

Modelo para la detección e identificación de fallas, basado en un algoritmo de inteligencia artificial, para el horno H150 en el departamento de refinación de crudos de la Refinería de Barrancabermeja.

Néstor Francisco Hernández Fernández

Trabajo de Grado para Optar al Título de Magister en Gerencia de Mantenimiento

Director

Heller Guillermo Sánchez Acevedo

Coordinador Posgrados en Gerencia de Mantenimiento UIS

Universidad Industrial de Santander

Escuela de Ingeniería Mecánica

Posgrado en Gerencia de Mantenimiento

Barrancabermeja

2022

### **Dedicatoria**

A Dios por darme las facultades para ejercer mi profesión y las bendiciones que de él recibo que me permiten prepararme para mejorar.

A mi Esposa Carol Bareño.....por su apoyo incondicional, su paciencia y comprensión, en cada fase de mi vida, a Juan Diego, mi Hijo..... por su Amor y el tiempo para el desarrollo de este proyecto que será de ejemplo en el crecimiento profesional de su padre. Ellos son lo más importante y la más grande motivación para superarme y mejorar cada día.

### **Agradecimientos**

A Dios por la salud y perseverancia en cada meta trazada para mi crecimiento profesional, laboral y personal, que me brinda la vida.

Al ingeniero Héller Sánchez por su apoyo constante y su atención durante el desarrollo de este proyecto.

A Ecopetrol SA por el apoyo para la realización de esta Maestría aportando para mi vida personal y profesional conocimientos que se pueden colocar al servicio de la comunidad.

## Tabla de Contenido

	<b>Pág.</b>
Introducción.....	12
2. Objetivos.....	14
2.1 Objetivo General .....	14
2.2 Objetivos Específicos .....	14
3. Planteamiento del Problema .....	15
3.1 Antecedentes .....	15
3.2 Problema de Investigación.....	18
3.3 Definición del Problema .....	19
3.4 Justificación .....	21
4. Marco Teórico.....	26
4.1 Análisis de componentes principales (PCA) .....	26
4.2 Algoritmo de inteligencia artificial .....	30
5. Metodología del Proyecto.....	36
5.1 Diseño de la investigación .....	36
5.2 Muestra .....	38
5.3 Técnicas de recolección de Datos .....	38
6.1 Fase 1: Recolección de datos .....	41
6.2 Fase 2: Preparación de los datos. ....	45
6.3 Fase 3: Ingreso de datos.....	46
6.4 Fase 4: Procesamiento de datos. ....	52

Para el desarrollo de esta fase, se toman en cuenta todas las variables dependientes e independientes del proceso. Después de tener 12 variables del serpentín B, al realizar la correlación de variables, los resultados muestran que no hay correlación entre ellas. Para el análisis de componentes principales se trabaja con todas las variables que intervienen en el proceso que son las presentadas en la Tabla 8. ....52

*Nota.* Variables del serpentín B (12) del flujo de gas a quemadores. Tomado del Análisis de los Componentes Principales (ACP). ....53

6.5 Fase 5: Interpretación de los datos y Entrenamiento de la Red SOM.....56

6.6 Fase 6: Almacenamiento de datos.....59

7. Análisis del Desarrollo del Proyecto.....60

8. Conclusiones.....64

9. Recomendaciones.....65

Referencias .....66

**Lista de Tablas**

	<b>Pág.</b>
Tabla 1. Producción de la Refinería de Barrancabermeja .....	16
Tabla 2. Inversiones de los proyectos en la Refinería de Barrancabermeja .....	17
Tabla 3. Variables de proceso del Horno H-150 .....	42
Tabla 4. Datos atípicos .....	45
Tabla 5. Cálculo de correlación con la presión de entrada serpentín B. ....	47
Tabla 6. Cálculo de correlación con la variable de presión entrada de crudo .....	48
Tabla 7. Coeficiente de correlación con la variable de temperatura de salida serpentín B .....	48
Tabla 8. Coeficiente de correlación con la variable de temperatura del tubo 1 serpentín B .....	49
Tabla 9. Coeficiente de correlación con la variable de temperatura del tubo 12 serpentín B .....	49
Tabla 10. Coeficiente de correlación con la variable de temperatura de choque serpentín B .....	50
Tabla 11. Coeficiente de correlación con la variable de temperatura de la cámara de combustión serpentín B .....	50
Tabla 12. Coeficiente de correlación con la variable de temperatura de gases de chimenea .....	51
Tabla 13. Coeficiente de correlación con la variable de temperatura del techo inclinado .....	51
Tabla 14. Variables de proceso para PCA .....	52
Tabla 15. Resultados del PCA datos en control .....	53
Tabla 16. Resultados del PCA datos por fuera de las ventanas operativas .....	53
Tabla 17. Porcentaje de información en cada componente .....	55
Tabla 18: Componentes principales utilizadas para entrenamiento de la red .....	56
Tabla 19. Datos actualizados para Test .....	60

