plt.xlabel('Sensores')
plt.grid(True, alpha=0.3)
plt.show()

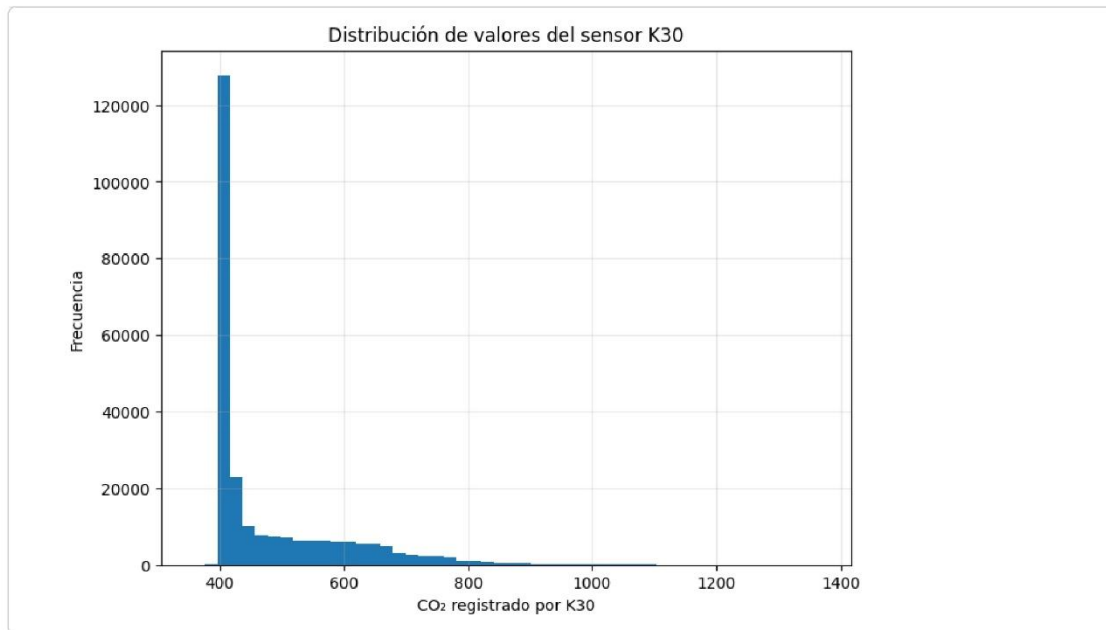
```



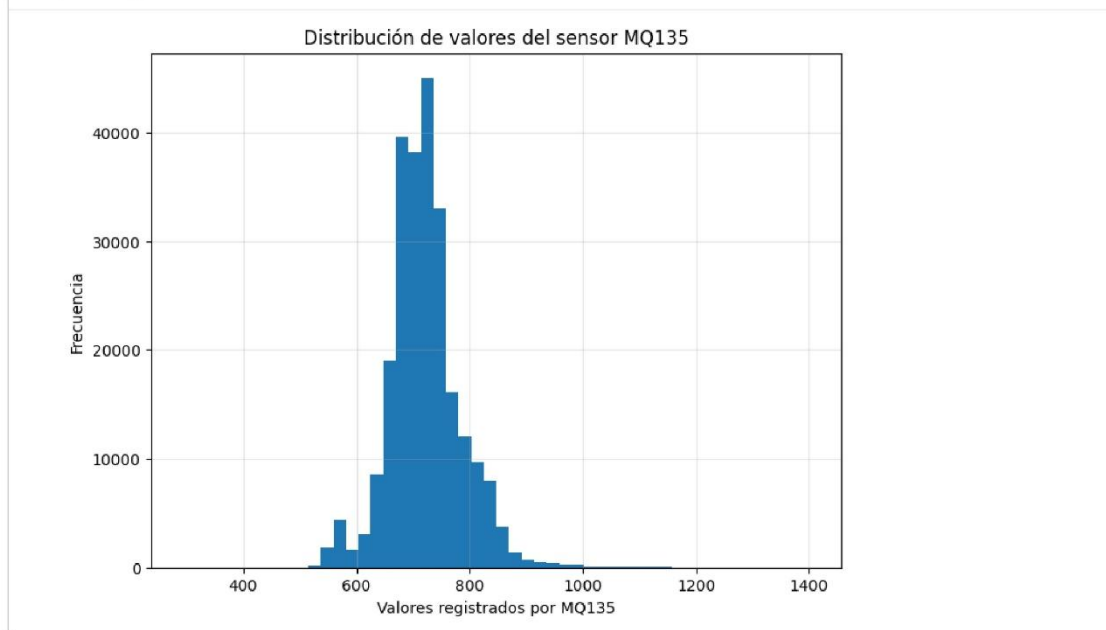
```

