



**Análisis y detección de acumulación de gases en espacios cerrados, enfocados en
la mitigación de riesgos para la salud.**

Ing. Rivadeneira Zapata, Andrés Vinicio

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación previo a la obtención del título de Magíster en Gestión de Sistemas
de Información e Inteligencia de Negocios

Dr. Fonseca Carrera, Efraín Rodrigo

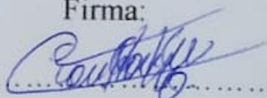
11 de mayo del 2020

URKUND**Document Information**

Analyzed document 12-Tesis-master-Rivadeneira-Andres-BI_URKUND.docx (D78249791)
Submitted 8/27/2020 9:13:00 PM
Submitted by Gualotuña Alvarez Tatiana Marisol
Submitter email tmgualotunia@espe.edu.ec
Similarity 0%
Analysis address tmgualotunia.espe@analysis.orkund.com

Sources included in the report

Firma:

**Dr. Fonseca Carrera Efraín Rodrigo****DIRECTOR**



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA**

CENTRO DE POSGRADOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, “**Análisis y detección de acumulación de gases en espacios cerrados, enfocados en la mitigación de riesgos para la salud**” fue realizado por el señor **Rivadeneira Zapata Andrés Vinicio** el mismo que ha sido revisado y analizado en su totalidad, por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 02 septiembre 2020

Firma:

Dr. Fonseca Carrera Efraín Rodrigo

Director

C.C.: 1710979574



VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA
CENTRO DE POSGRADOS

RESPONSABILIDAD DE AUTORÍA

Yo **Rivadeneira Zapata Andrés Vinicio**, con cédula de ciudadanía n° 1726247958, declaro que el contenido, ideas y criterios del trabajo de titulación: **Análisis y detección de acumulación de gases en espacios cerrados, enfocados en la mitigación de riesgos para la salud** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 02 septiembre 2020

Firma (s)

Rivadeneira Zapata Andrés Vinicio

C.C 1726247958



VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA

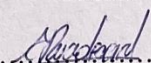
CENTRO DE POSGRADOS

AUTORIZACIÓN DE PUBLICACIÓN

Yo **Rivadeneira Zapata, Andrés Vinicio** autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **Análisis y detección de acumulación de gases en espacios cerrados, enfocados en la mitigación de riesgos para la salud** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 02 septiembre 2020

Firma


.....

Rivadeneira Zapata Andrés Vinicio

C.C 1726247958

DEDICATORIA

A mi futuro yo, al que me antecedió y al que hoy por hoy intenta ganarse la vida en un mundo de locuras y especulaciones, de verdades y mentiras, de amores y desengaños, pero por sobretodo busca el perfeccionamiento de su propio ser mas allá del valor del dinero o de una calificación, siempre depositando su confianza en la virtud mas grande del ser humano, la sabiduría, misma que nos impulsa a tomar buenas o malas decisión ya que de eso se trata la vida, equivocarse tanto hasta el momento en el esos errores sean el motor de que algo mucho mas grande esta por llegar.

AGRADECIMIENTO

A mi madre por siempre apoyarme incluso a veces obligarme a arriesgarme, ya que sin su ayuda jamás hubiese continuado con mi camino académico, el cual hoy por hoy me impulsa a otro nivel, a Tatiana y Carmen que son dos grandes maestras que confiaron en mi por sobre todas las cosas dándome mis primeras oportunidades de mostrar mi valor y agradezco a mi trabajo actual por mostrarme que el camino no es por ahí y que cada día debo buscar la manera de mejorar para encontrar un futuro mejor. En la vida es mandatorio aprender 3 cosas importantes: ser buenas personas, inglés y Excel.

Índice de contenidos

Urkund	2
Certificación	3
Responsabilidad de autoría.....	4
Autorización de publicación.....	5
Resumen	14
Abstract.....	15
Capítulo 1	16
Antecedentes	16
Planteamiento del problema.....	17
Objetivos	22
Objetivo general	22
Objetivos específicos	22
Justificación, importancia y alcance del proyecto	23
Justificación.....	23
Alcance	24
Hipótesis de investigación.....	26
Capítulo 2	27
Estudio del estado del arte.....	27
Objetivo.....	27
Preguntas de investigación	27
Criterios de inclusión	28
Criterios de exclusión	28

	9
Grupo de control	28
Cadena de búsqueda y resultado del pilotaje	29
Selección de estudios	31
Estudios primarios.....	31
Extracción de la información	32
Características del estado del arte	34
Metodología	34
Capítulo 3	40
Categorización de las variables.....	40
Fundamentación de la variable independiente	41
Machine learning.....	41
Deep learning.....	42
Minería de datos	42
Técnicas de aprendizaje automático	44
Fundamentación de la variable independiente:	44
Contaminación ambiental.....	44
Contaminación del aire.....	46
Fugas de gas en el aire.....	47
Fugas de gases en espacios cerrados	48
Capítulo 4	50
Construcción del prototipo.....	50
Requerimientos para el desarrollo de software.....	50
Componentes hardware	53
Herramientas de software	56
Diseño.....	57

	10
Diseño de casos de uso	59
Diseño secuencias de la aplicación	63
Construcción de la aplicación	64
Recolección y análisis de datos	65
Monóxido de carbono (CO)	68
Calidad de aire	70
Mitigación de riesgos para la salud en base a los resultados	72
Monóxido de carbono (CO)	72
Amoníaco, sensor de calidad de aire.....	74
Tolueno, sensor de calidad de aire.....	75
Selección de algoritmos predictivos	76
Escenario 1 (Rapid Minner):.....	77
Escenario 2:	79
Escenario 3:	79
Escenario 4:	80
Capítulo 5	82
Evaluación de los modelos.....	82
Validación de los modelos seleccionados.	84
Exploración	87
Ubicación de la aplicación dentro de la arquitectura de software	91
Diagramas de secuencia de la aplicación.....	92
Codificación	93
Pruebas	94
Capítulo 6	95

	11
Conclusiones:	96
Recomendaciones:	97
Trabajos Futuros	98
BIBLIOGRAFÍA.....	98

Índice de tablas

Tabla 1 <i>Objetivos específicos - preguntas de investigación</i>	24
Tabla 2 <i>Grupo de control</i>	29
Tabla 3 <i>Términos clave</i>	30
Tabla 4 <i>Cadena de búsqueda</i>	30
Tabla 5 <i>Resumen de estudios candidatos</i>	31
Tabla 6 <i>Estudios primarios</i>	31
Tabla 7 <i>Historia de usuario 1</i>	51
Tabla 8 <i>Historia de usuario 2</i>	51
Tabla 9 <i>Historia de usuario 3</i>	52
Tabla 10 <i>Historia de usuario 4</i>	52
Tabla 11 <i>Historia de usuario 5</i>	53
Tabla 12 <i>Componentes y características del NSCI</i>	54
Tabla 13 <i>Herramientas de software</i>	56
Tabla 14 <i>Resumen métricas algoritmos predictivos</i>	83
Tabla 15 <i>Historia de usuario 1 aplicación de presentación</i>	87
Tabla 16 <i>Historia 2 aplicación de presentación</i>	88
Tabla 17 <i>Historia de usuario 3 aplicación de presentación</i>	89
Tabla 18 <i>Historia de usuario 4 aplicación de presentación</i>	90

Tabla 19 <i>Caso de prueba APW1</i>	94
Tabla 20 <i>Caso de prueba APW2</i>	94
Tabla 21 <i>Caso de prueba APW1</i>	95

Índice de Figuras

Figura 1 Causas principales	17
Figura 2 Causas derivadas.....	18
Figura 3 Causas derivadas de fallas humanas	19
Figura 4 Efectos principales	19
Figura 5 Árbol de problemas	21
Figura 6 Metodología Ad-hoc.....	35
Figura 7 Categorización de las variables	40
Figura 8 Ciclo de vida de machine learning.....	41
Figura 9 Red neuronal deep learning.....	42
Figura 10 Proceso de minería de datos.....	43
Figura 11 Contaminación ambiental.....	45
Figura 12 Contaminación del aire.....	47
Figura 13 Dispositivo NSCI.....	55
Figura 14 Modelo de almacenamiento de datos	58
Figura 15 Arquitectura NCSI	59
Figura 16 Casos de uso SB-NCSI	61
Figura 17 Casos de uso SB-DRA.....	62
Figura 18 Casos de uso SB-MDRA.....	62
Figura 19 Diagrama de secuencia SB-NCSI	63
Figura 20 Diagrama de secuencia SB-DRA.....	63

Figura 21 Diagrama de secuencia SB-MRDA	64
Figura 22 GLP en el tiempo.....	66
Figura 23 Humo en el tiempo	66
Figura 24 Monóxido de carbono en el tiempo	68
Figura 25 Distribución de valores para CO	69
Figura 26 Correlación de las variables para la predicción del CO	69
Figura 27 Calidad de aire en el tiempo.....	71
Figura 28 Distribución de valores para calidad de aire	71
Figura 29 Correlación de las variables de la calidad de aire.....	72
Figura 30 Concentraciones y efectos del CO	73
Figura 31 Concentraciones aceptables de amoníaco	74
Figura 32 Niveles de toxicidad de Tolueno	75
Figura 33 Error medio absoluto Escenario 1	78
Figura 34 Error medio cuadrático del escenario 1	78
Figura 35 Errores modelo escenario 2.....	79
Figura 36 Error medio absoluto escenario 3.....	80
Figura 37 Error medio cuadrático escenario 3.....	80
Figura 38 Resumen de variables escenario 4	81
Figura 39 Datos de calidad de aire para pronóstico	84
Figura 40 Predicción Calidad de aire	85
Figura 41 Datos de CO para pronóstico.....	85
Figura 42 Predicción CO.....	86
Figura 43 Diagrama de arquitectura de software.....	92
Figura 44 Diagrama de secuencia visualización de datos.....	93
Figura 45 Diagrama de secuencia predicción de datos	93

Resumen

Antecedentes: los resultados del último censo realizado por el Instituto Ecuatoriano de Estadísticas y Censos (INEC) en 2010 muestran que en el Ecuador un 90.66% de los hogares y locales comerciales utilizan cilindros de gas para las actividades de cocina. **Problema:** en los últimos años en la ciudad de Quito han existido varias emergencias atendidas por temas de fugas de gas; además, de una gran cantidad de afectaciones a la salud de las personas por contaminación del aire en espacios cerrados producto de la combustión y la falta de ventilación. **Objetivo:** implementar un sistema predictivo de bajo costo enfocado en el análisis y monitorización de la acumulación de gases en espacios cerrados, utilizando técnicas de minería de datos y sistemas empujados basados en hardware libre. **Metodología:** se utilizó una metodología propia (Ad-Hoc) que consta de cinco fases: evaluación de la situación actual, estudio de viabilidad de la solución, diseño y construcción de la solución, validación de la solución y presentación de resultados. **Resultados:** se creó un sistema predictivo de bajo costo con Raspberry Pi y Arduino que alcanzó un porcentaje de precisión que oscila entre el 92% y el 99%. **Conclusión:** la construcción de modelos predictivos utilizando redes neuronales recurrentes (RNN) permitieron analizar, monitorizar y predecir la acumulación de monóxido de carbono (CO) y determinar la concentración conglomerada de gases como el amoníaco, sulfuro y tolueno.

Palabras clave:

- **SISTEMAS PREDICTIVO**
- **CIENCIA DE DATOS**
- **HARDWARE LIBRE**
- **ACUMULACIÓN DE GASES**

Abstract

Background: the results of the last census realized by the Instituto Ecuatoriano de Estadísticas y Censos (INEC) in 2010 show that in Ecuador 90.66% of households and commences use gas cylinders for cooking activities. **Issue:** Quito in the last years have been having several emergencies attended by issues of gas leaks; in addition, people health was in danger as consequence of air pollution in closed spaces caused by combustion of gas and lack of ventilation. **Aim:** Implement a low-cost predictive system focused on the analysis and monitoring of the accumulation of gases in closed spaces, using data mining techniques and embedded systems based on open hardware. **Methodology:** An own methodology (Ad-Hoc) consisting of five phases was used: evaluation of the current situation, feasibility study of the solution, design and construction of the solution, validation of the solution and presentation of results. **Result:** A low-cost predictive system was created with Raspberry Pi and Arduino that reached a precision percentage that ranges between 92% and 99%. **Conclusion:** The construction of predictive models using recurrent neural networks (RNN) allowed to analyze, monitor and predict the accumulation of carbon monoxide (CO) and to determine the conglomerate concentration of gases such as ammonia, sulfur and toluene.

Key words

- **PREDICTIVE SYSTEM**
- **DATA SCIENCE**
- **OPEN HARDWARE**
- **GASES ACCUMULATION**

Capítulo 1

Evaluación de la situación actual

Antecedentes

Los accidentes en torno a la utilización del gas licuado de petróleo como combustible son muy comunes alrededor del mundo. Por ejemplo, en Europa estos accidentes han ocurrido en varios países (es.euronews.com):

- Una estación de esquí en Polonia estalló en diciembre de 2019, quitando la vida a 8 personas (Euronews en español, 2019).
- En Francia 3 personas resultaron heridas tras el estallido de bombonas de gas en un laboratorio mientras se daba mantenimiento, (es.euronews.com, 2019).
- Eslovaquia, 5 persona perdieron la vida luego de que en un edificio de 12 pisos estallar una bombona de gas (es.euronews.com, 2019)
- Rusia, desde el 2018 en (es.euronews.com) aparecen en los titulares de las noticias algunos incidentes producidos por el uso de gas como combustible. Una explosión de gas destruyó alrededor de 35 apartamentos de un edificio y dañó seriamente 10 viviendas a la redonda, perdiendo la vida 37 personas en el suceso (Pascual, euronews, 2019), 7 muertos y 30 desaparecido al colapsar parcialmente un edificio luego de una explosión de gas (es.euronews.com, 2018), 22 muertos por un derrumbe producido por una explosión de gas en un edificio (Pascual, es.euronews.com, 2019), se desplomó un edificio tras estallar una bombona de gas (es.euronews.com, 2018).

El continente americano no es la excepción. Por ejemplo, en EEUU alrededor de 25 personas resultaron heridas tras múltiples explosiones de bombonas de gas en 2018, debido a un problema de presurización (Martín-Fradejas, 2018). En el del Valle del Cauca en Colombia se registró una explosión de gas en una notaria que dejó alrededor de 6 heridos (AseoYumbo, 2019).

En el Ecuador los datos del último censo realizado por el INEC en 2010 muestran que un 90.66% de los hogares y locales comerciales utilizan cilindros de gas para las actividades de cocina (Censos, 2011). Por lo que los accidentes son muy comunes, un reporte relacionado menciona que hay una explosión de gas cada 82 horas solo en la ciudad de Quito (Jácome, 2015), a continuación, se detalla un análisis causa efecto de la problemática presentada en torno a este tema.

Planteamiento del problema

La problemática que será abordada en esta investigación es la referente a las fugas de gas en espacios cerrados. En base a una revisión básica de literatura, determinamos que este problema se debe principalmente a la carencia de monitoreo y control de las instalaciones de gas, y a los errores humanos (ver Figura 1) (Vásconez, Celi, Ayabaca, & Rocha-Hoyos, 2018).

Figura 1

Causas principales

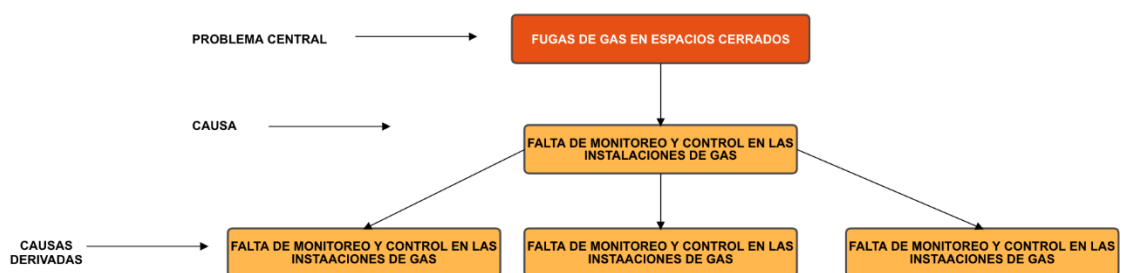


Pese a la existencia de la norma NTE INEN 2260 (2010) para instalaciones de gas, esta no se aplica a cabalidad al momento de implementar nuevas instalaciones, ya que se omiten parámetros adecuados para el diseño y almacenamiento de las

bombonas de gas. Si a esto le sumamos que únicamente la norma se enfoca en buenas practicas de instalación de los cilindros utilizados como combustible, pero no se hace referencia a ningún manual de parámetros de calidad o requerimientos mínimos que deben cumplir los espacios cerrados para calificar como bodegas de gas, tenemos un peligro potencial. Por otra parte, La falta de un seguimiento y mantenimiento frecuente de los componentes que integran una instalación de gas, también presenta un riesgo para las personas (ver Figura 2) (Vásconez, Celi, Ayabaca, & Rocha-Hoyos, 2018).

Figura 2

Causas derivadas.

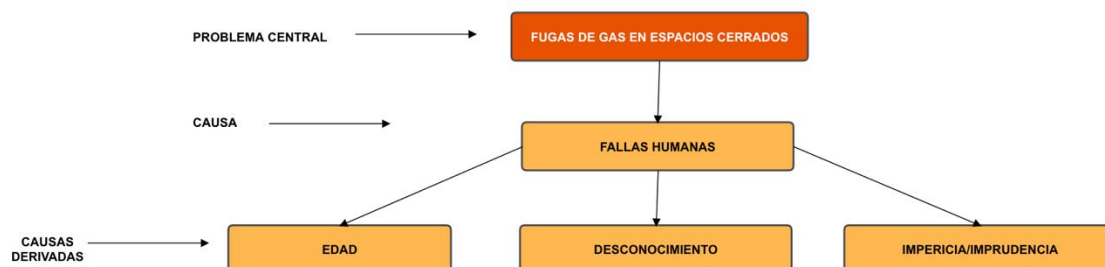


La falta de controles permite que el uso inadecuado de los cilindros de gas de uso domestico e industrial, provoque en las cocinas alta peligrosidad, pues al no contar con las medidas de seguridad adecuadas representan un alto riesgo de explosión (metroecuador, 2017) (lahora, 2017).

Las fallas humanas son otra de las causas principales de esta problemática (Ver Figura 3), las cuales se ocasionan principalmente por el desconocimiento, la impericia e imprudencia de los usuarios de cocinas a gas de uso domestico. Como consecuencia, los accidentes ocasionados por fugas de gas en la ciudad de Quito van en aumento (Jácome, 2015).

Figura 3

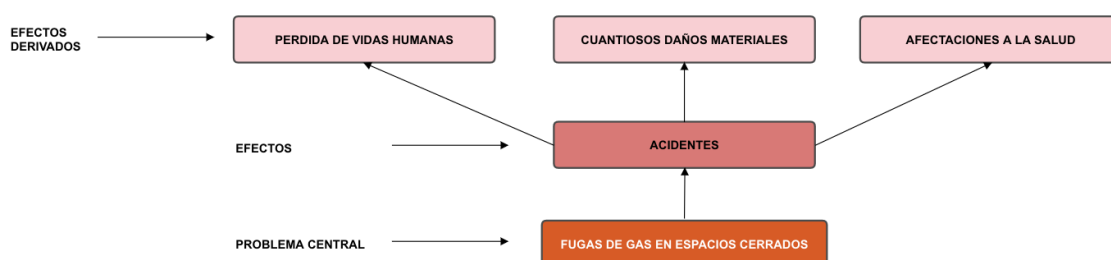
Causas derivadas de fallas humanas



De las causas presentadas anteriormente, los efectos más graves son los accidentes provocados por fugas de gas (ver Figura 4); un estudio realizado en 2015 por la Secretaría Nacional de Riesgos, encontró que existieron alrededor de 280 personas heridas, 33 fallecidos, 170 viviendas afectadas y 46 completamente destruidas, por situaciones relacionadas con las fugas de gas (Servicio Nacional de Gestión de Riesgos y Emergencias, 2015) (elcomercio, 2015). Además, en el año 2017 en la ciudad de Quito se registraron alrededor de 1012 casos de fugas de gas que fueron resueltos por el cuerpo de Bomberos, estos eventos ocasionaron la muerte de personas por intoxicación y heridos por explosiones derivadas (elcomercio, 2017). Si a esto le agregamos el caso del restaurante Toronto de diciembre de 2017, donde perdieron la vida 2 personas y se tuvieron cuantiosos daños materiales, se puede apreciar claramente que el uso, mantenimiento y evaluación del estado de las instalaciones de gas tanto doméstico como industrial, no es una práctica habitual que se realiza en el país (metroecuador, 2017).

Figura 4

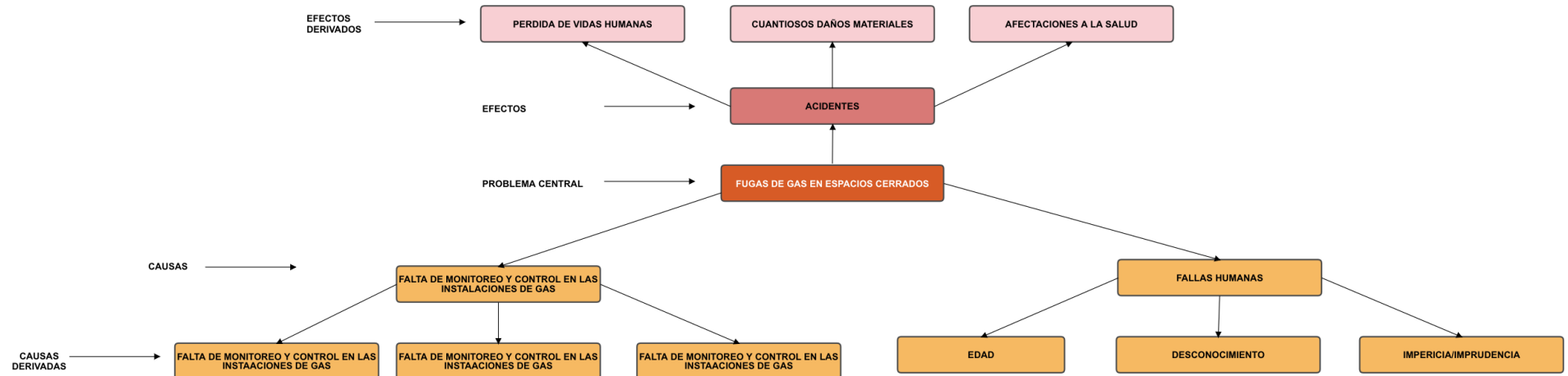
Efectos principales



Dadas estas circunstancias para conceptualizar la problemática de este trabajo, en la Figura 5 se presenta un árbol de problemas donde se unen todas las causas y efectos derivados de problema central: Fugas de gas en espacios cerrados.