## Lista de Figuras

	<b>Pág.</b>
<b>Figura 1.</b> Refinería de Barrancabermeja .....	15
<b>Figura 2.</b> Unidad de destilación U-150 .....	18
El último de coquizado efectuado fue en la parada de septiembre de 2022 donde se encontró fragmentos de coque de forma generalizada de hasta 10 cm como se muestra en la Figura 3. ....	19
<b>Figura 3.</b> Muestra Fragmento Coque Serpentes H150 .....	20
Allí, estudios en laboratorio asocian la causa del coquizamiento prematuro a la precipitación de asfáltenos por inestabilidad coloidal de la mezcla liviana parafínica cargada a la unidad, como lo muestra la Figura 4, vista a través de microscopio para el crudo cargado en la U150 donde se evidencia la alta precipitación de asfáltenos.....	21
<b>Figura 4.</b> Precipitación alta de asfáltenos U150 .....	22
<b>Figura 5.</b> <i>Tendencia IC Serpentes H150 2017 a 2019.</i> .....	23
<b>Figura 6.</b> Tendencia Temperatura Piel de Tubo H150.....	23
<b>Figura 7.</b> Tendencia Temperatura Gases de Chimenea H150 .....	24
<b>Figura 8.</b> Mapa auto-organizado de Kohonen (SOM) .....	32
<b>Figura 20.</b> Pantalla de control del DCS.....	42
<b>Figura 21.</b> Control DCS del H-150 .....	43
<b>Figura 22.</b> Información del PI DATA LINK.....	44
Nota. Se observa en la Figura 22 el archivo de Excel al descargar la información de PI DATA LINK, permitiendo ubicar toda la instrumentación de la planta por el número de tag que tiene cada	

instrumento, y el historial de datos que se requiera y en que intervalos de tiempo entre dato y dato.

Tomado del complemento de MS Excel PI DATA LINK .....	44
<b>Figura 23.</b> Componentes principales PCA .....	54
<b>Figura 24.</b> Cargue de los datos en la red SOM.....	57
<b>Figura 25.</b> Pesos de las neuronas cercanas .....	57
<b>Figura 26.</b> Pesos de cada una de las entradas .....	58
<b>Figura 27.</b> Evaluación del modelo según los residuos .....	58
<b>Figura 28.</b> Rotación posiciones de neuronas y pesos de la información en 3D. ....	59
<b>Figura 29.</b> Datos y resultados en el gráfico de respuesta .....	63

## Glosario

**DCS (Sistema de Control Distribuido de proceso):** consiste, en un sistema de control automatizado conformado con elementos de control instalados de forma geográfica en la planta o en el área que se desea controlar. Este sistema de control de proceso, es la interacción entre el proceso y el operador, el cual toma valores exactos del proceso.

**Horno de crudo H150:** equipo de refinación de crudo cuya función es calentar el producto llevándolo a una temperatura promedio entre 650 F y 750 F, para generar despojo de fracciones livianas en la torre de destilación.

**Inteligencia artificial:** hace parte de la investigación y la informática buscando diversas formas de imitar el funcionamiento del cerebro humano y sus neuronas con las máquinas para aportar en la solución de problemas imitando procesos de la inteligencia humana mediante mecanismos lógicos y matemáticos.

**PI Data Link:** repositorio de información, herramienta de enlace, que trabaja con la aplicación de Excel, toma los datos del DCS y los almacena para análisis operacional tomando valores promedio.

**Redes Neuronales:** son sistemas de información inspirados en el comportamiento de las redes neuronales biológicas, o del ser humano, que consta de un conjunto de procesamiento llamado neuronas las cuales están conectadas mediante un valor número modificable llamado peso.