#HISTOGRAMA DEL K30
plt.figure(figsize=(8,6))
plt.hist(df_final['K30'].dropna(), bins=50)
plt.title('Distribución de valores del sensor K30')
plt.xlabel('CO2 registrado por K30')
plt.ylabel('Frecuencia')
plt.grid(True, alpha=0.3)
plt.show()

```



```
# HISTOGRAMA MQ135
plt.figure(figsize=(8,6))
plt.hist(df_final['MQ135'].dropna(), bins=50)
plt.title('Distribución de valores del sensor MQ135')
plt.xlabel('Valores registrados por MQ135')
plt.ylabel('Frecuencia')
plt.grid(True, alpha=0.3)
plt.show()
```



```
# seleccionar datos
datos_k30 = df_final[['K30']].dropna()

# modelo KNN
modelo_knn = KNN(contamination=0.02)

# entrenar
modelo_knn.fit(datos_k30)

# predicción
outliers_k30 = modelo_knn.labels_
```

```
# agregar columna
datos_k30['OUTLIER'] = outliers_k30

print(datos_k30.head())
```

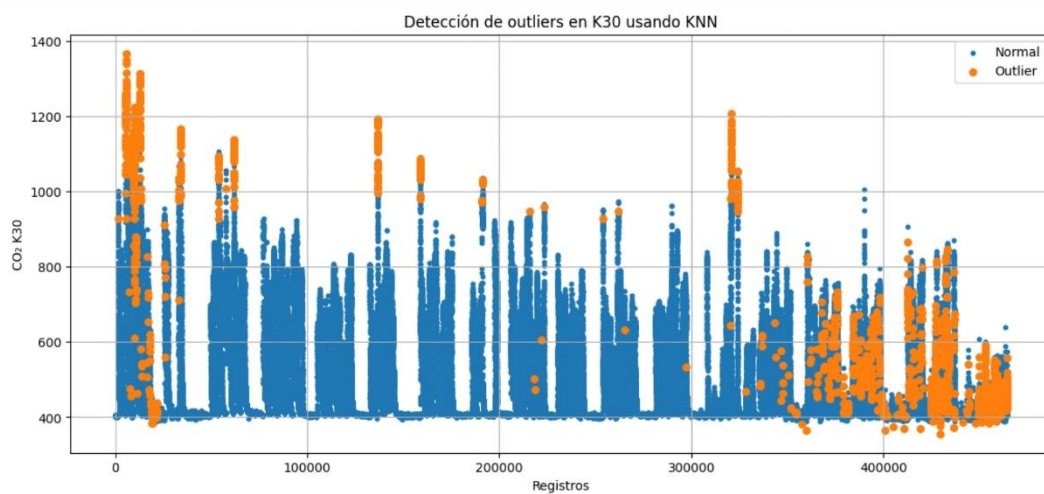
```
   K30  OUTLIER
0  403.0       0
6  403.0       0
7  403.0       0
10 402.0       0
11 403.0       0
```

```
plt.figure(figsize=(14,6))

# normales
normales = datos_k30[datos_k30['OUTLIER'] == 0]
plt.scatter(normales.index,
            normales['K30'],
            label='Normal',
            s=10)