Figura 5

Árbol de problemas



En el presente proyecto se abarcó la falta de monitoreo y control de las instalaciones de gas, de tal manera que se pueda establecer el estado actual de las mismas, descubriendo a qué gases son expuestas las personas y qué consecuencias podrían presentarse al respecto; la otra causa del árbol de problemas, no se tomó en cuenta para este estudio, ya que se involucran un sin número de factores humanos que corresponden a variables externas al experimento asociado.

Objetivos

Objetivo general

Implementar un sistema predictivo de bajo costo enfocado en el análisis y monitorización de la acumulación de gases en espacios cerrados, utilizando técnicas de minería de datos y sistemas empujados basados en hardware libre.

Objetivos específicos

OE1: Definir la problemática relacionada a la acumulación de gases en espacios cerrados mediante un análisis causa efecto por medio de una investigación preliminar de literatura.

OE2: Estudiar la viabilidad de una propuesta de solución a los problemas en torno a la acumulación de gas en espacios cerrados a través de una revisión sistemática de literatura preliminar.

OE3: Diseñar y construir la solución de hardware y software que permita la recolección y análisis de datos referente a la acumulación de gases en espacios cerrados de un escenario real.

OE4: Evaluar algoritmos que permitan establecer modelos predictivos de los datos del proyecto referentes a la acumulación de gases en espacios cerrados, por medio de herramientas utilizadas en ciencia de datos.

OE5: Validar la precisión de los algoritmos seleccionados para la predicción de los datos de entrada referentes a la acumulación de gases en espacios cerrados, estableciendo métricas que permitan evaluar el error de los datos reales contra los predichos.

OE6: Presentar los resultados de la experimentación a través de una aplicación web que permita la visualización y pronóstico, en base a los modelos de predicción para los datos referentes a la acumulación de gases en espacios cerrados.

Justificación, importancia y alcance del proyecto

Justificación

La falta de políticas de control derivadas de una escasa asignación de recursos a los proyectos que fomenten la revisión de las instalaciones de gas licuado de petróleo (GLP) de uso doméstico o industrial, por parte de los organismos de control estatal y la poca o nula monitorización de los parámetros idóneos de calidad de aire en estos sitios provocada por el desconocimiento de los organismos de control que dictan estos estándares, sumado a las fallas humanas causadas por el desconocimiento, la impericia y la imprudencia de las personas que diariamente interactúan con este combustible, han provocado que las alarmantes cifras presentadas anteriormente aún sigan en aumento. Por lo tanto, la implementación de este proyecto se justifica ampliamente, ya que al no realizarlo posiblemente cientos de personas seguirán siendo afectadas por los accidentes derivados del uso inadecuado del GLP.

Alcance

Con el propósito de volver alcanzables a los objetivos específicos fue necesario plantear preguntas de investigación para abarcar una sección particular de la solución, lo que se puede apreciar en la Tabla 1.

Tabla 1

Objetivos específicos - preguntas de investigación

Objetivos	Preguntas de investigación
OE1: Definir la problemática relacionada a la acumulación de gases en espacios cerrados mediante un análisis causa efecto por medio de una investigación preliminar de literatura.	OE1-PI1: ¿Cuáles son las causas principales por las cuales se producen fugas de gas en espacios cerrados? OE1-PI2: ¿Cuáles son las consecuencias más tangibles en torno a la acumulación de gas en espacios cerrados?
OE2: Estudiar la viabilidad de una propuesta de solución a los problemas en torno a la acumulación de gas en espacios cerrados a través de una revisión sistemática de literatura preliminar.	OE2-PI3: ¿Qué dispositivos son los más utilizados en las soluciones propuestas en los artículos encontrados? OE2-PI4: ¿Con qué frecuencia se han presentado propuestas relacionadas con el control de gases en ambientes cerrados que incluyan un modelo predictivo?
OE3: Diseñar y construir la solución de hardware y software que permita la recolección y análisis de datos referente a la acumulación de gases en espacios cerrados de un escenario real	OE3-PI5: ¿Qué requerimientos son necesarios para construir un sistema predictivo de bajo costo enfocado en el análisis y monitorización de la acumulación de gases en espacios cerrados? OE3-PI6: ¿Qué componentes de hardware y software son requeridos para diseñar un sistema predictivo de bajo costo enfocado en el análisis y monitorización de la acumulación de gases en espacios cerrados?

Objetivos	Preguntas de investigación
<p>OE4: Evaluar algoritmos que permitan establecer modelos predictivos de los datos del proyecto referentes a la acumulación de gases en espacios cerrados, por medio de herramientas utilizadas en ciencia de datos.</p>	<p>OE3-PI7: ¿Qué variables finalmente se utilizaron en el experimento luego del culminado el proceso de descubrimiento de datos?</p> <p>OE3-PI8: ¿Qué características esenciales presentan los datos de las variables del experimento?</p> <p>OE3-P9: ¿Qué efectos puede producir en la salud de las personas los datos recopilados en este experimento?</p> <p>OE3-PI10: ¿De qué manera se puede reducir los efectos encontrados luego del análisis de los datos recopilados en este experimento?</p> <p>OE4-PI11: ¿Qué algoritmos son los más comunes para realizar predicciones según las características de los datos del experimento?</p> <p>OE4-PI12: ¿Con qué herramientas se puede realizar modelos predictivos según las características de los datos del experimento?</p>
<p>OE5: Validar la precisión de los algoritmos seleccionados para la predicción de los datos de entrada referentes a la acumulación de gases en espacios cerrados, estableciendo métricas que permitan evaluar el error de los datos reales contra los predecidos.</p>	<p>OE5-PI13: ¿Cuáles son las métricas que permiten determinar si un modelo predictivo es más adecuado que otro en función del conjunto inicial de datos?</p> <p>OE5-PI14: ¿Qué algoritmo es el más preciso para realizar la predicción de datos de la solución?</p>
<p>OE6: Presentar los resultados de la experimentación a través de una aplicación web que permita la visualización y pronóstico, en base a los modelos de predicción para los datos referentes a la acumulación de gases en espacios cerrados.</p>	<p>OE6-P15: ¿Cuáles son los requerimientos para construir una aplicación que permita visualizar la situación actual del sistema predictivo de bajo costo enfocado en el análisis y monitorización de la acumulación de gases en espacios cerrados?</p> <p>OE6-P116: ¿Qué información es la más destacada para presentar en la aplicación que permita visualizar la situación actual del sistema predictivo</p>

Objetivos	Preguntas de investigación
	de bajo costo enfocado en el análisis y monitorización de la acumulación de gases en espacios cerrados?

Hipótesis de investigación

H1: Un modelo de gestión de información basado en técnicas de aprendizaje de automático permite establecer modelos de predicción de los datos de fugas de gases en espacios cerrados.

Capítulo 2

Marco metodológico

Estudio del estado del arte

A menudo gran cantidad de las ideas de solución de problemáticas de un área de interés específico se ven solventadas en otras partes del mundo de forma total o parcial, la intención de revisar los trabajos relacionados al tema de investigación es precisamente conocer que han realizado otros autores en torno a las innumerables fugas de gas que provocan accidentes y cuantiosos daños materiales o descubrir los acercamientos realizados sobre éste tema, para lo cual se realizó una revisión sistemática preliminar de literatura en la base de datos digital IEEE Xplorer siguiendo los lineamientos de Kitchenham & otros (Velásquez, 2015), que se detalla a continuación.

Objetivo

Estudiar la viabilidad de una propuesta de solución a los problemas en torno a la acumulación de gas en espacios cerrados a través de una revisión sistemática preliminar de literatura para determinar una línea base en la implementación de esta.

Preguntas de investigación

Las preguntas de investigación planteadas para el estudio del estado del arte corresponden con las preguntas de investigación de OE2 (Ver Capítulo 1).

Criterios de inclusión

Los criterios de inclusión y exclusión se definen para identificar estudios primarios que respondan a las características deseadas para la revisión. Los criterios de inclusión definidos para este caso son los siguientes:

- Artículo debe presentar los resultados con pruebas funcionales en ambientes reales controlados.
- Artículo con soluciones basadas en internet de las cosas (IOT) y referentes a la temática de gases.
- Artículos con soluciones que permitan detectar y monitorizar cuando ocurren fugas de gas.
- Artículos que empleen prototipos de bajo costo utilizando plataformas de hardware libre

Criterios de exclusión

Los criterios de exclusión definidos para este caso son los siguientes:

- Artículos que implementen únicamente soluciones basadas en IOT con pruebas en laboratorio sin resultados en ambientes reales.
- Artículos enfocados en la temática de IOT de forma general.
- Artículos que presenten una solución válida para controlar el tema de la fuga de gases que no sea replicable.

Grupo de control

La conformación del grupo de control es el punto de partida de la investigación ya que permite identificar una serie de estudios relacionados que servirán como línea base. El proceso empieza con una revisión básica de literatura, efectuada por cada uno de los investigadores, para este caso particular el maestrante y el tutor de proyecto de titulación, seleccionaron algunos artículos de fuentes fiables que se

encontraron muy relacionados con la temática de estudio, éstos fueron sometidos a una validación cruzada, obteniéndose como resultado aquellos trabajos que formaran parte del grupo de control.

En total se seleccionaron 4 artículos técnicos de control de la base de datos bibliográficos de IEEE (ver Tabla 2).

Tabla 2

Grupo de control

#	Artículos
S1	Joao B. A. Gomes, Joel J. P. C. Rodrigues, Jalal Al-Muhtadi, N Arunkumar, Ricardo A. L. Rabêlo and Vasco Furtado " An IoT-Based Smart Solution for Preventing Domestic CO and LPG Gas Accidents", 2018 IEEE 10th Latin-American Conference on Communications (LATINCOM) (Joao B. A. Gomes, 2018)
S2	Gabriel V. da Silva Medeiros ; Matheus Ricardo dos Santos ; Alba Sandyra Bezerra Lopes ; Edmilson C. B, "Smartgas: a smart platform for cooking gas monitoring" 2017 IEEE First Summer School on Smart Cities (S3C) (Medeiros, Santos, Lopes, & Edmilson C. B, 2017)
S3	Metta Santiputri, Muhammad Tio" IoT-based Gas Leak Detection Device", 2018 International Conference on Applied Engineering (ICAE) (Metta Santiputri, 2018)
S4	Shankey Garg, Jyotir Moy Chatterjee, Raghvendra Kumar Agrawal "Design of a Simple Gas Knob: 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)

Cadena de búsqueda y resultado del pilotaje

En base a la lectura de cada uno de los artículos del grupo de control, se seleccionaron palabras que permitieron armar una posible cadena de búsqueda, después de una validación cruzada entre los integrantes del grupo de trabajo se llegó a un consenso determinando la lista de términos clave (ver Tabla 3).

Tabla 3*Términos clave*

#	Referencias	Términos
S1	Joao B. A. Gomes, Joel J. P. C. Rodrigues, Jalal Al-Muhtadi, N Arunkumar, Ricardo A. L. Rabêlo and Vasco Furtado " An IoT-Based Smart Solution for Preventing Domestic CO and LPG Gas Accidents", 2018 IEEE 10th Latin-American Conference on Communications (LATINCOM) (Joao B. A. Gomes, 2018)	Gases, protocols, sensors, internet of Things, gas monitoring systems, detection system, gas Accidents
S2	Gabriel V. da Silva Medeiros ; Matheus Ricardo dos Santos ; Alba Sandrya Bezerra Lopes ; Edmilson C. B, "Smartgas: a smart platform for cooking gas monitoring" 2017 IEEE First Summer School on Smart Cities (S3C) (Medeiros, Santos, Lopes, & Edmilson C. B, 2017)	Intelligent support, real-time information, gas sensors
S3	Metta Santiputri, Muhammad Tio" IoT-based Gas Leak Detection Device", 2018 International Conference on Applied Engineering (ICAE) (Metta Santiputri, 2018)	Accidents
S4	Shankey Garg, Jyotir Moy Chatterjee, Raghvendra Kumar Agrawal "Design of a Simple Gas Knob: 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)	IoT, real-time, WiFi, Accidents, Internet of things, Smart phones, Embedded device

Se extrajeron 14 palabras luego del consenso en el grupo, dando como resultado la siguiente cadena de búsqueda para el pilotaje en IEEE (ver Tabla 4).

Tabla 4*Cadena de búsqueda*

#	Cadena	Art	Grupo Control
5	((gas sensors) OR (Gas Sensing) OR (Gas monitoring systems)) AND (detection system) AND (efficient) AND (iot OR Internet of Things) AND (Gas Accidents) AND (Gases) AND (WiFi) AND (Smart phones) AND (Embedded device)	60	S1, S2, S3, S4

De la cadena de búsqueda se encontraron 60 artículos dentro de los cuales constaban los 4 del grupo de control, al tener los títulos coherencia con el tema, se acepta la cadena de búsqueda como válida.

Selección de estudios

Se realizó la lectura de los títulos de los artículos, palabras clave y abstract, tomando en cuenta los criterios de inclusión y exclusión, se obtuvieron los estudios candidatos (ver Tabla 5).

Tabla 5

Resumen de estudios candidatos

Cantidad	Rechazados	Aceptados	Motivo
60	51	10	Los títulos, palabras clave y abstract no cumplen con los criterios de inclusión y exclusión

La mayoría de los artículos encontrados, son de la temática de IOT y sensores, mas no se enfocan en el tema de gases, por lo que fueron rechazados, quedando estudios bastante actuales, mayormente de Brasil e Indonesia.

Estudios primarios

De los resultados de los estudios candidatos, se descartaron los 4 artículos que correspondían a un tema de control, ya que no se encontró información relacionada con sensores de gas, pues la información era mas acorde con IOT (internet de las cosas). De esta manera se seleccionaron 6 artículos como estudios primarios, para este caso particular únicamente se añadieron 2 artículos más a aquellos que forman parte del grupo de control (ver Tabla 6).

Tabla 6

Estudios primarios

Título	Motivo
Design of a Simple Gas Knob: An Application of IoT	Relacionado al tema de gases
A WSN based system for CO2 concentration monitoring in large-scale petrochemical plants	Relacionado al tema de gases
An IoT-Based Smart Solution for Preventing Domestic CO and LPG Gas Accidents	Relacionado al tema de gases
Smartgas: a smart platform for cooking gas monitoring	Relacionado al tema de gases
Gas Leakage Detection Based on IOT	Relacionado al tema de gases

Titulo	Motivo
IoT-based Gas Leak Detection Device	Relacionado al tema de gases

Extracción de la información

En (Garg, Chatterjee, & KumarAgrawal, 2018) se construyó un dispositivo capaz de monitorizar el estado de estufas de gas para prevenir accidentes, utilizando componentes de bajo costo, el proyecto se basa en un dispositivo IOT que permite monitorizar remotamente si existen escapes de gas, a través de una aplicación móvil para Android, el proyecto se involucra en la temática que se buscó abordar, más no se presentan indicios de almacenamiento de información, análisis y recolección de datos histórico o incluso pronósticos a futuro sobre los mismo, fue un muy buen punto de partida para iniciar el trabajo de investigación.

El trabajo realizado por (Zhang, et al., 2015) menciona la utilización de la arquitectura de distribución de información basada en dispositivos que conforman una red bajo el protocolo Zigbee y Wireless Sensor Network (WSN), el trabajo se realizó para medir concentraciones de CO2 principalmente, utilizando un monitor dentro una aplicación para móviles, los esfuerzos realizados por este grupo de investigadores permitieron medir además la temperatura, la humedad y la dirección del viento, sin embargo, el trabajo solo se relaciona con la recolección y presentación de la información, más no se enuncia ninguna evidencia de análisis o tratamiento o predicción sobre el grupo de datos de entrada.

Sobre el trabajo realizado por (Joao B. A. Gomes, 2018) se presenta como medida de solución para la problemática derivada del escape de gas, la construcción de un sensor para IOT plug and play (de entrada y salida), que transmite datos utilizando el protocolo IEEE 802.15.4 para comunicarse con un router coordinador central, este artículo únicamente se enfoca en una solución que permite la recolección de información como base para resolver el problema, es un gran aporte para una

segunda fase donde posiblemente se pueda realizar análisis de la información o incluso predicciones sobre la misma.

En (Suma, Shekar, & Akshay, 2019) se presenta una solución para monitorizar específicamente las bombonas de gas, de esta manera se puede detectar la presión, la cantidad de contenido dentro de la misma, así como fugas en el interior, además la particularidad de este dispositivo es que envía un mensaje SMS al usuario cuando alguna alerta sucede y se integra con los servicios de la compañía de gas de la India para notificar que se requiere reabastecimiento; es un trabajo muy interesante controlar directamente los cilindros de gas, pero en países donde se ocupa varios cilindros para actividades principalmente de cocina, podría resultar muy costoso, además en ciertos países como Ecuador, las bombonas se reemplazan constantemente cuando se vacían, lo cual obligaría a reinstalar siempre el dispositivo. El proyecto tiene un enfoque direccionado a optimizar el reabastecimiento de gas, si bien es cierto, aunque se permite la detección de fugas, no se menciona ningún análisis de datos o recolección o almacenamiento de estos para a un futuro utilizar técnicas que permitan pronosticar estos valores.

En (Metta Santiputri, 2018) se analizaron los parámetros del estándar de diseño de las bombonas de gas en Brasil y su peligrosidad o riesgo de explosión, no se presentan parámetros que dictaminen cuando se debe o no emitir una alerta por fugas de gas, pues los algoritmos utilizados en este estudio se enfocan únicamente en una respuesta booleana que indique si se debe o no enviar una alerta sin ninguna guía o parametrización que determine rangos confiables, se debería perfeccionar este tema en base a información que brinde alguna organización gubernamental de regulación de concentración de gas en el aire.

Características del estado del arte

Para responder las preguntas de investigación fue necesario recabar la información necesaria de los pocos artículos que se encontraron para éste tema, respondiendo a OE2-PI3, los dispositivos mas utilizados en (Joao B. A. Gomes, 2018) y (Medeiros, Santos, Lopes, & Edmilson C. B, 2017) y (Garg, Chatterjee, & KumarAgrawal, 2018) son derivaciones de la distribución de Hardware Libre Arduino.

Para OE2-PI4 se encontraron apenas 6 estudios relacionados al tema de gases, éstos se enfocan en IOT (Internet de las cosas), pero ninguno hace referencia a ciencia de datos o algo similar, por tanto, el estudio es viable para su realización, tomando como punto de partida que en cada uno de los artículos se utilizaron sensores para capturar los datos.

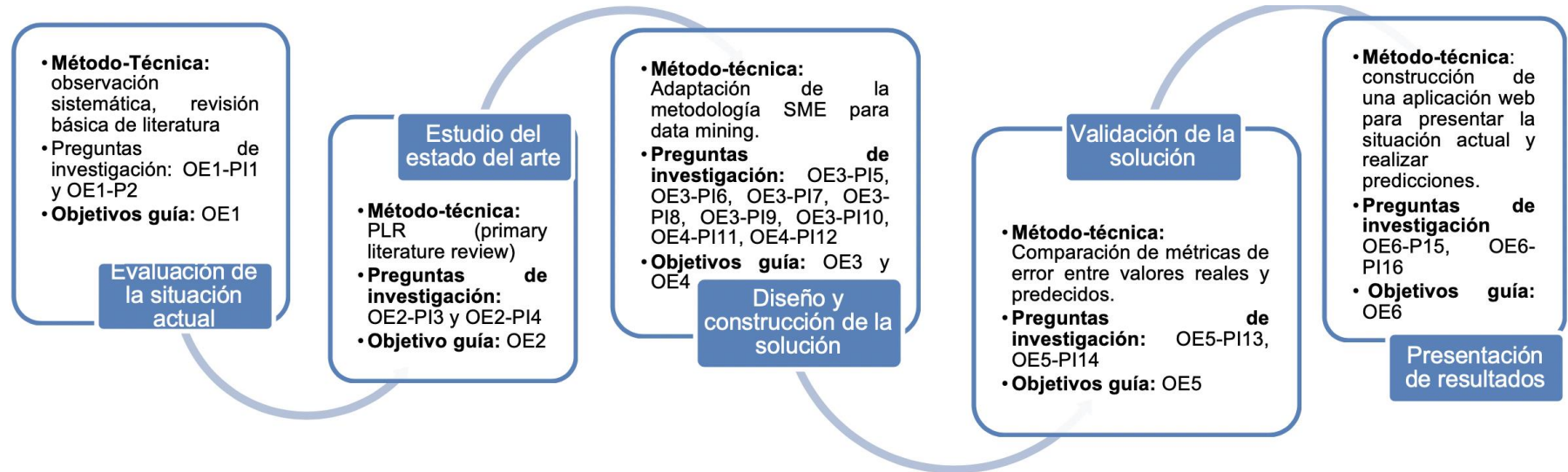
Metodología

El enfoque de investigación para el proyecto en curso se desarrolló en el marco cuantitativo ya que los datos fueron recolectados de un entorno real controlado que opera actualmente en la ciudad de Quito. Dado que el proyecto tiene componentes tanto de IOT como de minería de datos, fue necesario un diseño flexible adaptable al cambio, por lo que se utilizó una metodología AD-HOC propia derivada de: el marco teórico de procesos para SME propuesto por Dittert et, esta combina lo mejor de las metodologías de analítica de datos KDD-Process¹, SEMMA y CRISP-DM; para los elementos de IOT se añadirán algunas fases necesarias que permitan el prototipo rápido de la solución y la recolección de información necesaria para el análisis, de esta manera, se presenta en la Figura 6. Metodología AD-HOC.

¹ KDD-Process (Knowledge Data Discovery Process) metodología para la implementación de proyectos de inteligencia de negocios que consiste en nueve pasos los cuales son iterativos la cual no depende de una herramienta de software específica.

Figura 6

Metodología Ad-hoc



Las fases de la metodología planteada se describen a continuación:

- **Evaluación de la situación actual:**

La investigación parte de una observación sistemática y de una revisión de literatura básica; a partir de lo cual se planteó la problemática en torno a la acumulación de gases en espacios cerrados, se evaluaron sus causas y efectos mediante un diagrama de árbol de problemas, dando solución al OE1 y las preguntas de investigación OE1-PI1 y OE1-PI2.

- **Estudio de viabilidad de la solución:**

Se revisó aquello que otros autores a nivel del mundo han realizado acerca del tema, que aporte tuvieron y que fue mejorable en sus soluciones, determinando de esta manera la viabilidad del proyecto, se utilizó para esto una revisión de literatura inicial. En este punto detallado en el Capítulo II, se respondieron las preguntas de investigación OE2-PI3 y OE2-PI4 determinando la línea base del proyecto.