## Resumen

**Título:** Modelo para la detección e identificación de fallas, basado en un algoritmo de inteligencia artificial, para el horno h150 en el departamento de refinación de crudos de la Refinería de Barrancabermeja \*

**Autor:** Néstor Francisco Hernández Fernández \*\*

**Palabras Clave:** Confidencialidad de activos, PCA, Algoritmos de Inteligencia Artificial Industrial, Mantenimiento de Planta y Equipo

### Descripción:

Esta investigación, se elaboró para analizar y dar respuesta a la necesidad de identificar las fallas presentadas en el horno H150, implementando un algoritmo de inteligencia artificial, basado en redes neuronales que permita la identificación de fallas y comportamiento del horno, seguido de la discriminación y análisis de datos estadísticos arrojados, para finalizar, con la presentación del modelo que se llevó a cabo, en el que se utilizó el análisis de componentes Principales (PCA), para analizar el comportamiento de las variables de entrada y salida en el tiempo, para el horno H150 de la refinería de Barrancabermeja.

---

\* Trabajo de Grado

\*\* Escuela de Ingeniería Mecánica. Posgrado Gerencia de Mantenimiento. Director: Héller Guillermo Sánchez Acevedo UIS

## Abstract

**Title:** Model for the detection and identification of failures, based on an artificial intelligence algorithm, for the H150 furnace in the crude refining department of the Barrancabermeja refinery.

\*

**Author(s):** Nestor Francisco Hernández Fernández\*\*

**Key Words:** Asset Reliability, PCA, Artificial Intelligence Algorithms, Industrial Production, Plan and Equipment Maintenance.

### Description:

This research was developed to analyze and respond to the need to identify the failures presented in the H150 furnace, implementing an artificial intelligence algorithm based on neural networks that allows the identification of failures and furnace behavior, followed by the discrimination and analysis of statistical data, and finally, with the presentation of the model that was carried out, in which the Principal Components Analysis (PCA) was used to analyze the behavior of the input and output variables over time, for the H150 furnace of the Barrancabermeja refinery.

---

\* Undergraduate Work

\*\* School of Mechanical Engineering. Graduate Degree in Maintenance Management. Director: Héller Guillermo Sánchez Acevedo UIS

## Introducción

El presente documento ha sido elaborado en el marco de la tesis de maestría en gerencia de mantenimiento. Su desarrollo corresponde a un ejercicio académico inicial en la exploración y búsqueda de soluciones a problemas industriales empleando herramientas computacionales que conduzcan a hacer más eficiente la tarea de identificación y diagnóstico de fallas en los activos productivos de una organización. A nivel de este proceso el trabajo se enfoca en una exploración inicial que permite entender el funcionamiento de herramientas como el análisis de sistemas de componentes y las redes neuronales artificiales. En el documento se reportan los resultados de dicha exploración, y por lo tanto se recomienda al lector, analizarlo desde dicha perspectiva.

La ciencia y la tecnología se han constituido como el pilar de la vida humana, para el desarrollo cultural, social y económico de la sociedad moderna en este siglo XXI. La ciencia aporta el conocimiento y la tecnología adopta formas materiales específicas, generando productos que completan y adecuan la solución a la necesidad del hombre, para que sus actividades sean más eficaces, eficientes y fiables; llevándolo a ser más competitivo en este mundo moderno. Lo mencionado anteriormente, genera a su vez, la necesidad de contar con empresas competitivas que reporten altos estándares de calidad, productividad, fiabilidad, seguridad y viabilidad ambiental a bajo costo, para sostenerse en el mercado.

Este desafío plantea una apuesta obligada: el uso de la Inteligencia Artificial (IA), llevando a las empresas y al mundo a la era digital, con programas de cálculo inteligente que benefician la capacidad de procesamiento, el análisis de gran volumen de datos, que permiten detectar y predecir futuros problemas, reducción de costos, mayor precisión en la toma de decisiones y visualizar el desempeño en tiempo real de sus activos. En ese sentido, el especialista (Rovisora, 2020) aborda

para el sector petróleo & gas dos (2) aplicaciones, el Data Science que estudia los datos combinando el área de la estadística, con el apoyo de las matemáticas y la informática para mejorar la toma de decisiones y Machine Learning o Aprendizaje Automático, que permite a una maquina o software identificar patrones de comportamiento entre millones de datos, prosperando la ejecución de las operaciones de forma rápida y eficiente.

Gracias a esto el presente trabajo de investigación, pretende plasmar los resultados luego de la aplicación de las técnicas estadísticas y los modelos matemáticos (PCA), para el desarrollo de un sistema de detección e identificación de fallas y dar solución a un problema en una de las plantas del área de refinación de crudos, el horno H150 de la refinería de Barrancabermeja. Esta técnica estadística multivariante de detección de fallas PCA (Análisis de Componentes Principales), se usa para describir en nuevas variables el conjunto de los datos históricos del horno H150 reduciendo la dimensional de los mismos. De esta manera se busca detectar condiciones anormales de operación, llevando estas nuevas variables a un modelo matemático basado en redes neuronales.