# outliers
outliers = datos_k30[datos_k30['OUTLIER'] == 1]
plt.scatter(outliers.index,
            outliers['K30'],
            label='Outlier',
            s=30)

plt.title('Detección de outliers en K30 usando KNN')
plt.xlabel('Registros')
plt.ylabel('CO2 K30')
plt.legend()
plt.grid(True)
plt.show()
```



```
datos_mq = df_final[['MQ135']].dropna()

modelo_knn_mq = KNN(contamination=0.02)

modelo_knn_mq.fit(datos_mq)

outliers_mq = modelo_knn_mq.labels_

datos_mq['OUTLIER'] = outliers_mq
```

```
plt.figure(figsize=(14,6))

normales = datos_mq[datos_mq['OUTLIER'] == 0]
plt.scatter(normales.index,
            normales['MQ135'],
            label='Normal',
```

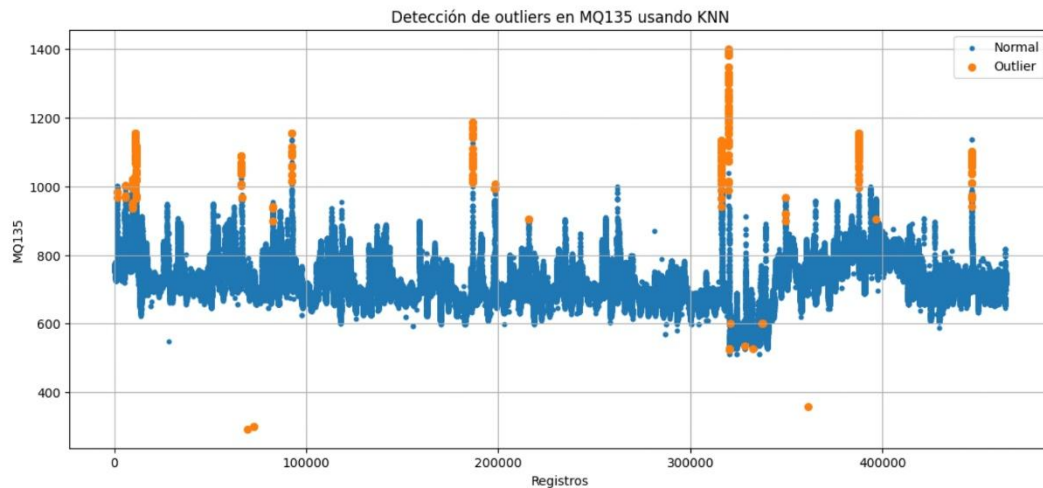
```

s=10)

outliers = datos_mq[datos_mq['OUTLIER'] == 1]
plt.scatter(outliers.index,
            outliers['MQ135'],
            label='Outlier',
            s=30)

plt.title('Detección de outliers en MQ135 usando KNN')
plt.xlabel('Registros')
plt.ylabel('MQ135')
plt.legend()
plt.grid(True)
plt.show()

```



```

cantidad_k30 = datos_k30['OUTLIER'].sum()
cantidad_mq = datos_mq['OUTLIER'].sum()