- **Diseño y construcción de la solución:**

En esta fase se realizó la implementación de la solución propuesta, utilizando en su gran mayoría partes de la metodología SME mencionada anteriormente, se dio solución al objetivo OE3 y OE4, los pasos que se cumplieron en este punto fueron:

- **Construcción del prototipo:**

Se realizó la implementación del nodo sensor capturador de información (NSCI) utilizando sistemas empotrados Hardware Libre en base a los parámetros definidos anteriormente, esta fase se realiza por medio de una metodología ad-hoc derivada de la metodología ágil de desarrollo software XP, aquí se dan solución a las preguntas de investigación OE3-PI5 y OE3-PI6.

- **Recolección y análisis de datos:**

Dado que en la fase de “**Construcción del prototipo**” se determina la estructura de los datos que se va a almacenar en la base NoSQL, no fue necesario un proceso de limpieza y depuración. A continuación, se realizó un proceso de descubrimiento de datos (Aguilar, 2013), que permite conocer el estado actual del escenario real de estudio, mediante el cual se determina que variables, de las que fueron consideradas al iniciar, son realmente representativas para la solución final, además se estableció una análisis de las posibles consecuencias para las personas en base a los datos recopilados para el experimento y de que manera se podrían mitigar, aquí se dieron solución a las preguntas de investigación OE3-PI7, OE3-PI8, OE3-PI9 y PE3-PI10.

- **Selección de algoritmos predictivos:**

Dependiendo de la tarea de BI a ejecutar se va a contar con varias técnicas de implementación como: clasificación, “Collaborative Filtering” (CF), reglas de asociación, clusterización, redes neuronales, arboles de decisión, regresión lineal, Smoreg (Lorenzo, 2008). Para esta etapa se realizó un proceso de evaluación de varios algoritmos predictivos utilizando herramientas de minería de datos en conjunto con la librería Tensorflow de Google. Luego de la evaluación de varios algoritmos se obtuvieron modelos en base al conjunto de datos inicial; aquí se respondieron las preguntas de investigación OE4-PI11 y OE4-PI12.

- **Validación de la solución:**

La evaluación de la solución se enfocó principalmente en determinar cuán buena o mala fue una predicción utilizando el grupo de datos de validación, de esta manera se buscó obtener un resultado muy similar a los datos de entrada. Un modelo fue aceptado como válido cuando la predicción fue lo más cercana al valor real tomado por el NSCI, para lo cual se realizó la evaluación de ciertas métricas representativas. En esta etapa se dio solución a las preguntas de investigación OE5-PI13, OE5P14 y por consiguiente al OE5.

- **Presentación de resultados:**

Una vez analizada la información y construido el modelo de la solución con la selección del algoritmo más preciso, se construyó un prototipo de aplicación para presentar de una manera más amigable los resultados al usuario final, esta etapa es orientada al desarrollo de software utilizando una metodología propia ad-hoc que tomó las mejores prácticas de la metodología Ágil XP, esta consta de las siguientes fases:

- Exploración:

En este paso se establecen las historias de usuario de la aplicación, se escriben en lenguaje natural y permiten orientar de una manera sencilla el desarrollo de las funcionalidades (ONess, 2020). En esta parte se da solución a la pregunta de investigación OE6-P15.

- Diseño:

Aquí se estructura un diagrama arquitectura de software de alto nivel, que incluye la distribución de la aplicación de frontend con los servicios Rest de las aplicaciones backend, se realiza un diagrama de casos de uso de la aplicación para describir las historias de usuario, un diagrama de secuencia de como opera la aplicación internamente en cada

proceso y las maquetas de la aplicación, aquí también se selecciona la información que se presentará al usuario final; se dio respuesta a la pregunta de investigación OE6-PI16 y por consiguiente al objetivo específico OE6.

- Codificación

Utilizando el lenguaje de programación JavaScript y el Framework de frontend React bajo el patrón Redux se codificó la aplicación de presentación de datos.

- Pruebas

Se realizaron pruebas de funcionalidad en la interfaz de usuario.

Capítulo 3

Marco teórico referencial

Del análisis de la hipótesis planteada en el Capítulo I se identifica la variable dependiente e independiente:

Variable Dependiente: fugas de gases en espacios.

Variable Independiente: técnicas de aprendizaje de automático.

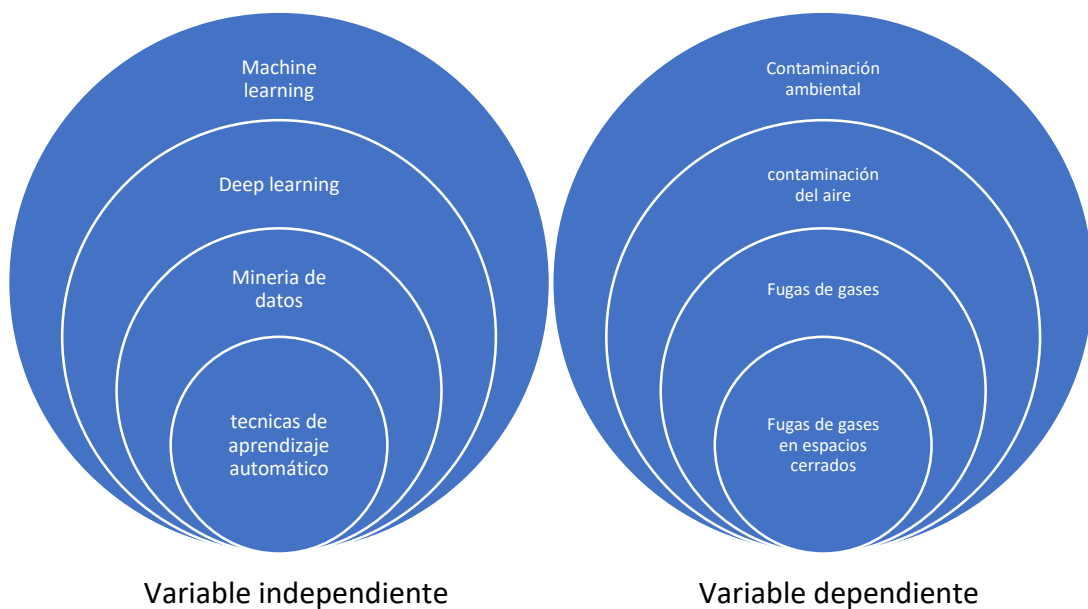
En base a estas variables y a su categorización se detalla el marco teórico referencial necesario para la realización del presente proyecto.

Categorización de las variables

En la Figura. 7, se representa la jerarquización de las variables dependiente e independientes.

Figura 7

Categorización de las variables



Fundamentación de la variable independiente

Machine learning

Se ocupa de las técnicas y algoritmos para realizar tareas que mejoren con la experiencia de la realización de tareas específicas. Las principales categorías son: clasificación, regresión, clasificación, entrenamiento, descubrimiento de relaciones, y aprendizaje reforzado (Gallagher, 2015).

Tiene como propósito proveer a los sistemas la capacidad de aprender sin ser programados de forma explícita, no todos los algoritmos de machine learning funcionan de manera eficiente para un mismo problema, por lo que es necesario hacer un estudio de los más adecuados (Lopez & Cadavid, 2016).

En la Figura 8 se presenta el ciclo de vida de machine learning que única con la adquisición de los datos a procesar, estos se someten a n proceso limpieza y calidad; después utilizando una herramienta se realizan los procesos de machine learning y al final presentamos informes con los descubrimientos realizados durante todo el ciclo de vida (Bugarin, 2019).

Figura 8

Ciclo de vida de machine learning

Ciclo de Vida Machine Learning



Nota. Tomado de (Bugarin, 2019)

Deep learning

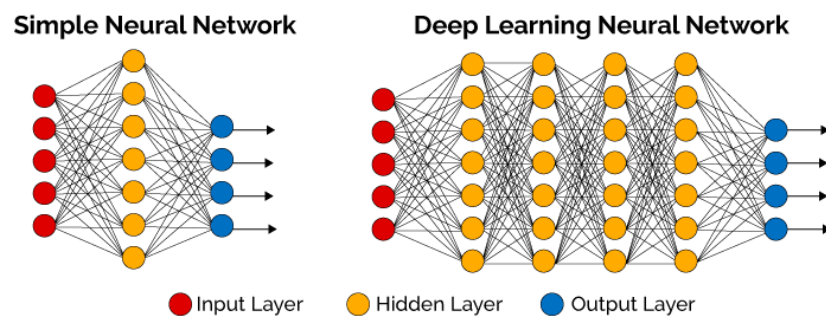
“Permite que los modelos computacionales compuestos de múltiples capas de procesamiento aprendan representaciones de datos con múltiples niveles de abstracción. Estos métodos han mejorado dramáticamente el estado del arte en reconocimiento de voz, reconocimiento visual de objetos, detección de objetos y muchos otros dominios como el descubrimiento de fármacos y la genómica.

Deep learning descubre una estructura compleja en grandes conjuntos de datos utilizando el algoritmo de propagación hacia atrás para indicar cómo una máquina debe cambiar sus parámetros internos que se utilizan para calcular la representación en cada capa a partir de la representación en la capa anterior” (Yann LeCun, 2015).

Como diferenciación de los procesos de Machine y Deep learning en la Figura 9 observamos que la diferencia mas notoria entre ambas redes son las capas ocultas de procesamiento que integran la red neuronal para deep learning.

Figura 9

Red neuronal deep learning



Nota. Tomado de (Vázquez, 2018)

Minería de datos

“Puede definirse como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos. La disponibilidad de grandes volúmenes de información y el uso generalizado de

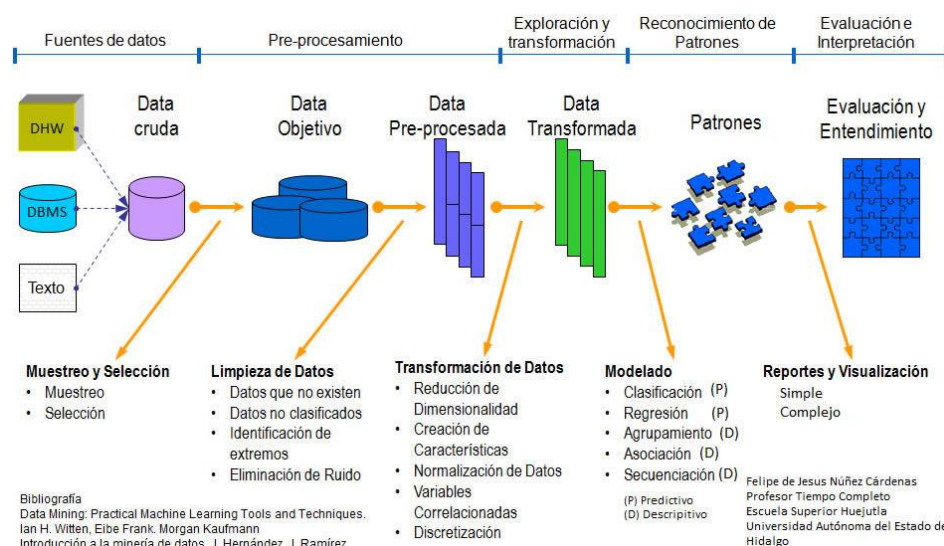
herramientas informáticas ha transformado el análisis de datos orientándolo hacia determinadas técnicas especializadas.

Las técnicas de minería de datos persiguen el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos. Estas técnicas tienen como objetivo descubrir patrones, perfiles y tendencias a través del análisis de datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas” (Lorenzo, 2008).

En la figura 10, se presenta el proceso usual utilizado para minería de datos, que inicia con la extracción de los datos de diversas fuentes de información como: bases de datos relacionales, no relacionales o archivos planos; con los datos obtenidos se deben seleccionar las fuentes de datos prioritarias, agruparlas y someterlas a un pre procesamiento que permita limpiar y estandarizar la información, una vez se cumple esta etapa los datos deben transformarse en nuevas estructuras para su presentación, aquí se pueden buscar patrones y asociaciones entre los mismos, para posteriormente presentar informes relacionados a este proceso (Cárdenas).

Figura 10

Proceso de minería de datos



Técnicas de aprendizaje automático

Las técnicas de aprendizaje automático se dividen en: técnicas de aprendizaje supervisado y técnicas de aprendizaje no supervisado. **Técnicas de aprendizaje supervisado:** se tiene variables de entrada y salida. Lo que se hace es ajustar las variables de entrada para predecir las variables de salida. **Técnicas de aprendizaje no supervisado:** no tienen una variable de salida para predecir, solo se tiene variables de entrada. En vez de ajustar el modelo a las variables de entrada para predecir la variable de salida, estas técnicas buscan descubrir patrones dentro de los volúmenes de información (Gorakala, 2015).

Fundamentación de la variable independiente:

Contaminación ambiental

“El crecimiento económico y la globalización han originado evidentes beneficios, pero al mismo tiempo han provocado la aparición de nuevos riesgos. Existen dificultades e incertidumbres para identificar con exactitud la relación causal entre medio ambiente y salud. La medición de la exposición a numerosos factores ambientales es compleja porque no disponemos de sistemas adecuados de información y vigilancia sanitaria que permitan valorar la magnitud y gravedad de los riesgos. La información disponible sobre las enfermedades relacionadas con el medio ambiente procede de la experimentación en animales, estudios de laboratorio, estudios epidemiológicos y toxicológicos. Los resultados de estos trabajos de investigación permiten extrapolar y estimar posibles riesgos para la salud pública. Sabemos, además, que algunas sustancias ambientales por debajo de ciertos niveles no son peligrosas.

Sin embargo, otros agentes, tales como alérgenos, radiaciones ionizantes, contaminantes del aire, preparados químicos carcinógenos, pueden suponer un riesgo a niveles más bajos de los observados. A pesar de ello existen algunos trabajos que han identificado la relación entre determinados agentes ambientales y la salud humana.

Las enfermedades respiratorias, el asma y las alergias están asociadas con la contaminación del aire externo e interno. La relación entre la contaminación atmosférica y la salud es cada día más conocida. Los agentes ambientales implicados son los óxidos de nitrógeno y azufre, las partículas en suspensión, ozono, metales, compuestos orgánicos volátiles (COV) y los hidrocarburos” (Marcos, n.d.).

La figura 11 presenta una imagen de la contaminación realizada por fabricas, explotaciones petroleras que contaminan el agua y en la parte inferior el pésimo tratamiento de la basura.

Figura 11

Contaminación ambiental



Nota. Tomado de (Portillo., 2020).

Contaminación del aire

“La contaminación representa uno de los conflictos ambientales más graves, principalmente en las zonas urbanas por el crecimiento acelerado de la población, el incremento del parque automotor y el desarrollo industrial, entre otros” (Ministerio de Ambiente, n.d.).

“La niebla tóxica que flota por encima de las ciudades es la forma de contaminación del aire más común y evidente. No obstante, existen diferentes tipos de contaminación, visibles e invisibles, que contribuyen al calentamiento global. Por lo general, se considera contaminación del aire a cualquier sustancia, introducida en la atmósfera por las personas, que tenga un efecto perjudicial sobre los seres vivos y el medio ambiente.

El dióxido de carbono, un gas de efecto invernadero, es el contaminante que está causando en mayor medida el calentamiento de la Tierra. Si bien todos los seres vivos emiten dióxido de carbono al respirar, éste se considera por lo general contaminante cuando se asocia con coches, aviones, centrales eléctricas y otras actividades humanas que requieren el uso de combustibles fósiles como la gasolina y el gas natural. Durante los últimos 150 años, estas actividades han enviado a la atmósfera una cantidad de dióxido de carbono suficiente para aumentar los niveles de éste por encima de donde habían estado durante cientos de miles de años.

Existen otros gases de efecto invernadero, como el metano (que proviene de fuentes como ciénagas y gases emitidos por el Ganado) y los clorofluorocarbonos (CFCs), que se utilizaban para refrigerantes hasta que se prohibieron por su efecto perjudicial sobre la capa de ozono de la Tierra” (nationalgeographic, 2010).

En la figura 12 se muestra a uno de los más grandes involucrados en la contaminación del aire, las plantas nucleares.

Figura 12*Contaminación del aire*

Nota. Tomado de (Uriarte, 2019).

Fugas de gas en el aire

“Es bien conocida la importancia que desempeña el oxígeno para los seres vivientes, importancia tal, que sin su presencia sería imposible la existencia de vida: el ser vivo toma el oxígeno del aire que le rodea, cuya composición, salvo leves oscilaciones, es del 21% de oxígeno, 78,1 % de nitrógeno, 0,9% de argón y pequeñas cantidades de otros gases como el anhídrido carbónico, ozono, etc.

Toda disminución sobre el citado porcentaje del 21 % de oxígeno, da lugar a la aparición de una atmósfera sub oxigenada con el consiguiente riesgo para el ser humano, situación que puede considerarse como peligrosa para concentraciones inferiores al 16% y que cuando desciende al 10%, el riesgo de asfixia mortal es casi cierto.

Los gases inertes son incoloros, inodoros e insípidos, por lo que su efecto asfixiante al desplazar al aire se produce sin ningún signo fisiológico preliminar que señale su presencia; en este sentido son por tanto mucho más peligrosos que gases tóxicos como el cloro, amoníaco, etc., de los que basta una pequeña concentración ambiental para que su olor característico y penetrante delaten su presencia.

En muchas instalaciones es corriente encontrar redes de distribución de gases inertes, como el nitrógeno, empleado para la neutralización y/o purga, simultáneamente a redes de aire comprimido utilizado en accionamiento de muy diversa maquinaria. Cualquier error en la conexión de uno u otro gas, puede dar lugar a crearse atmósferas peligrosas por sub oxigenación, teniendo en cuenta que muchas veces la descarga suele producirse "in situ", por lo que espacios cerrados como galerías subterráneas, cajas frías, almacenes, cámaras frigoríficas y espacio semi cerrados como: oficinas, locales comerciales, laboratorios, cocinas, bodegas de gas, entre otros están en riesgo de sufrir algún accidente" (Valle).

Fugas de gases en espacios cerrados

“Los riesgos y efectos en la salud no están distribuidos equitativamente en la población. Las personas con enfermedades previas, los niños menores de cinco años y los adultos entre 50 y 75 años son los más afectados. Las personas pobres y aquellas que viven en situación de vulnerabilidad, así como las mujeres y sus hijos que utilizan estufas tradicionales de biomasa para cocinar y calentarse, también corren mayor riesgo.

Hay efectos de la contaminación del aire sobre la salud a corto y largo plazo, siendo la exposición a largo plazo y de larga duración la más significativa para la salud pública. La mayoría de las muertes atribuibles a la contaminación en la población general están relacionadas con las enfermedades no transmisibles. En efecto, el 36% de las muertes por cáncer de pulmón, el 35% de la enfermedad pulmonar obstructiva crónica (COPD), el 34% de los accidentes cerebrovasculares y el 27% de las cardiopatías isquémicas son atribuibles a la contaminación atmosférica. Sin embargo, el mayor impacto es sobre la mortalidad infantil, ya que más de la mitad de las muertes de niños menores de 5 años por infecciones agudas de las vías respiratorias inferiores (ALRI) son debidas a partículas inhaladas por la contaminación del aire interior

producto del uso de combustibles sólidos, además de los riesgos derivados de la inhalación del GLP y gases inertes como el CO₂ y el CO, gran cantidad de accidentes por explosiones provocadas por la impericia y el desconocimiento son riesgos que se deben tener en cuenta" (.WHO Global air pollution Platform.)

Capítulo 4

Diseño y construcción de la solución

Se efectuó la construcción del prototipo utilizando componentes tanto hardware como software, en base a modelos de aprendizaje automático supervisado, detallado a continuación.

Construcción del prototipo

Para realizar el prototipo se efectuó una serie de pasos sistemáticos conforme a una metodología propia Ad-hoc de desarrollo de software derivado de la metodología ágil XP, los cuales se describen a continuación:

Requerimientos para el desarrollo de software

Para describir los requerimientos de software se utilizaron historias de usuario que “son descripciones, simples muy cortas y esquemáticas, que resumen la necesidad concreta de un usuario al utilizar un producto o servicio, así como la solución que la satisface. Su función principal es identificar problemas percibidos, proponer soluciones y estimar el esfuerzo que requieren implementar las ideas propuestas” (Hoc, 2017). A continuación, se presentan los requerimientos de la aplicación de recolección de información:

Tabla 7

Historia de usuario 1

Historia:	SB-NCSI-H1: Recolectar información de gases
Descripción	Como usuario final quiero que por medio de un dispositivo de bajo costo basado en hardware libre se pueda recolectar la información de los gases de un espacio cerrado, sea una bodega de gas, una cocina, o algo similar, para poder analizarlos posteriormente.
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Se debe construir un dispositivo en base a hardware libre y debe utilizar componentes de fácil adquisición en el mercado. • El dispositivo deberá contener sensores que permitan medir gases como, por ejemplo: monóxido de carbono (CO), gas licuado de petróleo (LPG), amoníaco, sulfuro y tolueno, entre otros. • El dispositivo debe contener una arquitectura de software que pueda ser fácilmente montable y desmontable según se requiera. • El dispositivo deberá con los datos obtenidos de los sensores emitir un flujo constante de información cada minuto por medio de un protocolo ligero que evite retrasos en la transmisión.
Precondiciones:	El dispositivo deberá ser configurado e instalado en un lugar cerrado.
Pos condiciones:	El dispositivo emitirá un flujo de datos constante a un puerto específico de un servidor.

Tabla 8

Historia de usuario 2

Historia:	SB-NCSI-H2: Almacenar información en una base de datos
Descripción	Como usuario final quiero que el flujo de información constante de la historia SB-NCSI-H1 sea guardada en una base de datos NoSQL en la nube, para poder acceder a ella a través de internet.
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Construir o utilizar una herramienta que permita capturar el flujo de información de la historia SB-NCSI-H1. • Determinar la estructura de almacenamiento de los datos. • Codificar o formatear los datos al formato antes de su envío. • Proveer al dispositivo de un acceso con autenticación y permisos de escritura a una base de datos NoSQL en la nube. • Enviar el flujo de datos del dispositivo a la nube a través de internet.
Precondiciones:	<ul style="list-style-type: none"> • Se requiere que el dispositivo este conectado a internet para realizar la transmisión de los datos. • Configurar los permisos de lectura y escritura al usuario de conexión a la base de datos en la nube.
Pos condiciones:	La base de datos contendrá la información de los sensores del dispositivo NCSI.

Tabla 9

Historia de usuario 3

Historia:	SB-NCSI-H3: Analizar información
Descripción	Como usuario final quiero que la información de la base de datos sea analizada utilizando indicadores estadísticos básicos como: media, mediana, moda, valor mínimo, máximo y desviación estándar; realizando de esta manera un proceso de descubriendo de datos, para determinar las características esenciales de los mismo y aceptar o descartar los datos almacenados en la historia SB-NCSI-H2.
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Construir una aplicación en base a patrones de diseño de software que permita analizar la información estadísticamente. • Iniciar el análisis utilizando herramientas de minería de datos para realizar el proceso de descubriendo respectivo. • Determinar que datos se aceptan y que datos se rechazan. • Determinar las características esenciales del grupo de datos.
Precondiciones:	<ul style="list-style-type: none"> • Los datos deben estar almacenados en la base de datos en la nube • Configurar los permisos de lectura al usuario de conexión a la base de datos en la nube.
Pos condiciones:	Se determinará las variables que son aceptas y posteriormente pronosticadas.

Tabla 10

Historia de usuario 4

Historia:	SB-NCSI-H4: Construir modelos de predicción
Descripción	Como usuario final quiero que de los datos aceptados en la historia SB-NCSI se realicen modelos para que a futuro se permita predecir valores futuros.
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Construir modelos de predicción basados en herramientas de minería de datos y aprendizaje automático. • Calcular métricas de error como error medio absoluto y absoluto porcentual entre los valores reales y pronosticados por cada uno de los modelos. • Verificar que el modelo seleccionado pueda ser utilizado a futuro por algún recurso externo como una aplicación web.
Precondiciones:	Se debe determinar los datos pronosticables
Pos condiciones:	El modelo deberá permitir el pronóstico de datos futuros

Tabla 11*Historia de usuario 5*

Historia:	SB-NCSI-H5: Validar modelos de predicción
Descripción	Como usuario final quiero que de los modelos creados en SB-NCSI-H4, se valide qué tan cercana es la predicción efectuada por cada uno de ellos para determinar modelos de solución para los datos de entrada
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Pronosticar los valores reales utilizados para la creación de los modelos y determinar sus errores, el modelo o modelos que menor error presente será el aceptado. • Pronosticar los valores reales de un conjunto de datos diferente al utilizado en la construcción del modelo y determinar los valores de las métricas de error.
Precondiciones:	Se debe construir los modelos de pronóstico
Pos condiciones:	El modelo deberá permitir el pronóstico de datos futuros

En base a estos requerimientos en los siguientes pasos se detalla la construcción del prototipo de la solución propuesta.

Componentes hardware

Para el hardware se seleccionó las plataformas de modelado rápido Arduino y Raspberry Pi, por su ranking obtenido en 2019 otorgado por tiendas electrónicas como SparkFun (fabricante y distribuidor de componentes de hardware a nivel internacional) (sparkfun, 2019) y Amazon (All3DP, 2020) siendo las mejores opciones en relación costo beneficio, para la construcción del prototipo de éste trabajo se utilizó las especificaciones estándar de cada dispositivo, el detalle de los componentes utilizados y sus características se presenta en la Tabla 12.

Tabla 12

Componentes y características del NSCI

MQ2 (sensor de metano y humo) / variable: measure_smoke, measure_lpg	
Voltaje de operación	5V DC
Rango de detección	300 a 10000 ppm
Gas característico	1000ppm, Isobutano
Resistencia de censado	1KΩ 50ppm Tolueno a 20KΩ in
Tiempo de Respuesta	≤ 10s
Tiempo de recuperación	≤ 30s
Temperatura de trabajo	-20 °C ~ +55 °C
Humedad	≤ 95% RH
Contenido de oxígeno ambiental	21%
MQ7 (sensor de monóxido de carbono) / variable: measure_co	
Voltaje de operación	5V DC
Gas característico	1000ppm, Monóxido de carbono
Resistencia de censado	33Ω + o – 5%
Tiempo de Respuesta	≤ 10s
Tiempo de recuperación	≤ 30s
Temperatura de trabajo	-20 °C ~ +50 °C
Humedad	≤ 95% RH
Contenido de oxígeno ambiental	21%
MQ135 (sensor de calidad aire) / variable: measure_air_qa	
Voltaje de operación	5V DC
Rango de detección	10 a 10000 ppm
Resistencia de censado	33Ω + o – 5%
Tiempo de Respuesta	≤ 10s
Gases involucrados	amoniac, sulfuro y tolueno
Tiempo de recuperación	≤ 30s
Temperatura de trabajo	-20 °C ~ +50 °C
Humedad	≤ 95% RH
Contenido de oxígeno ambiental	21%
DTH11(humedad y temperatura) / variables: measure_temperature, measure_humidity	
Voltaje de operación	3.5V to 5.5V
Rango de medición temperatura	0°C to 50°C
Rango de medición humedad	20% to 90%
Exactitud	±1°C and ±1%
Arduino Uno	
Microcontrolador	ATmega328
Voltaje Operativo	5v
Voltaje de Entrada (Recomendado)	7 – 12 v
Pines de Entradas/Salidas Digital	14 (De las cuales 6 son salidas PWM)
Pines de Entradas Análogas	6
Memoria Flash	32 KB (ATmega328) de los cuales 0,5 KB es usado por Bootloader
SRAM	2 KB (ATmega328)
EEPROM	1 KB (ATmega328)
Velocidad del Reloj	16 MHZ
Raspberry Pi 3 B+	
Procesador	Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64-bit SoC
Frecuencia de reloj	1,4 GHz
Memoria	1GB LPDDR2 SDRAM

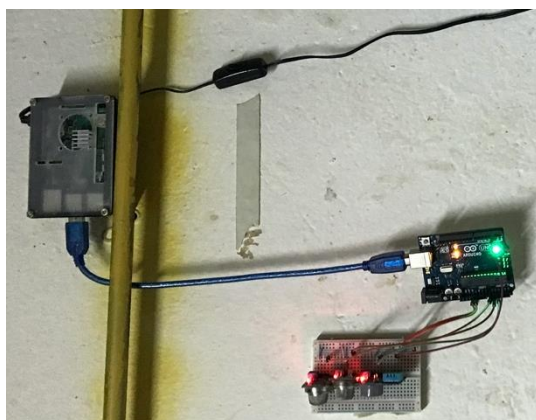
MQ2 (sensor de metano y humo) / variable: measure_smoke, measure_lpg	
Conectividad inalámbrica	2.4GHz / 5GHz IEEE 802.11.b/g/n/ac Bluetooth 4.2, BLE
Conectividad de red	Gigabit Ethernet over USB 2.0 (300 Mbps de máximo teórico)
Puertos	GPIO 40 pines HDMI 4 x USB 2.0 CSI (cámara Raspberry Pi) DSI (pantalla táctil) Toma auriculares / vídeo compuesto Micro SD Micro USB (alimentación) Power-over-Ethernet (PoE)

Nota. Referencias: (pololu) (PASTOR, 2018) (arduino)

Para el modelo físico del NSCI (ver en la Figura 13) se instalaron todos los componentes descritos en la tabla anterior, y se ubicó el mismo en una cocina semi industrial del bar de un colegio del sur de la ciudad de Quito.

Figura 13

Dispositivo NSCI



Se configuró el dispositivo para que realice la captura de la información cada minuto y efectuó la transmisión de los datos a un cluster de la base de datos MongoDB en la nube bajo el modelo base de datos como servicio (SBAAS) a través de Atlas y Amazon Web Services (aws); se utilizó la capa gratuita que consta de un arreglo de 3 nodos de base de datos sin respaldos auto programados y con un almacenamiento base de 512 Gb (MongoDb).

Herramientas de software

Para la recolección de información se utilizaron las herramientas de software descritas en la Tabla 13 en conjunto con el hardware enunciado en el punto anterior.

Tabla 13

Herramientas de software

Herramienta	Descripción
Arduino IDE	“El software Arduino de código abierto (IDE) hace que sea fácil escribir código y subirlo a la placa. Se ejecuta en Windows, Mac OS X y Linux. El entorno está escrito en Java y se basa en el procesamiento y otro software de código abierto.” (Arduino)
Mosquitto mqtt	“Eclipse Mosquitto es un agente de mensajes de código abierto (con licencia EPL / EDL) que implementa el protocolo MQTT versiones 5.0, 3.1.1 y 3.1. Mosquitto es liviano y es adecuado para usar en todos los dispositivos, desde computadoras de una sola placa de baja potencia hasta servidores completos. El protocolo MQTT proporciona un método liviano para llevar a cabo mensajes utilizando un modelo de publicación / suscripción.” (Foundation)
Node red	“Node-RED es una herramienta de programación para conectar dispositivos de hardware, API y servicios en línea de formas nuevas e interesantes. Proporciona un editor basado en navegador que facilita la conexión de flujos utilizando la amplia gama de nodos en la paleta que se pueden implementar en tiempo de ejecución.” (OpenJS)
Docker	“Un contenedor es una unidad de software estándar que empaqueta el código y todas sus dependencias para que la aplicación se ejecute de manera rápida y confiable de un entorno informático a otro. Una imagen de contenedor de Docker es un paquete de software ligero, independiente y ejecutable que incluye todo lo necesario para ejecutar una aplicación: código, tiempo de ejecución, herramientas del sistema, bibliotecas y configuraciones del sistema.” (Docker, 2020)

Herramienta	Descripción
Raspbian	Raspbian es un sistema operativo gratuito basado en Debian optimizado para el hardware Raspberry Pi. Raspbian proporciona más que un sistema operativo puro: viene con más de 35,000 paquetes, software pre compilado incluido en un formato agradable para una fácil instalación en Raspberry Pi. (aspbian, 2020)
MongoDB/Atlas	MongoDB es una base de datos distribuida, basada en documentos y de uso general que ha sido diseñada para desarrolladores de aplicaciones modernas, posee una versión en la nube bajo el nombre e MongoDB Atlas, sus clústeres de recuperación automática están formados por instancias de bases de datos distribuidas geográficamente para garantizar que no haya un solo punto de falla, también incluye potentes funciones para mejorar la confiabilidad de sus bases de datos de producción de misión crítica, tales como respaldos continuos y recuperación en un punto en el tiempo.

La selección de estas herramientas de software permite cumplir los requerimientos planteados en las historias de usuario, ya que al utilizar Docker establecemos una serie de contenedores independientes para las distintas aplicaciones que se manejan dentro de cada subsistema, de esta manera configurar un nuevo dispositivo es muchos más sencillo. Como la velocidad de transmisión de datos es importante se utilizó el protocolo MQTT como se menciona anteriormente, ya que este permite trabajar con flujos de datos comprimidos y ligeros a través del método de suscripción/publicación (Foundation).

Diseño

La etapa de diseño corresponde a la estructura de alto nivel de los componentes de la solución propuesta, a continuación, se detallará la arquitectura de software.

Arquitectura

En base a las herramientas y las historias de usuario descritas anteriormente se estructuró la arquitectura de software de alto nivel (ver Figura 15) donde, el proceso

de recolección inicia con el **Dispositivo recolector de datos** aquí se encuentra el programa elaborado en Arduino que permite: calibrar los sensores, obtener la lectura en crudo de la información (sensores MQ7, MQ135 y MQ2), pasar la lectura analógica a partes por millón(ppm) (sensores MQ7 y MQ2), se tomaron 2 cifras significativas como referencia para éste estudio, por lo que todo aquello que este por debajo de este rango será interpretado como 0. Toda la información es transmitida a la Raspberry Pi mediante uno de sus cuatro puertos serie, aquí interviene el **Python middleware**, que se encarga de decodificar los datos del puerto y enviarlos al **Bróker de mensajería Mosquitto**, el cual publicará el flujo de datos constante por medio de un contexto predefinido en el puerto 8883, se utiliza éste específicamente porque es un puerto nativamente encriptado mediante SSL/TLS por el bróker; el **Dispositivo transmisor de información** se encarga de leer el contexto publicado, decodificar y formatear los datos a una estructura JavaScript Object Notation (JSON) (ver figura 14) y enviarla por medio de una API de mongodb hacia la base de datos en la nube, de ésta manera se puede acceder directamente desde otra aplicación a los datos para realizar el proceso de análisis correspondiente.

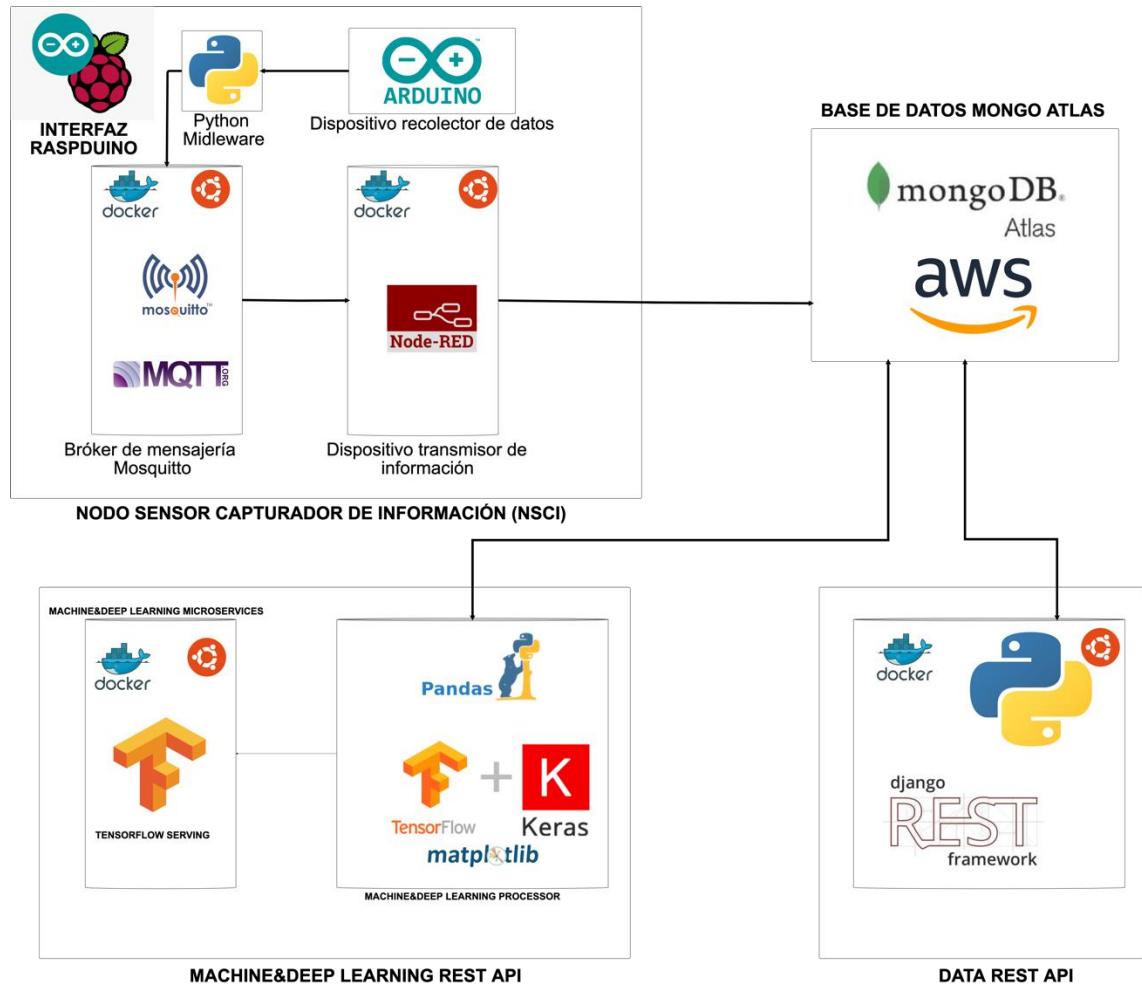
Figura 14

Modelo de almacenamiento de datos

```
{
  "measure_air_ga": "376",
  "measure_co": "15.5",
  "measure_humidity": "39",
  "measure_lgp": "0",
  "measure_smoke": "0",
  "measure_temperature": "26.5",
  "measure_timestamp": "1576628906"
}
```

Figura 15

Arquitectura NCSI



De esta manera los datos son almacenados y cuentan con alta disponibilidad para que cualquier aplicación pueda utilizarlos.

Diseño de casos de uso

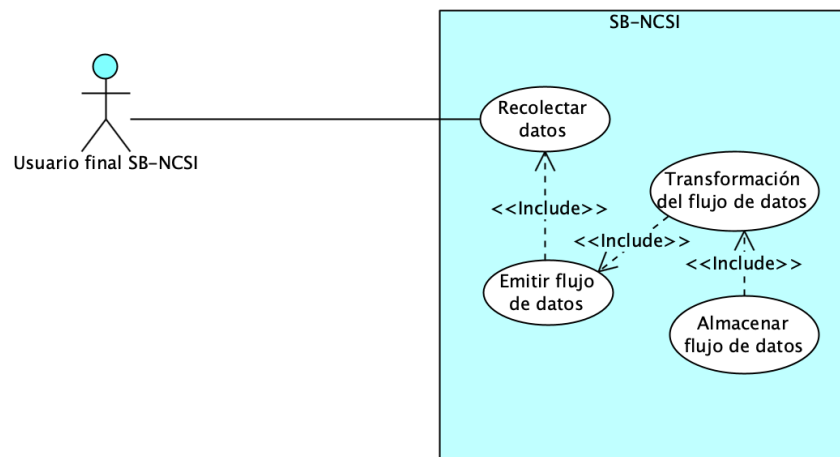
En el diagrama de arquitectura se presentan 3 subsistemas SB-NCSI (subsistema nodo sensor captador de información), SB-DRA (subsistema data rest API) y SB-MDRA (subsistema machine deep learning rest API), por medio de diagramas de caso de uso se ha determinado la operatividad de cada uno de ellos, estableciendo las acciones que se desencadenan con cada ejecución; este diagrama nos permite

determinar la relación directa entre el o los actores y las operaciones del sistema, se compone de los siguientes elementos: **actores** (rol del usuario que interactúa con el sistema), representado por medio de un monigote pequeño, **casos de uso** (operación o tarea específica del sistema), se representa por un elipse y **relaciones** (determina la interacción entre los componentes del diagrama), de estas últimas existen varios tipos como: **asociaciones** es la invocación de un proceso por parte de un actor, **generalización** representa la relación únicamente entre casos de uso y es de dos tipos **include** cuando la relación entre dos casos de uso es dependiente y **extends** cuando un caso de uso deriva de otro, las relaciones, se representan usualmente con líneas (openclassrooms, 2018) (stadium.unad.edu.co) .

Los casos de uso que se involucran en el SB-NCI (ver Figura 16) son, **recolectar datos**: con los sensores se recolectan los datos a través de un Arduino y se transmiten por medio de un puerto serie a la Raspberry Pi, aquí los datos se interpretan y decodifican; **emitir flujo de datos**: utilizando un bróker, los datos recolectados por los sensores se emiten y publican, este caso de uso tiene una relación include con el caso de uso anterior; **transformación del flujo de datos**: con la información publicada, se realizó un proceso de codificación de los datos a un formato legible por la base de datos y se envió los mismo para el guardado posterior, este caso de uso posee un relación include con el caso de uso anterior; **almacenar flujo de datos**: con los datos formateados se procede a almacenarlos en la base de datos MongoDB Atlas.

Figura 16

Casos de uso SB-NCSI



Para el subsistema SB-DRA los casos de uso (ver Figura 17) que intervienen con el manejo de la información son, **recuperar flujo de datos**: se definen las interacciones contra la base de datos utilizando una herramienta ORM(Object Relational Mapping)² y el patrón de diseño Data Access Object (DAO)³ para recuperar la información que se requiera; **analizar datos**: utilizando el patrón de diseño de arquitectura de software Modelo Vista Controlador (MVC) se realizó cada una de las

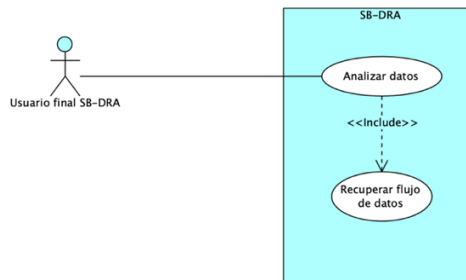
² **Herramientas ORM**: “es un modelo de programación cuya misión es transformar las tablas de una base de datos de forma que las tareas básicas, que realizan los programadores, estén simplificadas. Al hacer esto, se puede ahorrar mucho tiempo, ya que cada vez la información es mayor y más cambiante.” (Merino, 2018)

³ **DAO**: consiste en utilizar un objeto de acceso a datos para abstraer y encapsular todos los accesos a la fuente de datos. El DAO maneja la conexión con la fuente de datos para obtener y almacenar datos.

operaciones para el proceso de descubrimiento de datos y calculo de los indicadores estadísticos básicos, este caso de uso tiene una relación *include* con el anterior.

Figura 17

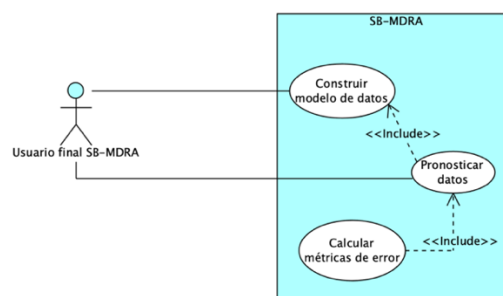
Casos de uso SB-DRA



En el subsistema SB-MDRA se describen los siguientes casos de uso, **construir modelo de datos**: a través de herramientas de minería de datos y algoritmos de aprendizaje automático se construyeron modelos de predicción utilizando los datos almacenados; **pronosticar datos**: con los modelos establecidos anteriormente se crearon funciones que realizan el calculo del siguiente valor de pronóstico, este caso de uso tiene una relación include con el anterior; **calcular métricas de error**: con los datos de pronóstico y los datos reales aquí se obtiene el valor de las métricas de error, este caso de uso tiene una relación include con el anterior.

Figura 18

Casos de uso SB-MDRA



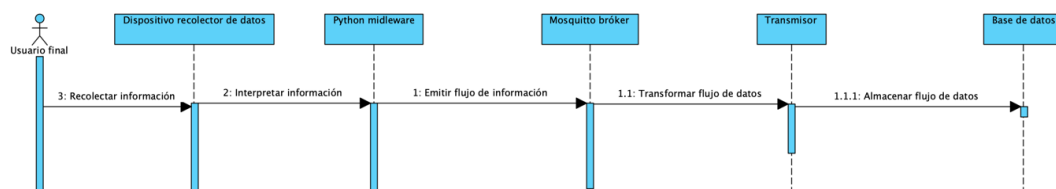
Una vez establecidos las funciones de cada uno de los subsistemas, a continuación, se representarán por medio de un diagrama de secuencia las interacciones entre éstos.

Diseño secuencias de la aplicación

Para el subsistema SB-NCSI en la Figura 19 se detallan la interacción de los componentes internos de la aplicación (ver figura 14) y las acciones presentadas en la sección anterior.