print("Outliers K30:", cantidad_k30)
print("Outliers MQ135:", cantidad_mq)

```

```

Outliers K30: 1810
Outliers MQ135: 389

```

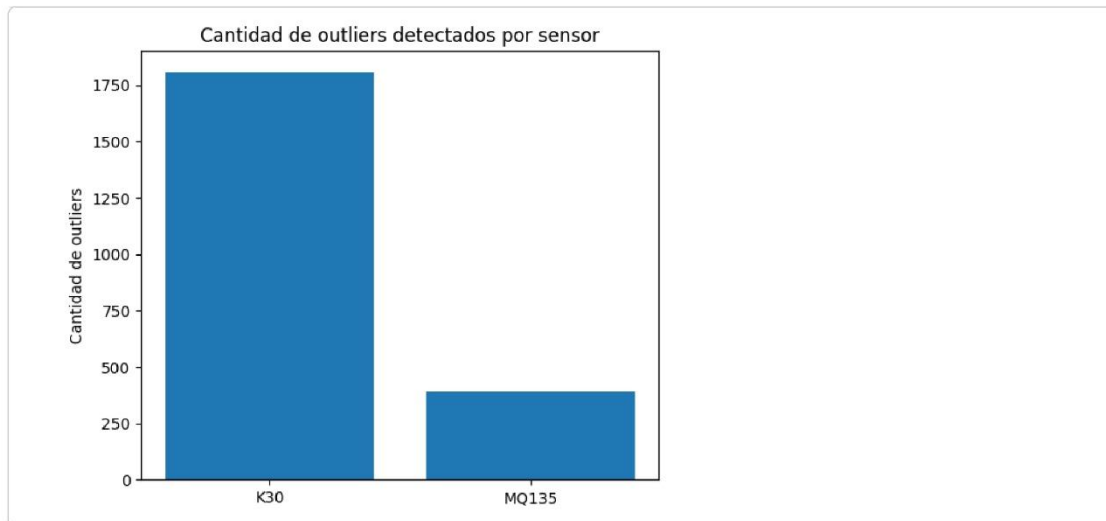
```

sensores = ['K30', 'MQ135']
outliers = [cantidad_k30, cantidad_mq]

plt.figure(figsize=(6,5))
plt.bar(sensores, outliers)

plt.title('Cantidad de outliers detectados por sensor')
plt.ylabel('Cantidad de outliers')
plt.show()

```



✓ PRECISIÓN DE PREDICCIÓN

✓ ARIMAX

K30 Y CANTIDAD DE ESTUDIANTES

```
y = df_final['K30']
df_final['HORA_NUM'] = pd.to_datetime(df_final['created_at']).dt.hour
df_final['DIA_SEMANA'] = pd.to_datetime(df_final['created_at']).dt.dayofweek

# Variable exógena
exog = df_final['CANTIDAD_ALUMNOS']

# División 80% - 20%
split = int(len(y) * 0.8)

y_train = y[:split]
y_test = y[split:]

exog_train = exog[:split]
exog_test = exog[split:]

print("Entrenamiento:", len(y_train))
print("Prueba:", len(y_test))
```

Entrenamiento: 198537
Prueba: 49635

```
model = SARIMAX(
    y_train,
    exog=exog_train,
    order=(3,1,1)
)

results = model.fit()
```

```
/usr/local/lib/python3.12/dist-packages/statsmodels/tsa/base/tsa_model.py:473: ValueWarning: An unsupported index was provided
self._init_dates(dates, freq)
/usr/local/lib/python3.12/dist-packages/statsmodels/tsa/base/tsa_model.py:473: ValueWarning: An unsupported index was provided
self._init_dates(dates, freq)
```

```
pred = results.predict(
    start=len(y_train),
    end=len(y_train)+len(y_test)-1,
    exog=exog_test
)
```

```
/usr/local/lib/python3.12/dist-packages/statsmodels/tsa/base/tsa_model.py:837: ValueWarning: No supported index is available
return get_prediction_index(
/usr/local/lib/python3.12/dist-packages/statsmodels/tsa/base/tsa_model.py:837: FutureWarning: No supported index is available
return get_prediction_index(
```

```

from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
import numpy as np

r2 = r2_score(y_test, pred)
mae = mean_absolute_error(y_test, pred)
rmse = np.sqrt(mean_squared_error(y_test, pred))

print("R²:", r2)
print("MAE:", mae)
print("RMSE:", rmse)

```

```

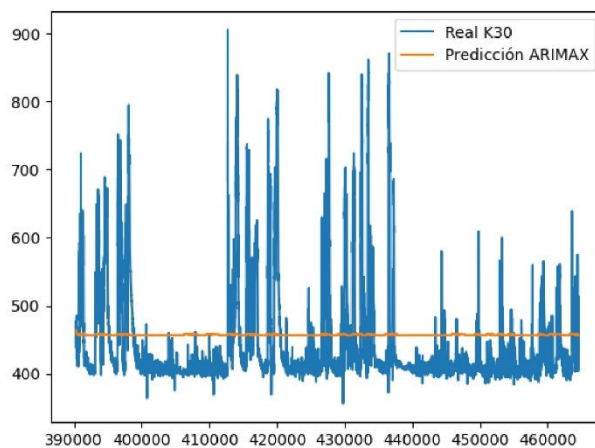
R²: -0.03734285774317425
MAE: 56.76776955214862
RMSE: 73.94436093561752

```