Figura 19

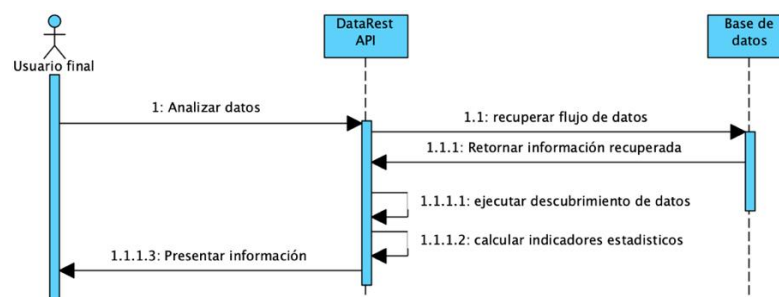
Diagrama de secuencia SB-NCSI



De la misma manera para el subsistema SB-DRA en la Figura 20 se detalla la comunicación entre los componentes de la aplicación (ver figura 14), los casos de uso presentados anteriormente (ver Figura 17) y los procedimientos internos dentro de cada componente como: *ejecutar el proceso de descubrimiento de datos y calcular indicadores estadísticos*, una vez recuperada la información desde la base de datos.

Figura 20

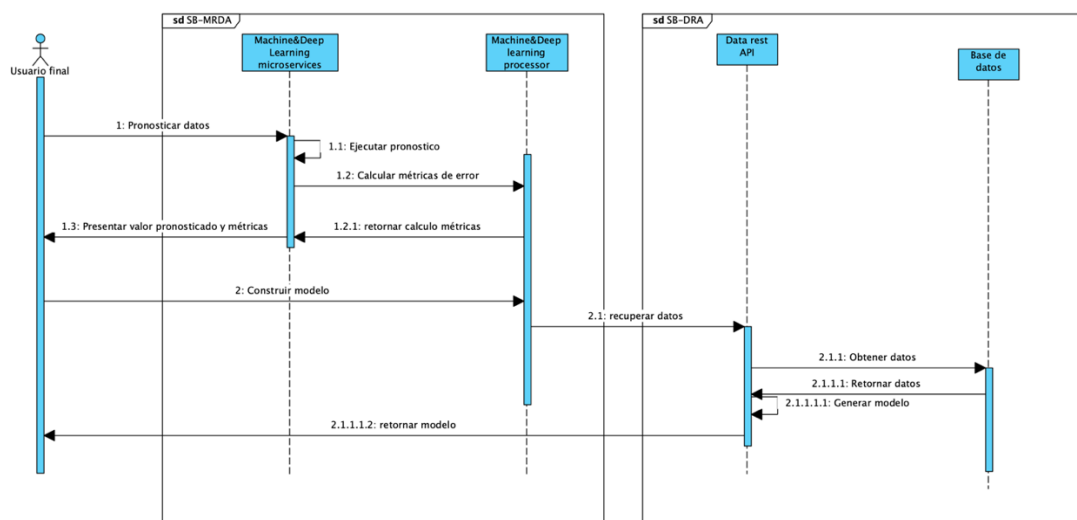
Diagrama de secuencia SB-DRA



Para el subsistema SB-MRDA, se representan las relaciones entre los componentes y las acciones (ver Figura 21), también se detalla la secuencia de eventos necesarios para completar estos procesos, que en este caso particular requiere de la interacción de otros componentes externos como el subsistema SB-DRA.

Figura 21

Diagrama de secuencia SB-MRDA



Construcción de la aplicación

Para efectuar la construcción de la solución se emplearon tecnologías como la herramienta Jupyter (“es un entorno de desarrollo interactivo basado en la web para portátiles programar mediante un cuaderno de trabajo” (Jupyter, 2020)) y la versión community de Rapid Minner que es una herramienta de minería de datos bastante utilizada, según (Gartner, 2020) ésta se ha posicionado este año en el cuadrante de los visionarios, lo cual lo vuelve una opción competitiva frente a los líderes, cuyo más grande problema es la difícil accesibilidad a una versión de prueba o community para estudiantes; además se utilizó el lenguaje de programación Python, el framework Django que provee de una herramienta ORM para la base de datos MongoDB, la plataforma especializada para machine learning y ciencia de datos Tensorflow

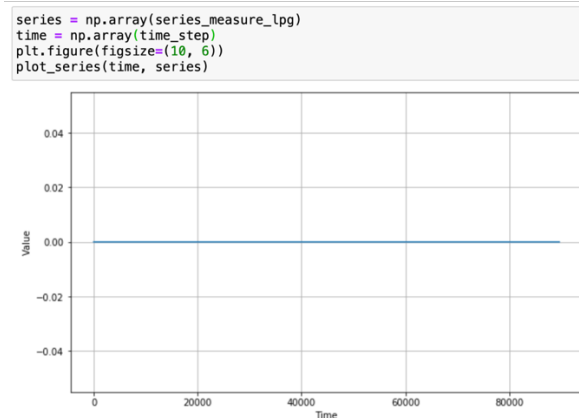
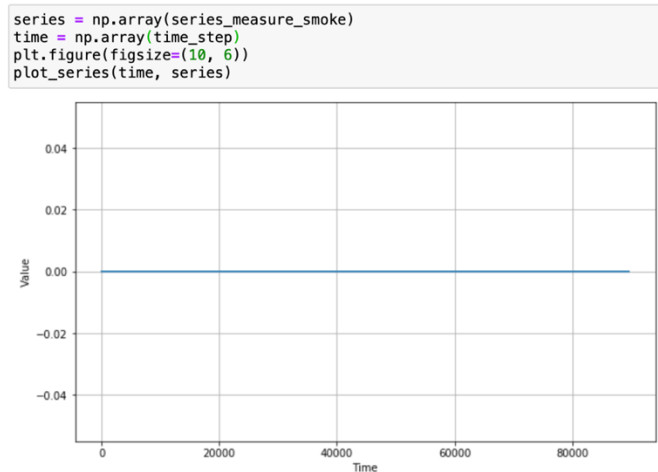
(tensorflow, 2020) y la librería matplotlib de Python que permite realizar gráficos de los datos obtenidos. Para encapsular cada uno de los componentes de los subsistemas se utilizó Docker y el diseño de la aplicación se basó en la metodología TDD⁴ (desarrollo dirigido por tests).

Recolección y análisis de datos

Se determinó que el grupo de datos a analizar corresponde a una serie de tiempo, definida como “una secuencia numérica de puntos de datos en orden sucesivo” (Seo, 201806). Los datos obtenidos a través de los sensores son consecutivos en el tiempo desde su inicio el 15 de noviembre de 2019 hasta el corte realizado el 20 de febrero de 2020. Para efectuar este trabajo, se realizó el 17 de enero de 2020 una primera extracción de datos a un repositorio diferente para comenzar el análisis, el volumen de información hasta este punto es de 89.611 registros. La validación de los resultados de los modelos se realizó en el grupo de datos restante desde el 17 de enero de 2020 hasta el 20 de febrero con un volumen de 17.409 registros.

En la Figura 22, 23 se presentan los gráficos en el tiempo de los datos obtenidos correspondientes a las variables de gas licuado de petróleo (lpg) y humo (smoke).

⁴ **TDD**: “es una práctica de programación que consiste en escribir primero las pruebas (generalmente unitarias), después escribir el código fuente que pase la prueba satisfactoriamente y, por último, refactorizar el código escrito” (Herranz, 2011).

Figura 22*GLP en el tiempo***Figura 23***Humo en el tiempo*

Al analizar las Figuras 22 y 23 tenemos que los valores de gas licuado de petróleo y humo dan como resultado un valor de 0 ppm en el tiempo, después de seleccionadas mas 89K muestras de la base de datos; en el apartado 4.1.3. Herramientas de software, se mencionó que la precisión de la lectura luego de la transformación a partes por millón era de 2 decimales y con un rango de oscilación de 300-1000ppm según la documentación técnica del sensor MQ2, por lo que, al presentar un valor extremadamente bajo, estas mediciones son 0 por lo que no tiene ningún sentido continuar su análisis; se decidió descartar estas 2 variables para la posteridad del estudio.

El valor de la señal analógica (bits) de los sensores de la serie MQ obtenida desde Arduino, requiere una transformación a un valor de voltaje (V) entre 0 y 1023 bits (VALLA, 2018), para lo que se utilizó la siguiente ecuación:

$$(1) \text{ sensor}_{volt} = \frac{sv(\text{bits}) * 5(v)}{1023(v)} = \frac{5sv}{1023} [v]$$

Donde sv (bits) corresponde al valor de la señal analógica, 5(v) es el voltaje de alimentación del sensor y 1023 (bits) es el valor al que queremos escalar. En las hojas técnicas de las 2 variables restantes de gases que nos quedan **calidad de aire** y **monóxido de carbono** (sparkfun) (sparkfun-MQ7), se detalla que para realizar la transformación del valor de sensor_{volt} a partes por millón⁵ (PPM), se requiere realizar una ecuación potencial de escalamiento en base a los valores predeterminados por el fabricante (VALLA, 2018), las ecuaciones para: amoníaco, tolueno y monóxido de carbono se detallan a continuación:

$$(2) \text{ ppm}_{NH_3} = 102.79x^{-2.424}; x = \frac{RS}{Ro}; \frac{RS}{Ro} = 0.77 \text{ (VALLA, 2018)}$$

$$(3) \text{ ppm}_{C_7H_8} = 50.286x^{-2.942}; x = \frac{RS}{Ro}; \frac{RS}{Ro} = 0.62$$

$$(4) \text{ ppm}_{CO} = 42.368x^{-1.099}; x = \frac{RS}{Ro}; \frac{RS}{Ro} = 0.09 \text{ (VALLA, 2018)}$$

$$(5) RS = RL * \frac{5(v) * \text{sensor}_{volt}}{\text{sensor}_{volt}}; RL = 1k\Omega \text{ (VALLA, 2018)}$$

Donde RS es la resistencia del sensor a la concentración de varios gases y RO para NH₃ y C₇H₈ significa la resistencia del sensor a 100 ppm de amoníaco (NH₃) en aire limpio y para CO significa resistencia del sensor a 100 ppm de CO en aire limpio.

Estas ecuaciones permitirán determinar los valores reales de las mediciones posteriormente. De las muestras tomadas para estos gases se realizó un análisis por

⁵ PPM: es una unidad de medida de concentración que mide la cantidad de unidades de sustancia que hay por cada millón de unidades del conjunto (medicina.usac.edu.gt, 2005)

métodos tradicionales de series temporales utilizando los factores de tendencia, estacionalidad y aleatoriedad en la distribución de los datos (Rivero, 2016), detalladas a continuación.

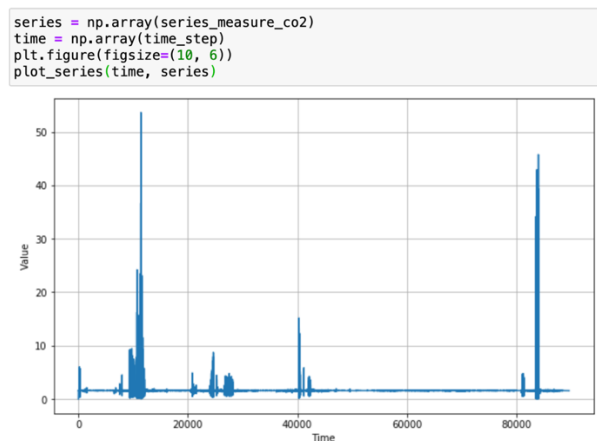
Monóxido de carbono (CO)

Se tomaron alrededor de 89K muestras mediante el NSCI, y se realizó la conversión a partes por millón utilizando las ecuaciones detalladas en el apartado 4.2 para el dióxido de carbono (ver Figura 24), donde se observan las siguientes características:

- **Estacionalidad:** existe un componente no lineal que cambia con el tiempo y no se repite, no se aprecian patrones definidos claramente, encontramos también un movimiento periódico que oscila en un rango pequeño definido entre 0 y 54ppm (Seo, 201806).
- **Ruido:** existe un componente no sistemático que no es tendencia ni estacionalidad dentro de los datos, no se refleja una curva suavizada y uniforme sino a manera de líneas consecutivas que suben y bajan (Seo, 201806).

Figura 24

Monóxido de carbono en el tiempo



La mayor concentración se encuentra ente 0 y 5.5ppm con un 89% de los valores, el resto de los datos se hallan distribuido en el intervalo]5.5; 54[(ver Figura 25).

Figura 25

Distribución de valores para CO



Para el pronóstico, se puede utilizar únicamente el histórico de la variable a predecir o el histórico de la variable a predecir y el resto de las variables que intervienen en el estudio. En este caso particular, podemos utilizar las variables: temperatura, humedad y calidad de aire para pronosticar este gas, dado que su correlación es bastante baja (ver Figura 26).

Figura 26

Correlación de las variables para la predicción del CO

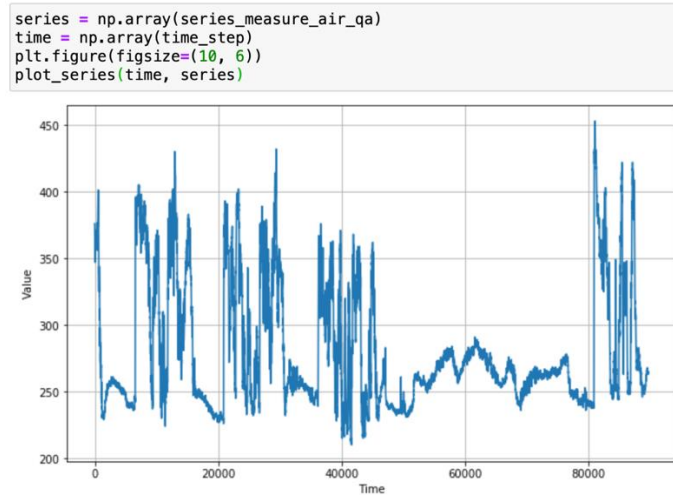
Quality	Name	Correlation
C I S M T	measure_air_qa	0.03%
C I S M T	measure_humidity	0.23%
C I S M T	measure_temperature	0.03%

La correlación máxima es del 0.23% lo cual es bastante bajo y permite realizar la predicción sin problemas, pues según la definición de correlación: un valor que tiende a 0 corresponde a variables independiente es decir no correlacionadas y sin asociación directa, una correlación muy elevada o medianamente elevada podría aportar a un sesgo que se limitaría únicamente a los datos de entrenamiento del modelo predictivo (Pedroza & Diovsky, 2006).

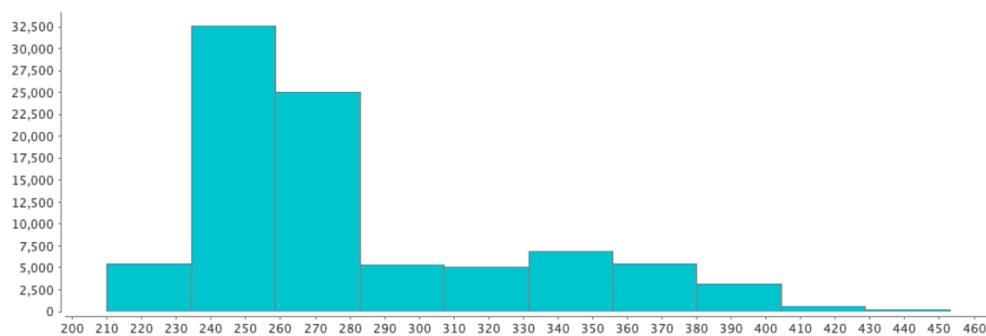
Calidad de aire

El sensor de calidad de aire aglutina varios gases, por lo que para un mejor análisis se tomó la medida directa del sensor (valor de la señal analógica en bits) (ver Tabla 12), para posteriormente realizar las transformaciones para cada uno de los gases. Se tomaron alrededor de 89K muestras mediante el NSCI (ver Figura 27), que presentan las siguientes características:

- **Ruido:** existe un componente no sistemático que no es tendencia o estacionalidad dentro de los datos, no se refleja una curva suavizada y uniforme sino a manera de líneas consecutivas que suben y bajan (Seo, 201806).
- **Aleatoriedad:** se encuentra también un componente no lineal que cambia con el tiempo y se repite en ocasiones, se pueden apreciar formas de W en ciertos instantes de tiempo, además existen movimientos erráticos que no persiguen un patrón específico (Rivero, 2016).

Figura 27*Calidad de aire en el tiempo*


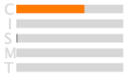
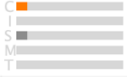

En la Figura 28 encontramos que los datos oscilan en un intervalo entre [210; 456], la mayor concentración de datos se encuentra entre los valores de [235; 258] que representan el 32.27% del valor total, seguido por un 27.90% comprendido entre [258; 282], en adelante la cifra mas alta esta entre [331; 356] con un 8.37%. En el gráfico también se presenta un ligero sesgo a la derecha.

Figura 28*Distribución de valores para calidad de aire*

Para las predicciones de este caso particular también fue analizada la correlación con las variables: temperatura, humedad y CO (ver Figura 29).

Figura 29

Correlación de las variables de la calidad de aire

Quality	Name	Correlation
	measure_timestamp	2.05%
	measure_temperature	63.16%
	measure_humidity	10.06%
	measure_co	0.03%

El valor máximo es del 63.16% para la variable de temperatura lo cual es bastante alto y no es recomendable incluirla dentro del modelo predictivo, para el caso de la humedad se tiene un 10.06% de correlación, el cual no es relativamente alto y podría incluirse en la predicción final. La variable CO es completamente independiente permitiendo incluirla en la realización de la predicción sin problemas (Pedroza & Diovsky, 2006).

Mitigación de riesgos para la salud en base a los resultados

En base a los datos obtenidos para la calidad de aire y monóxido de carbono a continuación se realizará un análisis en función de los valores en PPM, para establecer si la situación actual del escenario analizado en el estudio es apta para que las personas puedan desempeñarse con seguridad.

Monóxido de carbono (CO)

El CO es un gas inodoro, incoloro, inflamable y altamente tóxico, puede causar la muerte cuando se inhala en niveles muy elevados, es producto de la combustión de carburantes como el gas, la gasolina, el keroseno, el carbón y chimeneas, calderas, calentadores de agua y aparatos domésticos como estufas a gas puede producirlo cuando no están funcionando correctamente (Consejería de Salud de la Región de

Murcia). A continuación, en la Figura 30 se presenta un cuadro con las concentraciones y consecuencias derivadas a la exposición de este gas, considerado como base 8.7 ppm.

Figura 30

Concentraciones y efectos del CO

Concentración de monóxido de carbono	Efecto
0-229 mg/m ³ (0-200 ppm)	Ligero dolor de cabeza en algunos casos
10 mg/m ³ (8,7 ppm)	No se excede el nivel carboxihemoglobina del 2.5 %, aun cuando un sujeto normal realice ejercicio ligero o moderado durante 8 horas
30 mg/m ³ (26 ppm)	No se excede el nivel carboxihemoglobina del 2.5 %, aun cuando un sujeto normal realice ejercicio ligero o moderado durante una hora
34,4 mg/m ³ (30 ppm)	La exposición diaria a esta concentración es equivalente a fumar 20 cigarrillos al día
40,1 mg/m ³ (35 ppm)	Las personas que tienen enfermedades cardíacas no deben exponerse a niveles superiores a esta concentración
60 mg/m ³ (52 ppm)	No se excede el nivel carboxihemoglobina del 2.5 %, aun cuando un sujeto normal realice ejercicio ligero o moderado durante 30 minutos
100 mg/m ³ (87 ppm)	No se excede el nivel carboxihemoglobina del 2.5 %, aun cuando un sujeto normal realice ejercicio ligero o moderado durante 15 minutos
115 mg/m ³ (100 ppm)	Se informó del primer indicio de angina en sujetos que hacían ejercicio con cardiopatía coronaria expuestos a esta concentración
229-458 mg/m ³ (200-400 ppm)	Después de 5-6 horas se puede observar un leve dolor de cabeza, náuseas, vértigo y síntomas mentales
458-802 mg/m ³ (400-700 ppm)	Después de 4-5 horas se puede observar un fuerte dolor de cabeza, incoordinación muscular, debilidad, vómitos y colapso
802-1260 mg/m ³ (700-1100 ppm)	Después de 3-5 horas se puede observar un fuerte dolor de cabeza, debilidad, vómitos y colapso
1260-1832 mg/m ³ (1100-1600 ppm)	Después de 1.5-3 horas se puede observar coma. (la respiración es aún bastante buena a no ser que el envenenamiento se haya prolongado)
1832-2290 mg/m ³ (1600-2000 ppm)	Después de 1-1.5 horas hay posibilidad de muerte
5726-11452 mg/m ³ (5000-10000 ppm)	Después de 2-15 minutos se puede producir la muerte

Nota. Tomado de (*Consejería de Salud de la Región de Murcia*).

Como se menciona anteriormente (ver sección 4.2.1) para el “CO” el 89% de las mediciones corresponde al rango entre [0 y 5.5]ppm, lo que podría causar ligeros dolores de cabeza, pero no se muestran consecuencias más graves a la salud.

El 11% restante de las mediciones que oscila entre los valores comprendidos entre]5.5 ; 54[, equivalente a fumar 20 cigarrillos al día (concentraciones mayores a 30ppm), por lo que se recomienda realizar actividad física por lo menos durante 30 minutos al día como prevención; para corroborar esta información en (Grand Island Fire Department) se menciona que dolores de cabeza y mareos producto de concentraciones inferiores a 100ppm solo se presentan cuando los sujetos se encuentran de 6 a 8 horas en exposición constante. Por lo tanto, en base a los datos tomados se concluye que el lugar donde se tomó las muestras para este experimento

no presenta problemas de seguridad que pudieran ocasionar riesgos a la salud, sin embargo, se recomienda la monitorización de este por los valores atípicos encontrados durante el experimento.

Amoniaco, sensor de calidad de aire

“El amoniaco es un gas de altísimo riesgo para la salud porque es corrosivo para la piel, los ojos y los pulmones. La exposición a solo 300 partes por millón (ppm) es peligrosa para la vida y la salud. Adicionalmente el amoniaco es inflamable en concentraciones entre 15% y 28% del volumen” (MSA de CHILE).

En la Figura 31 se describen los límites establecidos por diversas organizaciones internacionales para la exposición en límites ponderados (8 horas diarias, 45 horas a la semana máximo) y temporales (15 minutos continuos) (MSA de CHILE).

Figura 31

Concentraciones aceptables de amoníaco

Norma / Organismo	Límite Ponderado	Límite Temporal
Ministerio de Salud de Chile - DS 594	LPP: 20 ppm (14mg/m ³)	LPT: 35 ppm (24mg/m ³)
Conferencia Americana de Higienistas Industriales del Gobierno (ACGIH) Estados Unidos	TWA: 25 ppm (17 mg/m ³)	STEL: 35 ppm (24mg/ m ³)
Organización de Salud y Seguridad Ocupacionales (OSHA)	PEL: 50 ppm (35mg/m ³)	
Instituto Nacional para la Salud y Seguridad Ocupacionales (NIOSH)	REL: 25 ppm (18mg/m ³)	STEL: 35 ppm(27mg/m ³) IDLH: 300 ppm

Nota. Tomado de (MSA de CHILE).

Para la calidad de aire se realizó el análisis del amoniaco, utilizando la ecuación descrita en la sección 4.2 para los valores más importantes mencionados en el apartado 4.2.2, los datos del experimento están comprendidos entre [210 ;456], aplicando la transformación descrita anteriormente para los límites inferior y superior del rango, tenemos un resultado único de 54.55ppm, esto significa que todas las mediciones comprendidos en este intervalo van a mantener este único valor.

Según la Figura 31 este se encuentra por encima de todos los estándares internacionales descritos, es necesario mencionar que en (MSA de CHILE) y (health) no se especifican riesgos a la salud a corto o mediano plazo por exposiciones de tan baja concentración, se advierte que este gas es muy inflamable, y es complicado de detectar con el olfato por ser muy liviano en el aire, las consecuencias mas notables por la inhalación de este gas es la irritación de nariz y la garganta causando tos y un silbido en la respiración, se indica también que este no contribuye un riesgo para desarrollar cáncer o infertilidad, por lo que esta concentración aunque esta sobre los niveles permitidos no representa una amenaza para la salud de las personas.

Tolueno, sensor de calidad de aire

Otro de los gases que es medible en base al sensor de calidad de aire es el tolueno, que “es un líquido incoloro con un característico olor aromático. Es menos denso que el agua, inmisible en ella, sus vapores son mas densos que el aire y se encuentra en derivados de petróleo” (TOLUENO). Los niveles toxicidad del tolueno se describen en la figura 32.

Figura 32

Niveles de toxicidad de Tolueno

NIVELES DE TOXICIDAD:	
RQ: 1000	
IDLH: 2000ppm	
LD ₅₀ (en ratones): 5300 ppm.	
LD ₅₀ (en piel de conejos): 14000 mg/Kg.	
LD ₅₀ (oral en ratas): 7.53 ml/Kg, 5000 mg/ Kg.	
LDLo (oral en humanos): 50 mg/Kg.	
LC ₅₀ (por inhalación en ratones): 5320 ppm/8 h	
Niveles de irritación a ojos: 300 ppm (humanos); 0.87 mg, leve y 2 mg/24 h, severa (conejos).	
Niveles de irritación a piel de conejos: 435 mg, leve; 500 mg, moderada	
México:	
CPT: 375 mg/m ³ (100 ppm)	
CCT: 560 mg/m ³ (150 ppm. Se absorbe a través de la piel).	
Estados Unidos:	
TLV TWA: 375 mg/m ³ (100 ppm)	
TLV STEL: 560 mg/m ³ (150 ppm)	
Reino Unido:	Alemania:
Periodos largos: 560 mg/m ³ (150 ppm)	MAK: 375 mg/m ³ (100 ppm)
Francia:	Suecia:
VME: 550 mg/m ³ (150 ppm)	Límitemáximo:200mg/m ³ (50ppm)

Nota. Tomado de (TOLUENO)

Para la calidad de aire se realizó el análisis del tolueno, utilizando la ecuación descrita en la sección 4.2 para los valores más importantes mencionados en el apartado 4.2.1, los datos del experimento están comprendidos entre [210 ;456], aplicando la transformación descrita anteriormente para los límites inferior y superior del rango, tenemos un resultado único de 12.6048ppm. Este valor implica un riesgo bastante bajo para la salud de las personas pues éste compuesto en forma gaseosa se considera tóxico desde las 50ppm (Estándar Sueco).

En base a los datos obtenidos en la sección 5.3, se concluye que actualmente el lugar donde fue realizado este experimento debe cuidar los valores de CO mientras realiza sus actividades diarias, es recomendable que las personas que encuentran expuestas a la jornada completa de trabajo de 8 horas, realicen 30 minutos de ejercicio al día para contrarrestar los efectos de este gas, el amoníaco podría convertirse en un peligro potencial de explosión si se eleva demasiado su concentración, actualmente se encuentra escasamente sobre el límite permitido por el estándar de la organización OSHA, se recomienda dotar de artefactos de ventilación adicional en este lugar y finalizando el tolueno no representa ningún problema en su concentración actual.

Selección de algoritmos predictivos

Para (Rivero, 2016) (Valero, 2013) las redes neuronales son muy utilizadas para el pronóstico de series temporales, en (Alvarez) además se presentan otros algoritmos como: árboles de decisión, métodos de grafos, agrupación, entre otros como alternativas viables. Para este caso particular se han evaluado bajo la herramienta Rapid Miner y la librería Tensorflow versión 2.1.0, las siguientes opciones de modelos para el conjunto de datos y las dos variables CO y Calidad de aire:

- Modelo Lineal (Generalized linear model)
- Aprendizaje profundo (Deep learning)
- Árbol de decisión (Decision tree)

- Bosque aleatorio (Random forest)
- Árboles impulsados por gradiente (Gradient boosted trees)
- Máquina de vectores de soporte (Support vector machine)
- Redes neuronales recurrentes (RNR)

Para la evaluación de los modelos se utilizó los siguientes escenarios (agrupación de variables) en conformidad con los apartados 3.2.1 y 3.2.2

- Calidad de aire:
 - Escenario 1: **measure_co, measure_air_qa**
 - Escenario 2: **measure_air_qa**
- Concentración de Co:
 - Escenario 3:
measure_temperature, measure_humidity, measure_air_qa,
measure_co
 - Escenario 4: **measure_co**

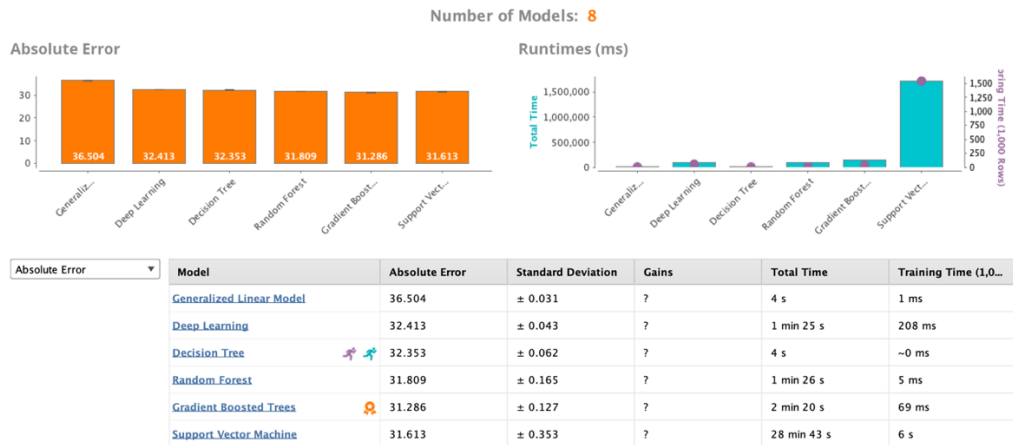
Escenario 1 (Rapid Minner):

Se utilizó la función auto modelado de Rapid Minner para realizar la estructuración de los 5 primeros algoritmos y determinar sus modelos de predicción, en la Figura 33 se presenta el resumen de los valores del error medio absoluto y los tiempos que tardó cada modelo en estructurar la solución, este último puede llegar a ser un factor determinante si en cada predicción es calculado, pero en la mayoría de las soluciones este proceso se realiza únicamente la primera vez que se entrena el modelo, ya que el resto de ocasiones se ocupa las funciones de predicción nativas según la herramienta que se utilice, evitando repetir el cálculo, ahorrando tiempos de respuesta en las aplicaciones utilizadas por los usuarios finales.

Figura 33

Error medio absoluto Escenario 1

Overview

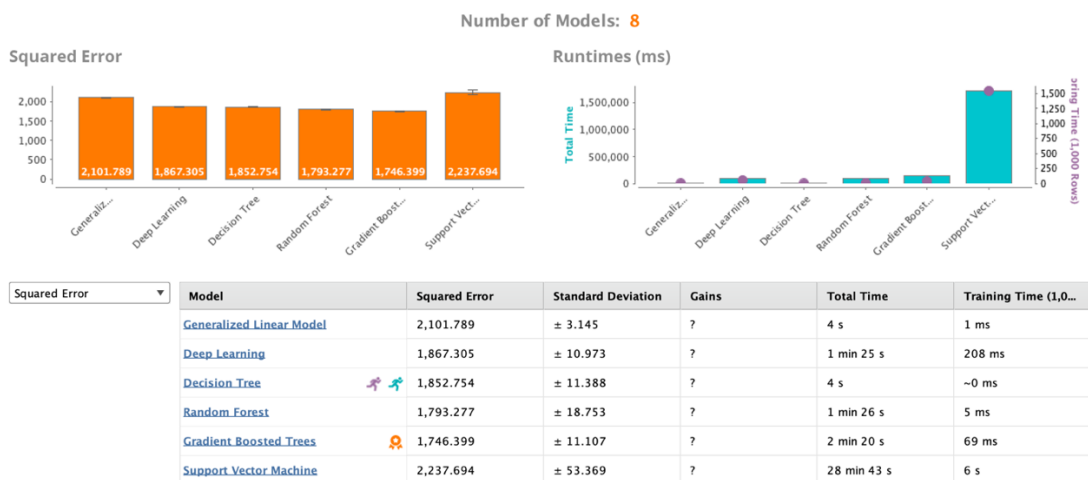


Para el error medio cuadrático (ver Figura 34), es importante mencionar que la herramienta Rapid Minner también determina en función de los resultados, cuál es el modelo mas rápido en efectuar una predicción y cual es el mejor modelo en función de métricas como los errores presentados en esta sección, la desviación estándar y el tiempo de entrenamiento.

Figura 34

Error medio cuadrático del escenario 1

Overview

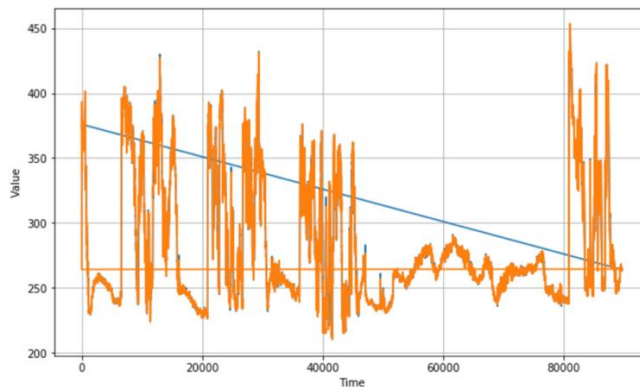


Escenario 2:

Para el análisis de el segundo escenario se utilizó Tensorflow ya que permite realizar análisis univariables (una sola variable), para este caso particular se evaluó la técnica de Redes neuronales recurrentes, en la Figura 35 se presenta un gráfico de los valores reales (color azul) y los valores predcidos por el modelo (color naranja), además de un resumen completo con los errores: medio absoluto, medio absoluto porcentual y medio cuadrático.

Figura 35

Errores modelo escenario 2



Error medio absoluto porcentual

```
tf.keras.metrics.mean_absolute_percentage_error(x_valid, results_ws).numpy()
```

0.29869681748720844

Error medio absoluto

```
tf.keras.metrics.mean_absolute_error(x_valid, results_ws).numpy()
```

0.8585339242205738

Error cuadrático medio

```
tf.keras.losses.MSE(x_valid, results_ws).numpy()
```

2.1315288679710425

Escenario 3:

En la figura 36 y 37 se describen nuevamente los 5 algoritmos iniciales mencionados en el principio del apartado 3.3, para los valores del error medio cuadrático y el error medio absoluto del escenario 3.

Figura 36

Error medio absoluto escenario 3

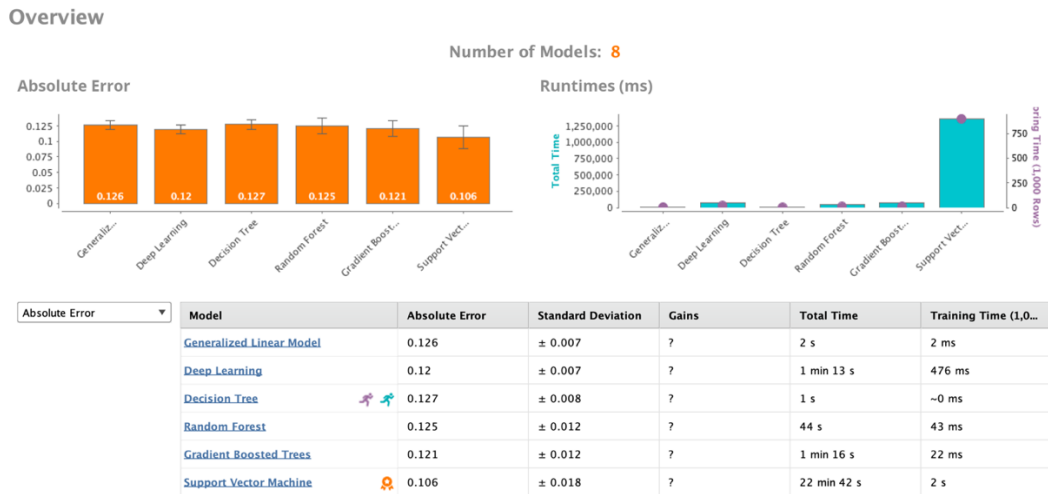
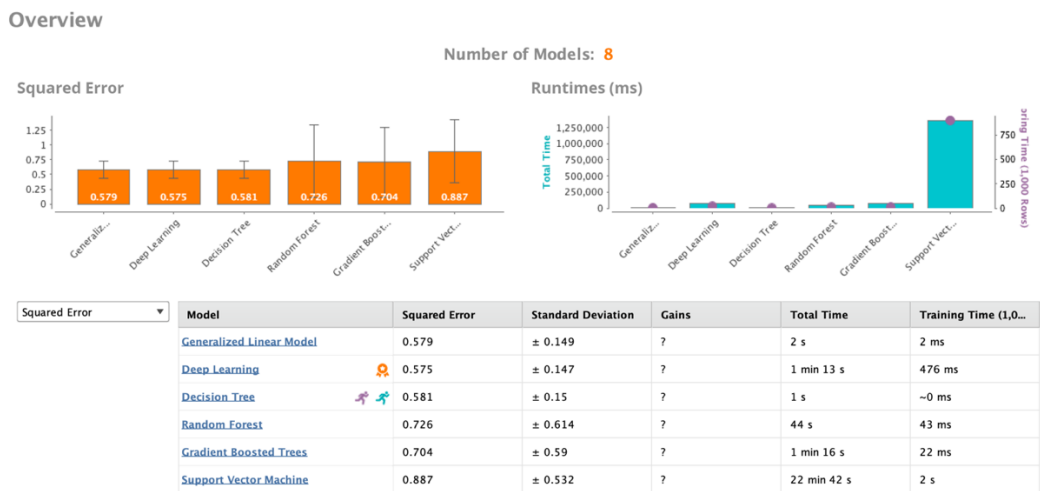


Figura 37

Error medio cuadrático escenario 3

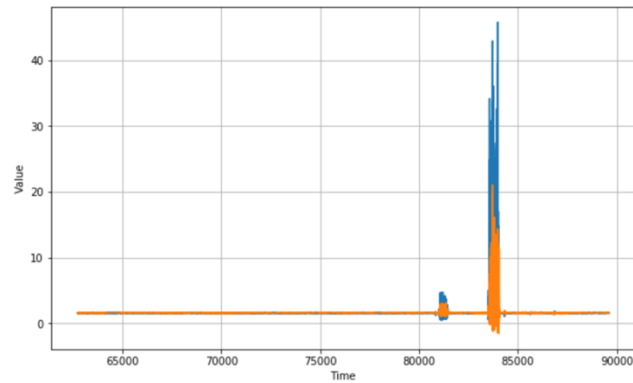


Escenario 4:

En la Figura 38, se presenta los valores de error descrito para la técnica de redes neuronales recurrentes del escenario 4, además de un grafico con los valores reales (azul) y predecidos (naranja), a simple vista se nota un acercamiento considerable entre ambas gráficas.

Figura 38

Resumen de variables escenario 4



Error medio absoluto porcentual

```
tf.keras.metrics.mean_absolute_percentage_error(x_valid, results).numpy()
8.078333
```

Error medio absoluto

```
tf.keras.metrics.mean_absolute_error(x_valid, results).numpy()
0.10313328
```

Error cuadrático medio

```
tf.keras.losses.MSE(x_valid, results).numpy()
0.89250284
```

Para la creación de los modelos acorde con las variables de los escenarios, se dividió la evaluación de los algoritmos en **multivariables**: para la herramienta Rapid Minner por la rápida funcionalidad en su opción de auto modelado y para los **univariables** que incluyan en sus soluciones la utilización de redes neuronales se utilizó la librería Tensor Flow por su capacidad de computo y especialización en este tipo de algoritmos.

Capítulo 5

VALIDACIÓN DE LA SOLUCIÓN

Evaluación de los modelos

Para validar los modelos enunciados en el Capítulo III se utilizaron las siguientes métricas, que permitieron estimar el error de predicción cometido por los modelos:

- Error medio absoluto (MAE):
 - Corresponde a la diferencia entre el valor real y el valor predicho dividido para la cantidad total de observaciones, sirve para cuantificar la precisión de una técnica de predicción (Hamilton, 1994).
 - El error absoluto medio es una medida común del error de pronóstico en análisis de series de tiempo (Hamilton, 1994).
- Error medio absoluto porcentual (MAPE):
 - Es una medición del tamaño del MAE en términos porcentuales, estima la magnitud del error porcentual (gestiondeoperaciones, 2015).
- Error medio cuadrático (MSE):
 - Miden la dispersión de los errores de la predicción, si el valor es pequeño la predicción se acerca a los valores reales, un valor grande anuncia errores de pronóstico considerable (Krajewski & Ritzman, 2000).

En la Tabla 14 se presenta un cuadro comparativo con los errores medio absoluto y medio cuadrático para los 6 modelos y los 4 escenarios propuestos en el apartado 3.3.

Tabla 14*Resumen métricas algoritmos predictivos*

Algoritmo	Error medio absoluto	Error medio cuadrático
Modelo Lineal (Generalized linear model)	Escenario 1: 36.404 Escenario 3: 0.126	Escenario 1: 2101.79 Escenario 3: 0.579
Aprendizaje profundo (Deep learning)	Escenario 1: 32.513 Escenario 3: 0.12	Escenario 1: 1867.305 Escenario 3: 0.575
Árbol de decisión (Decision tree)	Escenario 1: 32.353 Escenario 3: 0.127	Escenario 1: 1852.754 Escenario 3: 0.581
Bosque aleatorio (Random forest)	Escenario 1: 31.809 Escenario 3: 0.125	Escenario 1: 1793.277 Escenario 3: 0.726
Árboles impulsados por gradiente (Gradient boosted trees)	Escenario 1: 31.286 Escenario 3: 0.121	Escenario 1: 1746.399 Escenario 3: 0.704
Máquina de vectores de soporte (Support vector machine)	Escenario 1: 31.613 Escenario 3: 0.106	Escenario 1: 2237.694 Escenario 3: 0.89
Redes neuronales recurrentes (RNR)	Escenario 2: 0.8585 Escenario 4: 0.103	Escenario 2: 2.1315 Escenario 4: 0.89

Evaluando los resultados de las métricas de los algoritmos enunciados anteriormente, para la predicción de la variable **calidad de aire** los mejores resultados del error medio absoluto y del error medio cuadrático se dan el **escenario 2** y el algoritmo de redes neuronales recurrentes con un valor de 0.8585 y 2.1315 respectivamente.

Para el caso del CO se tuvo resultados bastante variables, los modelos: **lineal**, **redes neuronales recurrentes** y **máquina de vectores de soporte** se presentaron como las mejores alternativas. Al seleccionar el mejor modelo se utilizó la métrica: **error medio absoluto** como prioritaria para este estudio, pues permite obtener la diferencia entre los valores predichos y reales, lo cual es determinante a la hora de garantizar resultados fiables para los usuarios finales, por tanto, se selecciona al algoritmo de **redes neuronales recurrentes** y el escenario 4 por tener el valor más pequeño en la medida del error medio absoluto.

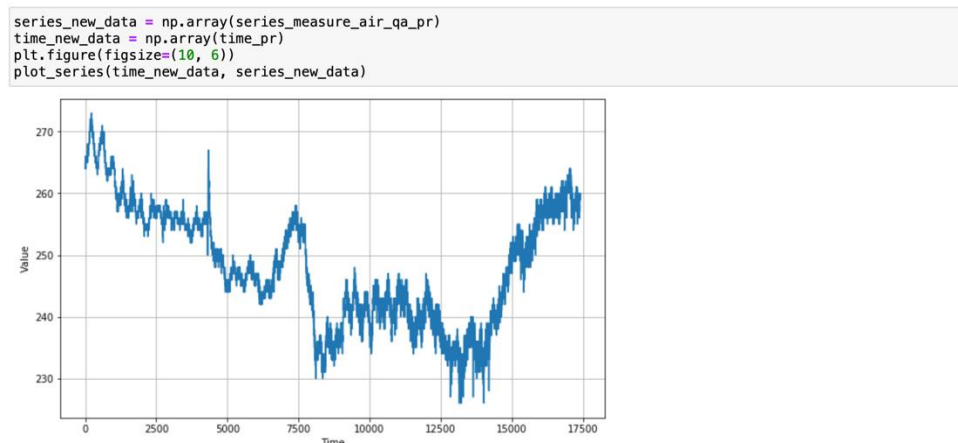
Validación de los modelos seleccionados.

Como se indico en el apartado 4.2, La validación de resultados se realizó en el grupo de datos diferente al que fue considerado para la construcción del modelo con un volumen de 17.409 registros.

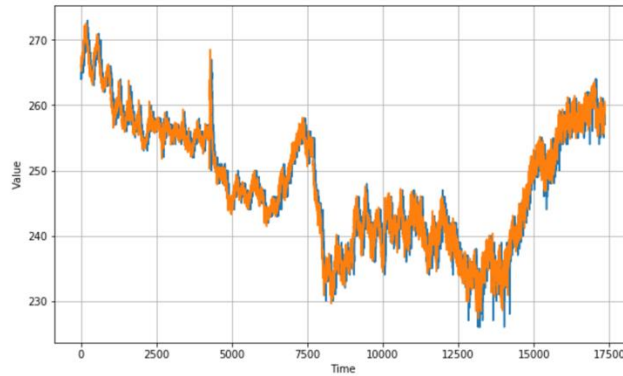
Para la variable calidad de aire, se presenta una distribución de datos que va en el intervalo comprendido entre [100; 275] (ver Figura 39), con un comportamiento similar al conjunto de datos que se utilizó en la construcción del modelo (ver apartado 4.2.2).

Figura 39

Datos de calidad de aire para pronóstico



Utilizando el modelo construido anteriormente con tensorflow, se pronosticó estos nuevos valores, obtenido que en los valores del error medio absoluto el valor se incrementa a 1.7292, presentando una diferencia de 0.8707 respecto al valor anterior, para el error medio cuadrático el valor se incrementa a 5.0882, con una diferencia de 2.9567 con relación al valor original (ver Figura 40), sin embargo aunque los valores se han incrementado el error absoluto porcentual aun es menor al 1% y si comparamos estos valores con respecto al resto de indicadores de los otros modelos el resultado esta muy por encima y es aceptable.