```

pred = pd.Series(pred.values, index=y_test.index)
plt.figure()
plt.plot(y_test, label='Real K30')
plt.plot(pred, label='Predicción ARIMAX')
plt.legend()
plt.show()

```



✓ ARIMAX

MQ135 Y CANTIDAD DE ESTUDIANTES

```

y1 = df_final['MQ135']

# Variable exógena
exog1 = df_final['CANTIDAD_ALUMNOS']

# División 80% - 20%
split1 = int(len(y1) * 0.8)

y_train1 = y1[:split1]
y_test1 = y1[split1:]

exog_train1 = exog1[:split1]
exog_test1 = exog1[split1:]

print("Entrenamiento:", len(y_train1))
print("Prueba:", len(y_test1))

```

```

Entrenamiento: 198537
Prueba: 49635

```

```

model1 = SARIMAX(
    y_train1,
    exog=exog_train1,
    order=(3,1,1)
)

results1 = model1.fit()

```

```

/usr/local/lib/python3.12/dist-packages/statsmodels/tsa/base/tsa_model.py:473: ValueWarning: An unsupported index was provided
self._init_dates(dates, freq)
/usr/local/lib/python3.12/dist-packages/statsmodels/tsa/base/tsa_model.py:473: ValueWarning: An unsupported index was provided
self._init_dates(dates, freq)

```

```

pred1 = results1.predict(
    start=len(y_train1),
    end=len(y_train1)+len(y_test1)-1,
    exog=exog_test1
)

```

```

/usr/local/lib/python3.12/dist-packages/statsmodels/tsa/base/tsa_model.py:837: ValueWarning: No supported index is available
return get_prediction_index(
/usr/local/lib/python3.12/dist-packages/statsmodels/tsa/base/tsa_model.py:837: FutureWarning: No supported index is available
return get_prediction_index(

```

```

from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
import numpy as np

r21 = r2_score(y_test1, pred1)
mae1 = mean_absolute_error(y_test1, pred1)
rmse1 = np.sqrt(mean_squared_error(y_test1, pred1))

print("R²:", r21)
print("MAE:", mae1)
print("RMSE:", rmse1)

```

```

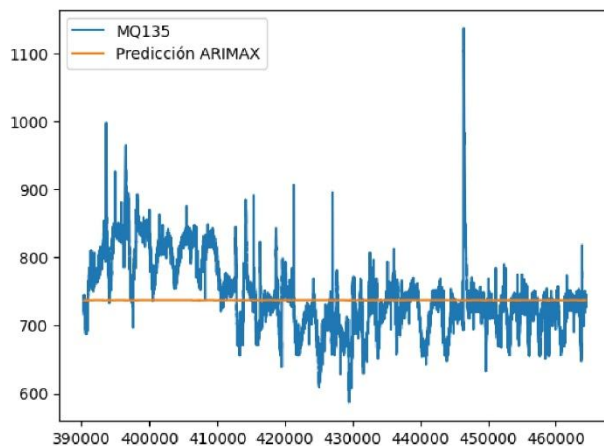
R²: -0.0049199809272653106
MAE: 40.26653064436078
RMSE: 53.90983318870251

```

```

pred1 = pd.Series(pred1.values, index=y_test1.index)
plt.figure()
plt.plot(y_test1, label='MQ135')
plt.plot(pred1, label='Predicción ARIMAX')
plt.legend()
plt.show()

```



✓ LSTM