Figura 40*Predicción Calidad de aire***Error medio absoluto porcentual**

```
tf.keras.metrics.mean_absolute_percentage_error(series_new_data_sp, results_new_data_ws).numpy()
0.6995621334283605
```

Error medio absoluto

```
tf.keras.metrics.mean_absolute_error(series_new_data_sp, results_new_data_ws).numpy()
1.729237838722693
```

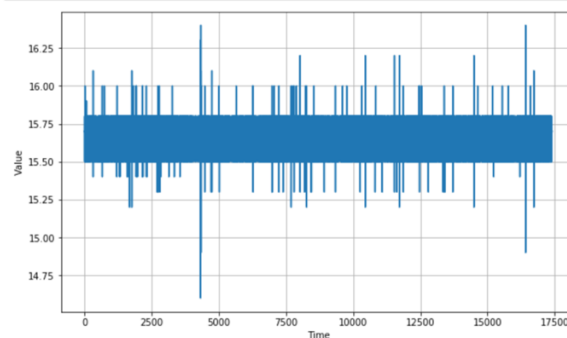
Error medio cuadratico

```
tf.keras.losses.MSE(series_new_data_sp, results_new_data_ws).numpy()
5.088235254083844
```

En relación con las emisiones de CO, tenemos que la distribución tiene una tendencia lineal en el intervalo [15.5; 15.8[con ciertos valores fuera del rango con un comportamiento aleatorio, encontramos también que el conjunto se desenvuelve en un rango muy reducido comprendido entre]0; 17[(ver Figura 41).

Figura 41*Datos de CO para pronóstico*

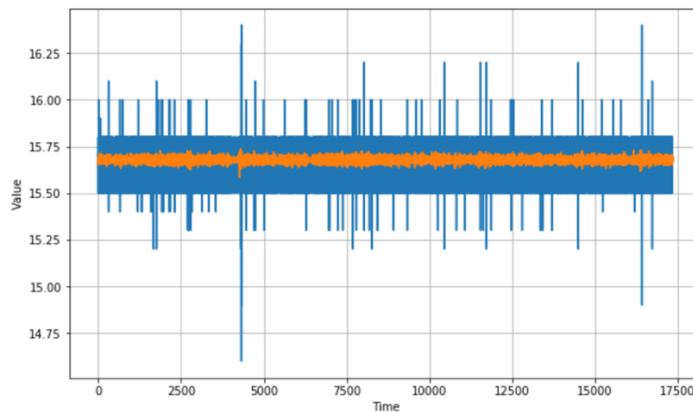
```
series_new_data = np.array(series_measure_co2_pr)
time_new_data = np.array(time_pr)
plt.figure(figsize=(10, 6))
plot_series(time_new_data, series_new_data)
```



Durante el pronóstico del CO, se encontró que a diferencia de la calidad de aire, los indicadores de error se reducen considerablemente, el error medio absoluto se redujo a 0.0063, con una diferencia de 0.097 menos que su valor anterior, y el error medio cuadrático paso a un valor de 8.5132E-5 lo cual es un reducción altamente considerable al valor anterior de 0.89, la reducción del error para el pronóstico va en relación al intervalo en el que los valores se encuentran, ya que es bastante reducido y lineal (ver Figura 42).

Figura 42

Predicción CO



Error medio absoluto porcentual

```
tf.keras.metrics.mean_absolute_percentage_error(series_new_data_sp, results_new_data_ws).numpy()
0.4007177661309915
```

Error medio absoluto

```
tf.keras.metrics.mean_absolute_error(series_new_data_sp, results_new_data_ws).numpy()
0.006267448704248126
```

Error medio cuadrático

```
tf.keras.losses.MSE(series_new_data_sp, results_new_data_ws).numpy()
8.513275828256641e-05
```

De esta manera se comprueba la efectividad de los modelos construidos en el capítulo IV, concluyendo que serán eficaces en conjuntos de datos que se encuentren en el intervalo que esté comprendido entre los valores mínimo y máximo de los datos iniciales utilizados para la construcción de los modelos.

Capítulo 5

Presentación de resultados

Para la visualización y control de la información recopilada por los sensores, se construyó una aplicación que permite monitorizar el estado actual y predicciones a futuro en base a los datos de las variables presentadas en el Capítulo IV, se utilizó una metodología ad-hoc derivada de XP (Extreme Programming) para completar este trabajo.

Exploración

Para el diseño de la aplicación web se requiere tomar en cuenta ciertos requisitos fundamentales detallados a continuación en las siguientes historias de usuario que adicional al formato presentado anteriormente contiene el prototipo (maqueta) del requerimiento.

Tabla 15

Historia de usuario 1 aplicación de presentación

Historia:	SB-APW-H1: Diseñar una aplicación web para visualizar los datos
Descripción	Como usuario final quiero que por medio de una aplicación web que contenga una cabecera (con los datos del nombre de la pagina, nombre del usuario autenticado en el sistema); una barra lateral de opciones de navegación (histórico, modelos y pronóstico) y una sección para desplegar cada una de las opciones del menú, para visualizar los datos actualmente recolectados.
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Crear un encabezado donde se presente el nombre de la pagina, alineado a la izquierda y a la derecha el encuentre el nombre del usuario del sistema. • Crear una barra lateral con las siguientes opciones de navegación: inicio, modelos y pronóstico • Crear una sección para el contenido de cada opción del menú de navegación

Historia:	SB-APW-H1: Diseñar una aplicación web para visualizar los datos
Prototipo:	
Precondiciones	Mantener activos los servicios de los subsistemas SB-DRA y SB-MDRA
Pos condiciones	La aplicación contará con el esqueleto base para la presentación de datos

Tabla 16

Historia 2 aplicación de presentación

Historia:	SB-APW-H2: Presentar resumen de los datos actuales
Descripción	Como usuario final quiero que por medio de una aplicación web se pueda presentar un resumen de como se encuentra distribuida la información actualmente, presentando los siguientes estadísticos descriptivos: media, mediana, moda y desviación estándar, además de gráficos que permitan ver la distribución de los valores en el tiempo, para analizar el estado actual de los datos.
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Presentar miniaturas con la siguiente información: tiempo de captura de la información, variables del experimento, variables para la predicción y modelos. • Presentar un grafico de línea y el resumen de los estadísticos descriptivos siguientes: promedio, mediana, moda y los indicadores del valor máximo, mínimo y la cantidad de observaciones. • Presentar dos gráficos de radar, uno con la cantidad de agrupaciones definidas por defecto por la librería empleada para el efecto y el otro con 16 agrupaciones. • Presentar el resumen de las variables: calidad de aire, monóxido de carbono, temperatura y humedad.

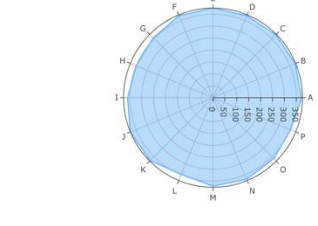
Historia:	SB-APW-H2: Presentar resumen de los datos actuales
Prototipo:	<div style="border: 1px solid #ccc; padding: 10px;"> <div style="text-align: right; font-size: small;">Andrés Rivaden</div> <div style="display: flex; justify-content: space-between; margin-bottom: 10px;"> <div style="border: 1px solid #ccc; padding: 5px; width: 20%;">Tiempo de captura (seg) 60</div> <div style="border: 1px solid #ccc; padding: 5px; width: 20%;">Variables 4</div> <div style="border: 1px solid #ccc; padding: 5px; width: 20%;">Variables para predicción 2</div> <div style="border: 1px solid #ccc; padding: 5px; width: 20%;">Modelos 2</div> </div> <div style="display: flex; align-items: flex-start;"> <div style="flex: 1;"> <p>Calidad de aire en el tiempo <small>Nov 15 2019, Ene 17 2020</small></p>  </div> <div style="flex: 0.5; border-left: 1px solid #ccc; padding-left: 10px; font-size: small;"> <p>Cantidad de observaciones 89611</p> <p>Valor mínimo 210</p> <p>Valor máximo 453</p> <p>Promedio 280.27</p> <p>Mediana 263</p> <p>Moda 256</p> <p>Desviación estándar 45.85</p> </div> </div> <div style="display: flex; justify-content: space-around; margin-top: 10px;"> <div style="width: 45%;"> <p>Gráfico de radar calidad de aire en el tiempo <small>Clase: grupo de defectos</small></p>  </div> <div style="width: 45%;"> <p>Gráfico de radar calidad de aire en el tiempo <small>Clase: grupo 17</small></p>  </div> </div> </div>
Precondiciones	Mantener activos los servicios de los subsistemas SB-DRA y SB-MDRA La aplicación deberá contener el esqueleto de la historia SB-APW-H1
Pos condiciones	ninguna

Tabla 17

Historia de usuario 3 aplicación de presentación

Historia:	SB-APW-H3: Presentar un resumen de los modelos de pronóstico
Descripción	Como usuario final quiero que por medio de una aplicación web se pueda presentar un resumen de los modelos creados de las variables de pronóstico, donde se indique la cantidad de registros que se requiere para la predicción, el error medio absoluto, error medio absoluto porcentual y el error medio cuadrático del modelo, al pronosticar los valores iniciales y de validación, para conocer la efectividad de estos.
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Presentar una tabla resumen por cada modelo que contenga la siguiente información: cantidad de registros para la predicción, error medio absoluto, error medio absoluto porcentual, error medio cuadrático. • Presentar un gráfico de radar de a información real contra la pronosticada. • Presentar la información de los modelos de las 2 variables de pronóstico para los datos iniciales y los datos de validación.

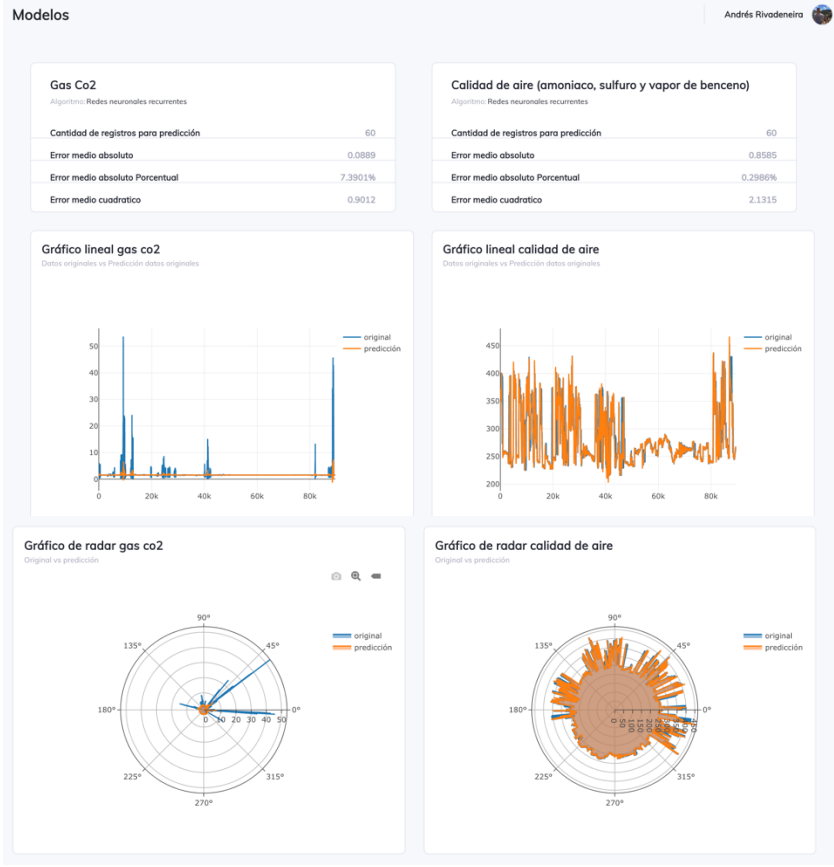

Historia:	SB-APW-H3: Presentar un resumen de los modelos de pronóstico
Prototipo:	 <p>The screenshot displays a dashboard titled 'Modelos' with the user 'Andrés Rivadeneira'. It features two main sections: 'Gas Co2' and 'Calidad de aire (amonaco, sulfuro y vapor de benceno)'. Each section includes a table of performance metrics and two types of visualizations: a line graph and a radar chart. The 'Gas Co2' metrics are: 60 records, 0.0889 absolute error, 7.3901% percentage error, and 0.9012 quadratic error. The 'Calidad de aire' metrics are: 60 records, 0.8505 absolute error, 0.2980% percentage error, and 2.1315 quadratic error. The line graphs compare 'original' data (blue) with 'predicción' (orange), and the radar charts compare 'original' (blue) with 'predicción' (orange) across various angles.</p>
Precondiciones:	Mantener activos los servicios de los subsistemas SB-DRA y SB-MDRA
Pos condiciones	La aplicación deberá contener el esqueleto de la historia SB-APW-H1 ninguna

Tabla 18

Historia de usuario 4 aplicación de presentación

Historia:	SB-APW-H4: Pronosticar siguiente valor futuro
Descripción	Como usuario final quiero que por medio de una aplicación web se presente una opción para que en base a los últimos datos almacenados se pueda predecir el siguiente valor.
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Presentar una tabla con los últimos 60 valores de las variables que se utilizarán el pronóstico • Presentar una opción que permita predecir el siguiente valor bajo demanda.

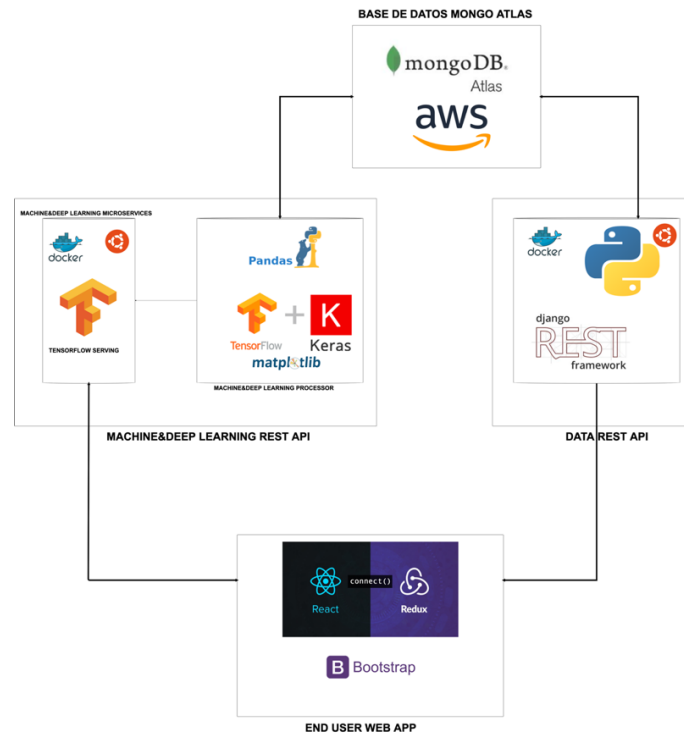
Historia:	SB-APW-H4: Pronosticar siguiente valor futuro
Prototipo:	
Precondiciones	<p>Mantener activos los servicios de los subsistemas SB-DRA y SB-MDRA</p> <p>La aplicación deberá contener el esqueleto de la historia SB-APW-H1</p>
Pos condiciones	ninguna

Ubicación de la aplicación dentro de la arquitectura de software

La aplicación de presentación de este capítulo, se alimenta de los servicios publicados por los subsistemas SB-DRA y SB-MDRA (ver Figura 43) utilizando peticiones asíncronas para la obtención de la información.

Figura 43

Diagrama de arquitectura de software

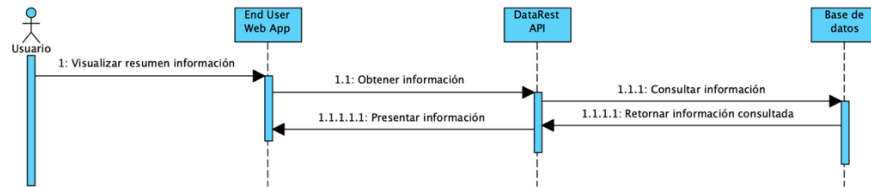


Diagramas de secuencia de la aplicación

Las opciones inicio y modelo de la aplicación se manejan bajo la siguiente secuencia (ver gráfico 44): cuando se carga la información el End User Web App se envía una petición asincrónica al micro servicio del DataRest API para obtener los datos que requiera la pantalla, utilizando una consulta a la Base de datos y efectuando el procesamiento según lo que la lógica del negocio requiera.

Figura 44

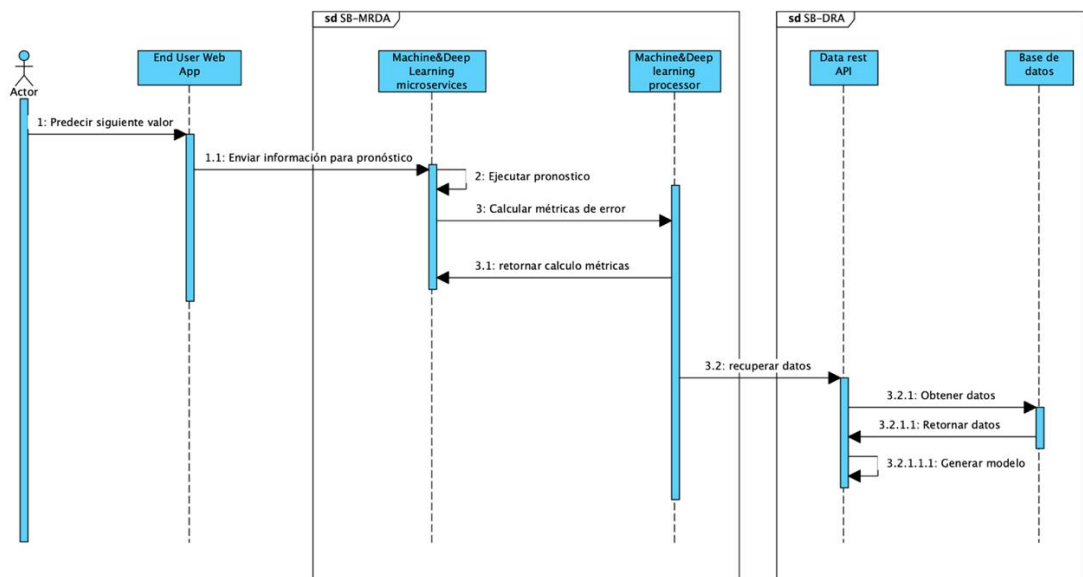
Diagrama de secuencia visualización de datos



Para el proceso de predicción de valores futuros se realiza la siguiente secuencia (ver Figura 45): se invoca a través del End User Web App una petición al SB-MRDA que por medio de un conjunto de datos de entrada realiza el calculo del valor futuro, obteniendo la información del modelo publicado en Tensorflow Server.

Figura 45

Diagrama de secuencia predicción de datos



Codificación

Para la construcción de la aplicación de presentación se utilizó el lenguaje de programación Javascript y el Framework React con el patrón de Redux que permite utilizar un contenedor que predice el estado de la aplicación (reactividad), cargando

solo lo que se requiere (es.[react.js.org](https://es.reactjs.org/)), como gestor de paquetes se utilizó NPM y como servidor de aplicaciones NodeJs.

Pruebas

Para validar el trabajo realizado, en la interfaz de usuario se verificó la funcionalidad desarrollada, mediante los siguientes casos de prueba:

Tabla 19

Caso de prueba APW1

Caso de prueba:	CP-APW-1: Probar presentación del resumen de los datos actuales
Descripción	En base al requerimiento SB-APW-H2, se requiere validar que la información presentada sea la correcta, validando con la información de los subsistemas
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Verificar que la aplicación inicie con la opción histórico al cargarse. • Verificar que al cambiar de opción no se vuelva a realizar la consulta para recolectar los datos. • Verificar que sea posible interactuar con los gráficos de la aplicación (ampliar, seccionar, regresar a los valores por defecto) • Verificar que los gráficos coincidan con los valores obtenidos por la librería matplotlib en el subsistema SB-DRA.
Estatus:	Aprobado.
Observaciones:	Se deberá validar con el cliente que gráficos son los necesarios presentar.

Tabla 20

Caso de prueba APW2

Caso de prueba:	CP-APW-2: Probar presentación del resumen de los modelos
------------------------	--

Descripción	En base al requerimiento SB-APW-H3, se requiere validar que la información presentada sea la correcta, validando con la información de los subsistemas.
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Verificar que la opción de menú de opciones al ser seleccionado cambie el contenido de la aplicación por el correspondiente a los modelos. • Verificar que al cambiar de opción no se vuelva a realizar la consulta para recolectar los datos. • Verificar que sea posible interactuar con los gráficos de la aplicación (ampliar, seccionar, regresar a los valores por defecto) • Verificar que los gráficos coincidan con los valores obtenidos por la librería matplotlib en el subsistema SB-DRA.
Estatus:	Aprobado.
Observaciones:	Se deberá validar con el cliente que gráficos son los necesarios presentar.

Tabla 21

Caso de prueba APW1

Caso de prueba:	CP-APW-3: Probar pronóstico del siguiente valor
Descripción	En base al requerimiento SB-APW-H3, se requiere validar que el pronóstico de los siguientes valores en base a los datos se realice correctamente.
Reglas de Negocio (criterios de aceptación):	<ul style="list-style-type: none"> • Verificar que la opción de menú de opciones al ser seleccionado cambie el contenido de la aplicación por el correspondiente al pronóstico. • Verificar que al cambiar de opción no se vuelva a realizar la consulta para recolectar los datos. • Verificar que los datos obtenidos para la predicción sean los últimos 60 registros de la base de datos. • Verificar que la aplicación mientras dure el proceso de obtención del pronóstico muestre un símbolo de carga hasta que la información sea recolectada.
Estatus:	Aprobado.
Observaciones:	Ninguna.

Capítulo 6

Conclusiones y recomendaciones

Conclusiones:

Al finalizar este trabajo a continuación se presentan las conclusiones derivadas de los objetivos específicos cumplidos a lo largo del desarrollo de la investigación.

- OE1-C1: La falta de monitorización y control de las instalaciones gas y las fallas humanas son las principales causas de las fugas de gas en espacios cerrados, esto ocasiona muertes, cuantiosos daños materiales y graves afectaciones a la salud de las personas.
- OE2-C2: Durante la revisión preliminar de literatura de este trabajo se estableció la viabilidad del proyecto, ya que esta problemática no ha sido abarcada en el ámbito de la inteligencia de negocios y la ciencia de datos, además existen muy pocas investigaciones al respecto.
- OE3-C3: La utilización de sensores de gas de bajo costo no permite captar valores muy pequeños, por lo que se descartaron 3 de las 5 variables de predicción del experimento ya que sus datos tendían a ser 0.
- OE3-C4: La construcción de software basada en un conjunto de contenedores con aplicaciones pequeñas interconectadas entre si permite, agilizar el proceso de configuración e instalación de nuevos dispositivos recolectores de información.
- OE4-C5: La función de auto modelado de herramientas como Rapid Miner permite evaluar algunas posibilidades de solución para un conjunto de datos de estudio, ya que utilizando varios algoritmos estándar (regresión lineal, árboles de decisión, etc.) provee de un conjunto de indicadores como: velocidad de respuesta, errores promedio, entre otros, que facilitan la toma de decisiones al momento de seleccionar aquel que se ajuste más a la realidad de un estudio particular.