```

from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

data_scaled = scaler.fit_transform(df_final[['K30', 'MQ135']])

```

```

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense

model_lstm = Sequential()

model_lstm.add(
    LSTM(

```

```

        50,
        activation='relu',
        input_shape=(X_train.shape[1], X_train.shape[2]))
    )
)

model_lstm.add(Dense(1))

model_lstm.compile(
    optimizer='adam',
    loss='mse'
)

```

/usr/local/lib/python3.12/dist-packages/keras/src/layers/rnn/rnn.py:199: UserWarning: Do not pass an `input\_shape` / `input\_dim` argument to the `__init__` method of `keras.layers.LSTM` or `keras.layers.SimpleRNN` layers. Use `input_shape` property instead.

```

def crear_secuencias(data, pasos=10):
    X, y = [], []
    for i in range(len(data)-pasos):
        X.append(data[i:i+pasos])
        y.append(data[i+pasos, 0]) # K30
    return np.array(X), np.array(y)

X, y = crear_secuencias(data_scaled)

```

```

split = int(len(X) * 0.8)

X_train = X[:split]
X_test = X[split:]

y_train = y[:split]
y_test = y[split:]

print("Entrenamiento:", len(X_train))
print("Prueba:", len(X_test))

```

Entrenamiento: 198529
Prueba: 49633

```

model_lstm.fit(
    X_train,
    y_train,
    epochs=10,
    verbose=1
)

```

```

Epoch 1/10
6205/6205 ————— 39s 6ms/step - loss: 2.1021e-04
Epoch 2/10
6205/6205 ————— 39s 6ms/step - loss: 6.3252e-05
Epoch 3/10
6205/6205 ————— 37s 6ms/step - loss: 5.9890e-05
Epoch 4/10
6205/6205 ————— 42s 6ms/step - loss: 5.9333e-05
Epoch 5/10
6205/6205 ————— 36s 6ms/step - loss: 5.8815e-05
Epoch 6/10
6205/6205 ————— 37s 6ms/step - loss: 5.8321e-05
Epoch 7/10
6205/6205 ————— 37s 6ms/step - loss: 5.8234e-05
Epoch 8/10
6205/6205 ————— 40s 6ms/step - loss: 5.8057e-05
Epoch 9/10
6205/6205 ————— 40s 6ms/step - loss: 5.7772e-05
Epoch 10/10
6205/6205 ————— 37s 6ms/step - loss: 5.7631e-05
<keras.src.callbacks.history.History at 0x7d4c67359ca0>

```

```
pred_lstm = model_lstm.predict(X_test)
```

```
1552/1552 ————— 3s 2ms/step
```

```

pred_lstm = scaler.inverse_transform(
    np.concatenate(
        (pred_lstm,
         np.zeros((len(pred_lstm),1))),
        axis=1
    )
)[:,:0]

real = scaler.inverse_transform(

```

```

np.concatenate(
    (y_test.reshape(-1,1),
     np.zeros((len(y_test),1))),
    axis=1
)[:,0]

```

```

mae_lstm = mean_absolute_error(real, pred_lstm)
rmse_lstm = np.sqrt(mean_squared_error(real, pred_lstm))
r2_lstm = r2_score(real, pred_lstm)

print("R²:", r2_lstm)
print("MAE:", mae_lstm)
print("RMSE:", rmse_lstm)

```

```

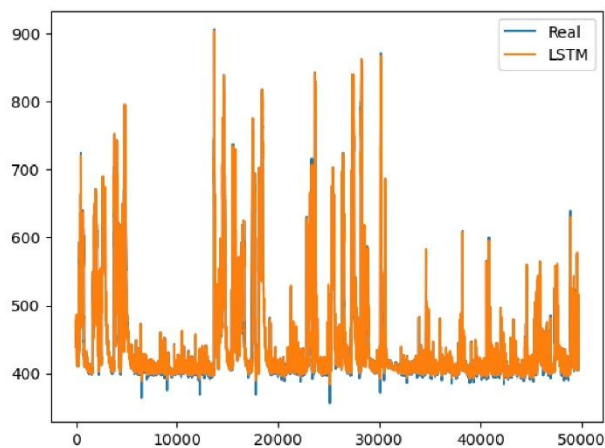
R²: 0.9867773284189413
MAE: 4.496380101184249
RMSE: 8.348559029557514

```

```

plt.figure()
plt.plot(real, label='Real')
plt.plot(pred_lstm, label='LSTM')
plt.legend()
plt.show()

```



✓ COMPLEMENTARIOS

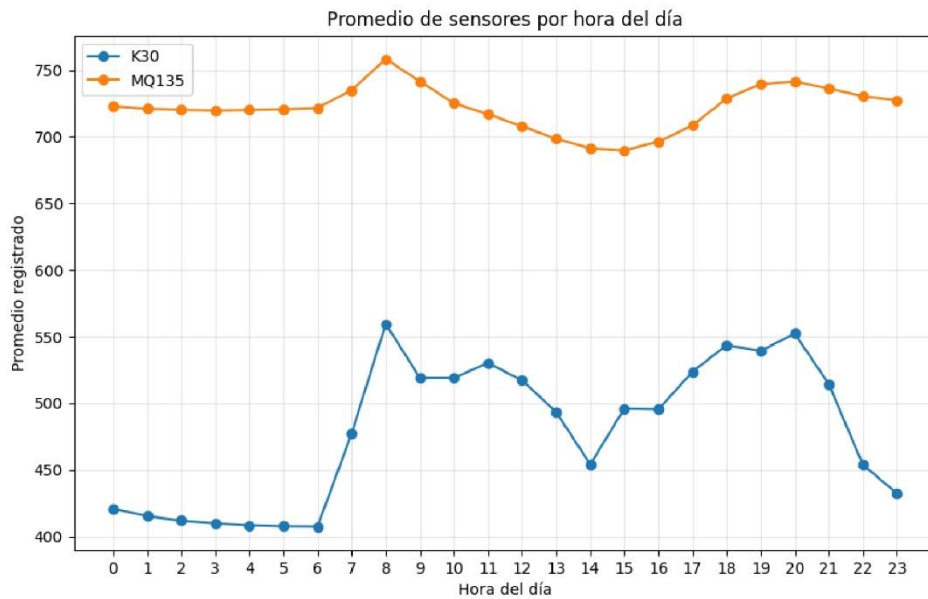
```

#PROMEDIO POR HORAS
df_final['HORA_NUM'] = df_final['created_at'].dt.hour

promedio_hora = df_final.groupby('HORA_NUM')[['K30', 'MQ135', 'CANTIDAD_ALUMNOS']].mean()

plt.figure(figsize=(10,6))
plt.plot(promedio_hora.index, promedio_hora['K30'], marker='o', label='K30')
plt.plot(promedio_hora.index, promedio_hora['MQ135'], marker='o', label='MQ135')
plt.title('Promedio de sensores por hora del día')
plt.xlabel('Hora del día')
plt.ylabel('Promedio registrado')
plt.xticks(range(0,24))
plt.legend()
plt.grid(True, alpha=0.3)
plt.show()

```



```
#PROMEDIO POR DIA A LA SEMANA
orden_dias = ['LUNES', 'MARTES', 'MIERCOLES', 'JUEVES', 'VIERNES', 'SABADO', 'DOMINGO']

promedio_dia = df_final.groupby('DIA')[['K30', 'MQ135', 'CANTIDAD_ALUMNOS']].mean()
promedio_dia = promedio_dia.reindex(orden_dias)

plt.figure(figsize=(10,6))
plt.plot(promedio_dia.index, promedio_dia['K30'], marker='o', label='K30')
plt.plot(promedio_dia.index, promedio_dia['MQ135'], marker='o', label='MQ135')
plt.title('Promedio de sensores por día de semana')
plt.xlabel('Día')
plt.ylabel('Promedio registrado')
plt.xticks(rotation=45)
plt.legend()
plt.grid(True, alpha=0.3)
plt.show()
```