- OE5-C6: Construir modelos predictivos utilizando redes neuronales recurrentes (RNN) permitió analizar, monitorizar y predecir la acumulación de CO y determinar la concentración de gases como el amoníaco, sulfuro y tolueno con los errores más bajos respecto a otros algoritmos de minería de datos; en las series temporales el error medio absoluto y el error medio cuadrático son indicadores importantes al evaluar modelos predictivos, pues se puede determinar la precisión de los algoritmos y la dispersión sus errores.
- OE6-C7: Conocer el estado actual de un espacio cerrado donde se utilice gas, implica una herramienta que ayudará a la toma de decisiones preventivas que permitan mitigar riesgos a la salud de las personas, así también una aplicación web que utilice reactividad y cargue sus componentes bajo demanda podría representar un medio adecuado para la difusión de la información gases en espacios cerrados a usuarios finales

Recomendaciones:

A continuación, de cada conclusión se realizan las siguientes recomendaciones, que se deberán tener en cuenta en el momento de replicar o mejorar el estudio realizado.

- C1-R1: Después de implementar el proyecto en un ambiente real y mantenerlo en funcionamiento durante algún tiempo, se debería realizar un nuevo estudio de validación que compruebe la efectividad de la solución propuesta en este trabajo.
- C2-R2: Ampliar el numero de repositorios digitales consultados utilizando fuentes como Springer o Scopus revisando posibles mejoras en proyectos futuros.

- C3-R3: Sustituir los sensores actuales por sensores de la misma serie, pero de una gama mucho mas alta y precisa.
- C4-R4: Incluir la administración de contenedor con Kubernetes para facilitar la administración del sistema de la solución propuesta.
- C5-R5: Utilizar funciones de auto modelado de otras herramientas como Google Cloud Platform para corroborar los resultados obtenidos.
- C6-R6: Recolectar información de distintos lugares y comprobar la eficacia de los modelos obtenidos a partir de esta solución.
- C7-R7: Para aligerar los tiempos de carga de grandes volúmenes de información se debería probar bases de datos columnares nativas como MemSql, también se recomienda reducir el peso de la aplicación web utilizando Svelte.
- C7-R8: Para aligerar los tiempos de carga de grandes volúmenes de información, migrar a un entorno Serverless on cloud utilizando GCP (Google Cloud Platform) o AWS (Amazon web services).

Trabajos Futuros

Cotejar los datos obtenidos con especialistas de la salud, que permitan determinar rangos óptimos de cada uno de los gases que pudiesen representar un riesgo potencial para la salud, también se debe incluir un modulo de alertas de seguridad basadas en notificaciones push, correo electrónico y mensajes sms a los usuarios finales, cuando un peligro potencial sea detectado. Se deberá incluir también un sistema de registro de afectaciones o posibles síntomas derivados de la exposición de gases, como un control para alertar y monitorizar la salud de los usuarios finales.

BIBLIOGRAFÍA

Censos, I. N. (12 de 2011). Recuperado el 15 de 02 de 2019, de <http://www.ecuadorencifras.gob.ec/documentos/web->

[inec/Encuestas_Ambientales/Ambientales_2011/Presentacion_hog2011.pdf](http://www.ecuadorencifras.gob.ec/documentos/web-inec/Encuestas_Ambientales/Ambientales_2011/Presentacion_hog2011.pdf)

elcomercio. (20 de 07 de 2017). Recuperado el 15 de 02 de 2019, de [https://www.elcomercio.com/actualidad/fuga-gas-calefon-quito-](https://www.elcomercio.com/actualidad/fuga-gas-calefon-quito-bomberos.html)

[bomberos.html](https://www.elcomercio.com/actualidad/fuga-gas-calefon-quito-bomberos.html)

Servicio Nacional de Gestión de Riesgos y Emergencias. (2015).

gestionderiesgos.gob.ec. Recuperado el 15 de 02 de 2019, de

<https://www.gestionderiesgos.gob.ec/sgr-cuida-de-ti-los-tanques-de-gas-asesinos-silenciosos/>

elcomercio. (16 de 03 de 2015). Recuperado el 15 de 02 de 2019, de [https://www.elcomercio.com/actualidad/explosiones-fugadegas-quito-](https://www.elcomercio.com/actualidad/explosiones-fugadegas-quito-heridos-bomberos.html)

[heridos-bomberos.html](https://www.elcomercio.com/actualidad/explosiones-fugadegas-quito-heridos-bomberos.html)

metroecuador. (25 de 12 de 2017). Recuperado el 15 de 02 de 2019, de [https://www.metroecuador.com.ec/ec/noticias/2017/12/25/dos-fallecidos-](https://www.metroecuador.com.ec/ec/noticias/2017/12/25/dos-fallecidos-tras-explosion-restaurante-toronto-quito.html)

[tras-explosion-restaurante-toronto-quito.html](https://www.metroecuador.com.ec/ec/noticias/2017/12/25/dos-fallecidos-tras-explosion-restaurante-toronto-quito.html)

Jácome, E. (16 de 03 de 2015). *www.elcomercio.com*. Recuperado el 22 de 04 de 2019, de [https://www.elcomercio.com/actualidad/quito-explosion-gas-vivienda-](https://www.elcomercio.com/actualidad/quito-explosion-gas-vivienda-ecuador.html)

[ecuador.html](https://www.elcomercio.com/actualidad/quito-explosion-gas-vivienda-ecuador.html)

Gallagher, C. M. (2015). A Bayesian Classification Approach to Improving Performance for a Real-World Sales Forecasting Application . *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*. Miami, FL, USA.

Lopez, C. C., & Cadavid, A. N. (2016). Machine learning classifiers for android malware analysis. *2016 IEEE Colombian Conference on Communications and Computing (COLCOM)*. Cartagena.

Yann LeCun, Y. B. (2015). *Deep learning*. Recuperado el 02 de 03 de 2019, de <https://www.nature.com/articles/nature14539>

Lorenzo, J. M. (2008). *Mineria de datos*. Madrid: Thomson Ediciones.

Gorakala, S. K. (2015). *Building a Recommendation System with R*. Packt Publishing.

Marcos, F. V. (s.f.). *LA CONTAMINACIÓN AMBIENTAL COMO FACTOR DETERMINANTE DE LA SALUD* . Recuperado el 01 de 03 de 2019, de https://www.scielosp.org/scielo.php?pid=S1135-57272005000200001&script=sci_arttext&tIng=es

ambiente, M. d. (s.f.). *Controlar la contaminación ambiental contribuye a mejorar la calidad de vida de la población*. Recuperado el 03 de 03 de 2019, de <http://www.ambiente.gob.ec/controlar-la-contaminacion-ambiental-contribuye-a-mejorar-la-calidad-de-vida-de-la-poblacion/>

nationalgeographic. (05 de 09 de 2010). *La contaminación del aire*. Recuperado el 02 de 03 de 2019, de <https://www.nationalgeographic.es/medio-ambiente/la-contaminacion-del-aire>

Valle, F. A. (s.f.). <http://www.insht.es>. Recuperado el 02 de 03 de 2019, de http://www.insht.es/InshtWeb/Contenidos/Documentacion/FichasTecnicas/NTP/Ficheros/301a400/ntp_340.pdf

.WHO Global air pollution Platform. . (s.f.). *Contaminación del Aire Ambiental*. Recuperado el 02 de 03 de 2019, de

https://www.paho.org/hq/index.php?option=com_content&view=article&id=12918:ambient-air-pollution&Itemid=72243&lang=es

Joao B. A. Gomes, J. J.-M. (2018). An IoT-Based Smart Solution for Preventing Domestic CO and LPG Gas Accidents . *2018 IEEE 10th Latin-American Conference on Communications (LATINCOM)*. Guadalajara, Mexico.

Medeiros, G. V., Santos, M. R., Lopes, A. S., & Edmilson C. B. (2017). Smartgas: a smart platform for cooking gas monitoring. *2017 IEEE First Summer School on Smart Cities (S3C)*. Natal, Brazil.

Metta Santiputri, M. T. (2018). IoT-based Gas Leak Detection Device. *2018 International Conference on Applied Engineering (ICAE)*. Batam, Indonesia.

pololu. (s.f.). Recuperado el 31 de 03 de 2019, de <https://www.pololu.com/file/0J309/MQ2.pdf>

PASTOR, J. (25 de 04 de 2018). *xataka*. Obtenido de Raspberry Pi 3 Model B+, análisis: más potencia y mejor WiFi para un miniPC que sigue asombrando: <https://www.xataka.com/ordenadores/raspberry-pi-3-model-b-analisis-mas-potencia-y-mejor-wifi-para-un-minipc-que-sigue-asombrando>

arduino. (s.f.). *arduino*. Recuperado el 22 de 03 de 2020, de arduino: <https://store.arduino.cc/usa/arduino-uno-rev3>

MongoDb. (s.f.). *MongoDb Atlas*. Recuperado el 22 de 03 de 2020, de https://www.mongodb.com/cloud/atlas/efficiency?utm_source=google&utm_campaign=gs_footprint_row_search_brand_atlas_desktop&utm_term=mongodb%20atlas&utm_medium=cpc_paid_search&utm_ad=e&gclid=Cj0KCQjwmdzzBRC7ARIsANDqRRlls53UH6qG2B3RBx9jMXQoeRSNZ_oOZ_y6OZ3sn5

- Arduino. (s.f.). *Arduino*. Recuperado el 22 de 03 de 2020, de Arduino:
<https://www.arduino.cc/en/main/software>
- Foundation, E. (s.f.). *Mosquitto*. Recuperado el 22 de 03 de 2020, de
<https://mosquitto.org/>
- OpenJS, F. (s.f.). *nodered.org*. Obtenido de nodered.org: <https://nodered.org/>
- Docker. (2020). *Docker*. Obtenido de Docker:
<https://www.docker.com/resources/what-container>
- aspbian. (2020). *aspbian*. Obtenido de <https://www.raspbian.org/>
- Jupyter, P. (2020). *jupyter.org*. Obtenido de <https://jupyter.org/>
- Gartner, I. (11 de 02 de 2020). *Magic Quadrant for Data Science and Machine Learning Platforms*. Obtenido de Magic Quadrant for Data Science and Machine Learning Platforms: <https://www.gartner.com/doc/reprints?id=1-1YCR6NY7&ct=200213&st=sb>
- Seo, J. D. (02 de 06 de 201806). *Trend, Seasonality, Moving Average, Auto Regressive Model : My Journey to Time Series Data with Interactive Code*. Obtenido de Trend, Seasonality, Moving Average, Auto Regressive Model : My Journey to Time Series Data with Interactive Code:
<https://towardsdatascience.com/trend-seasonality-moving-average-auto-regressive-model-my-journey-to-time-series-data-with-edc4c0c8284b>
- Pedroza, H., & Diovsky, L. (2006). Sistema de Analisis Estadistico con SPSS. En H. Pedroza, & L. Diovsky, *Sistema de Analisis Estadistico con SPSS* (pág. 97). Managua, Nicaragua.

Krajewski, L. J., & Ritzman, L. P. (2000). Administración de operaciones: estrategia y análisis. En L. J. Krajewski, & L. P. Ritzman, *Administración de operaciones: estrategia y análisis* (pág. 519). Naucalpan de Juárez: Camara de la industria editorial mexicana.

Hamilton, J. D. (1994). Time Series Analysis . En J. D. Hamilton, *Time Series Analysis* .

Obtenido de MAE - Error medio absoluto:
<https://support.numxl.com/hc/es/articles/215969423-MAE-Error-medio-absoluto>

gestiondeoperaciones. (26 de 01 de 2015). *Error Porcentual Absoluto Medio (MAPE) en un Pronóstico de Demanda*. Obtenido de Error Porcentual Absoluto Medio (MAPE) en un Pronóstico de Demanda: Error Porcentual Absoluto Medio (MAPE) en un Pronóstico de Demanda

Vásconez, D. F., Celi, S., Ayabaca, C., & Rocha-Hoyos, J. (2018). El riesgo en el almacenamiento de GLP en el Ecuador The risk of LPG storage in Ecuador. *INNOVA Research Journal 2018, Vol 3, No. 1, 19-29.*, 19-29.

lahora. (27 de 12 de 2017). *www.lahora.com.ec*. Obtenido de *www.lahora.com.ec*:
<https://www.lahora.com.ec/tungurahua/noticia/1102124136/preocupacion-por-uso-de-gas-domestico-en-locales-de-expendio-de-comidas->

Piñango, R. (s.f.). *Asegurando la Calidad de la Experiencia de Usuario*. Recuperado el 28 de 03 de 2020, de Asegurando la Calidad de la Experiencia de Usuario:
<https://sg.com.mx/revista/52/asegurando-la-calidad-la-experiencia-usuario>

Takuto Kimura, M. Y. (2015). QUVE: QoE Maximizing Framework for Video-Streaming. *IEEE JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015*, 16.

Simon N.B. Gunkel, M. S. (2015). A QoE study of different stream and layout configurations in video conferencing under limited network conditions. *QoMEX 2015*, 6.

Carlos Quadros, E. C. (2015). Beacon-less Video Streaming Management for VANETs Based on QoE and Link-Quality. *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM2015)*, 191-198.

Aguilar, L. J. (2013). Big Data. En *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Mexico: Alfaomega.

es.euronews.com. (s.f.). *es.euronews.com*. Recuperado el 03 de 04 de 2020, de <https://es.euronews.com/>

Euronews en español. (05 de 12 de 2019). *euronews*. Obtenido de <https://es.euronews.com/2019/12/05/tragica-explosion-de-gas-en-una-estacion-de-esqui-del-sur-de-polonia>

es.euronews.com. (17 de 01 de 2019). *es.euronews.com*. Obtenido de <https://es.euronews.com/2019/01/17/violento-incendio-y-explusiones-de-bombonas-en-la-universidad-de-lyon>

es.euronews.com. (06 de 12 de 2019). *es.euronews.com*. Obtenido de <https://es.euronews.com/2019/12/06/al-menos-5-muertos-en-un-incendio-en-eslovaquia>

Pascual, J. C. (01 de 01 de 2019). *euronews*. Obtenido de <https://es.euronews.com/2019/01/01/rescatan-un-bebe-entre-los-gelidos-escombros-de-la-explosion-en-los-urales-en-rusia>

es.euronews.com. (31 de 12 de 2018). *es.euronews.com*. Obtenido de es.euronews.com: <https://es.euronews.com/2018/12/31/siete-muertos-y-mas-de-30-desaparecidos-en-el-derrumbe-parcial-de-un-edificio-en-rusia>

Pascual, J. C. (02 de 01 de 2019). *es.euronews.com*. Obtenido de <https://es.euronews.com/2019/01/02/ascienden-a-21-los-muertos-en-el-derrumbe-producido-por-una-explosion-de-gas-en-los-urales>

es.euronews.com. (31 de 12 de 2018). *es.euronews.com*. Obtenido de es.euronews.com: <https://es.euronews.com/2018/12/31/decenas-de-desaparecidos-tras-el-desplome-de-un-edificio-en-rusia>

Martín-Fradejas, Y. (14 de 09 de 2018). *es.euronews.com*. Obtenido de <https://es.euronews.com/2018/09/14/decenas-de-explosiones-de-gas-simultaneas-en-boston>

AseoYumbo. (20 de 06 de 2019). Obtenido de <https://www.elheraldo.co/colombia/explosion-por-presunta-fuga-de-gas-en-notaria-vallecaucana-deja-seis-heridos-643373>

ONess. (2020). *Ciclo de vida de un proyecto XP*. Obtenido de Ciclo de vida de un proyecto XP: <http://oness.sourceforge.net/proyecto/html/ch05s02.html>

Bugarin, J. L. (26 de 03 de 2019). *CICLO DE VIDA DEL MACHINE LEARNING*. Obtenido de CICLO DE VIDA DEL MACHINE LEARNING : <https://consultorjava.com/blog/ciclo-de-vida-del-machine-learning/>

Portillo., S. R. (20 de 02 de 2020). *Tipos de contaminantes ambientales*. Obtenido de <https://www.ecologiaverde.com/tipos-de-contaminantes-ambientales-2477.html>

- Vázquez, F. (09 de 01 de 2018). *Deep Learning fácil con DeepCognition* . Obtenido de <https://planetachatbot.com/deep-learning-f%C3%A1cil-con-deeppcognition-9af43b2319ba>
- Uriarte, J. M. (18 de 08 de 2019). *Contaminación del Aire*. Obtenido de Contaminación del Aire: <https://www.caracteristicas.co/contaminacion-del-aire/>
- Garg, S., Chatterjee, J. M., & KumarAgrawal, R. (2018). Design of a Simple Gas Knob: An Application of IoT. *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*, 22-24.
- Zhang, F., Shu, L., Zhu, C., Wu, X., Li, K., Zeng, J., . . . Zhang, M. (2015). A WSN based system for CO2 concentration monitoring in large-scale petrochemical plants . *2015 IEEE/CIC International Conference on Communications in China - Workshops (CIC/ICCC)*, 1-4.
- Suma, V., Shekar, R. R., & Akshay, K. A. (2019). Gas Leakage Detection Based on IOT. *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, 12-14.
- Cárdenas, F. d. (s.f.). *El proceso de minería de datos*. Recuperado el 04 de 04 de 2020, de <https://repository.uaeh.edu.mx/revistas/index.php/huejutla/article/view/1009/4835>
- Hoc, S. A. (18 de 12 de 2017). *Qué son las historias de usuario y para qué se necesitan*. Obtenido de <https://solvingadhoc.com/las-historias-usuario-funcion-agilidad/>

- sparkfun. (2019). *ARDUINO COMPARISON GUIDE*. Obtenido de ARDUINO COMPARISON GUIDE:
https://www.sparkfun.com/standard_arduino_comparison_guide
- All3DP. (02 de 01 de 2020). *2020 Best Single Board Computers/ Raspberry Pi Alternatives*. Obtenido de <https://all3dp.com/1/single-board-computer-raspberry-pi-alternative/>
- openclassrooms. (30 de 03 de 2018). *Diagramas de Casos de Uso* . Obtenido de <https://openclassrooms.com/en/courses/4990961-planea-tu-proyecto-con-uml/4996511-diagramas-de-casos-de-uso>
- stadium.unad.edu.co. (s.f.). *Diagramas de Casos de Uso*. Recuperado el 01 de 05 de 2020, de http://stadium.unad.edu.co/ovas/10596_9839/diagramas_de_casos_de_uso.html
- Merino, T. A. (09 de 01 de 2018). Obtenido de <https://www.icemd.com/digital-knowledge/articulos/el-orm-como-herramienta-eficiente-de-trabajo/>
- tensorflow. (2020). *Tensorflow*. Obtenido de <https://www.tensorflow.org/>
- Herranz, J. I. (2011). *TDD como metodología de diseño de software*. Obtenido de <https://www.paradigmadigital.com/dev/tdd-como-metodologia-de-diseno-de-software/>
- Rivero, C. R. (12 de 2016). <https://rdu.unc.edu.ar/>. Obtenido de <https://rdu.unc.edu.ar/bitstream/handle/11086/4604/TesisCRodriguezRivero.pdf?sequence=1&isAllowed=y>

Valero, C. S. (2013). *dspace.uclv.edu.cu*. Obtenido de <https://dspace.uclv.edu.cu/bitstream/handle/123456789/1935/Miner%C3%ADa%20de%20datos%20para%20series%20temporales%20en%20Weka%20y%20su%20aplicaci%C3%B3n%20en%20el%20pron%C3%B3stico%20de%20precipitaciones.pdf?sequence=1&isAllowed=y>

Alvarez, F. M. (s.f.). *lsi.us.es*. Recuperado el 01 de 05 de 2020, de <http://www.lsi.us.es/docs/doctorado/memorias/Martinez,%20Francisco.pdf>

es.redux.js.org. (s.f.). *Redux*. Recuperado el 01 de 05 de 2020, de <https://es.redux.js.org/>

ecured. (s.f.). *Pruebas de caja blanca*. Recuperado el 01 de 05 de 2020, de https://www.ecured.cu/Pruebas_de_caja_blanca

sparkfun. (s.f.). *TECHNICAL DATA MQ-7 GAS SENSOR*. Recuperado el 03 de 05 de 2020, de <https://www.sparkfun.com/datasheets/Sensors/Biometric/MQ-7.pdf>

VALLA, E. J. (2018). *IMPLEMENTACIÓN DE UN PROTOTIPO ELECTRÓNICO INALÁMBRICO PARA SUPERVISIÓN, AVISO Y CONTROL DE GASES CONTAMINANTES EN INVERNADEROS ARTESANALES*. Obtenido de <http://dspace.esPOCH.edu.ec/bitstream/123456789/9209/1/108T0252.pdf>

Consejería de Salud de la Región de Murcia. (s.f.). *Monóxido de carbono*. Recuperado el 03 de 05 de 2020, de <https://www.murciasalud.es/pagina.php?id=180398&idsec=1573>

Grand Island Fire Department. (s.f.). *CO Los niveles y los efectos*. Recuperado el 03 de 05 de 2020, de <https://www.grand-island.com/home/showdocument?id=16269>

medicina.usac.edu.gt. (2005). (*ppm*), *Partes por millón*. Obtenido de

[http://medicina.usac.edu.gt/quimica/soluciones/Partes_por_mill_n_\(ppm\).htm](http://medicina.usac.edu.gt/quimica/soluciones/Partes_por_mill_n_(ppm).htm)

sparkfun-MQ7. (s.f.). *www.sparkfun.com*. Recuperado el 04 de 05 de 2020, de

<https://www.sparkfun.com/datasheets/Sensors/Biometric/MQ-7.pdf>

MSA de CHILE. (s.f.). *Prevención de Riesgos*. Recuperado el 04 de 05 de 2020, de

https://www.paritarios.cl/prevencion_de_riesgos_Amoniacos.htm

health, N. J. (s.f.). Recuperado el 04 de 05 de 2020, de Hoja Informativa sobre Sustancias

Peligrosas:

<https://www.nj.gov/health/eoh/rtkweb/documents/fs/0084sp.pdf>

TOLUENO, H. D. (s.f.). *HOJA DE SEGURIDAD XVII TOLUENO*. Recuperado el 04 de 05 de

2020, de <https://quimica.unam.mx/wp-content/uploads/2016/12/17tolueno.pdf>