## **2. Objetivos**

### **2.1 Objetivo General**

Plantear un modelo basado en redes neuronales para detectar e identificar los modos de falla del horno H-150, empleando un algoritmo basado en el análisis de componentes (PCA) y una red neuronal, en el departamento de refinación de crudos de la refinería de Barrancabermeja.

### **2.2 Objetivos Específicos**

Recopilar la información de todas las variables posibles que afectan el funcionamiento del horno en los últimos 3 años

Analizar el comportamiento de las variables de entrada y salida en el tiempo mediante el análisis de componentes principales (PCA).

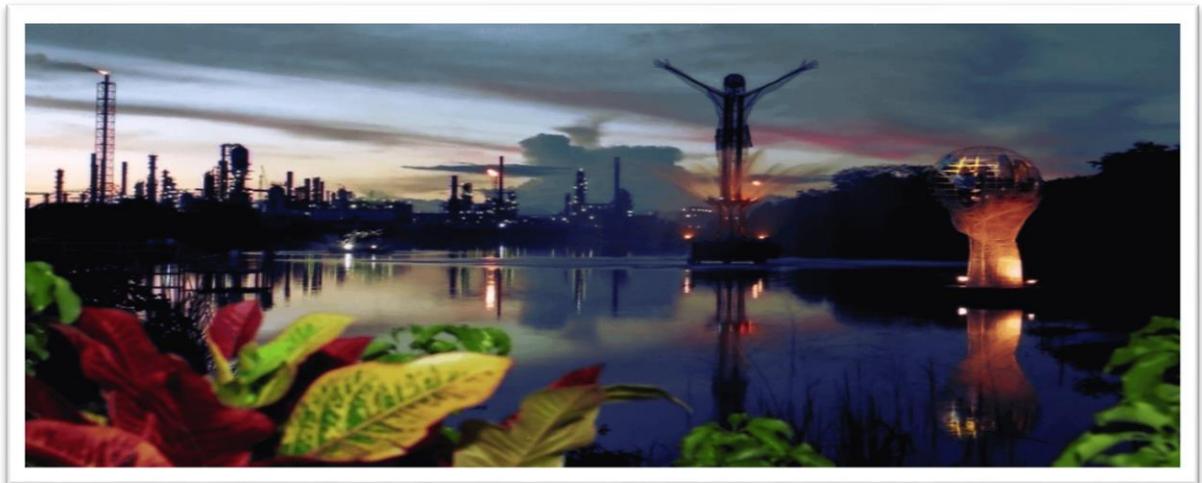
Implementar un algoritmo de inteligencia artificial, basado en redes neurales que permita la identificación de los modos de falla del horno H150 del departamento de refinación de crudos de la refinería de Barrancabermeja.

### 3. Planteamiento del Problema

#### 3.1 Antecedentes

La Figura 1, ilustra las 254 hectáreas en las que se extiende la Refinería de Barrancabermeja, a orillas del río Magdalena del Distrito de Barrancabermeja, de propiedad de la empresa (Ecopetrol, 2022) , una de las primeras zonas en las que comenzó la explotación de hidrocarburos en Colombia. Empezó sus operaciones en 1922 con unos alambiques traídos desde Talara en el Perú y con una capacidad instalada inicial de 1.500 barriles por día, dando inicio a la construcción de esta compleja infraestructura.

*Figura 1.* Refinería de Barrancabermeja



Fuente. Portal web Ecopetrol SA. Tomado de: <https://nuevoportal.ecopetrol.com.co/wps/portal/es/ecopetrol-web/nuestra-empresa/quienes-somos/lo-que-hacemos/refinacion/complejo-barrancabermeja> Citado (15 de julio 2022)

Ecopetrol, resalta que la importancia de esta refinería es su producción basada en una carga de crudo de 230 KBD, la refinación y exportación de crudos y petroquímica, así lo confirma los

estudios de benchmarking internacional realizados por la firma internacional (Salomon Associates, 2020) quien la avaló, como una de las mejores refinerías en Latinoamérica, por su desempeño, eficiencia y rentabilidad, gracias a su producción, inversión, y evolución permanente tanto en investigación como en uso de nuevas tecnologías en las operaciones de sus procesos industriales de bienes y servicios.

**Tabla 1.** Producción de la Refinería de Barrancabermeja

Productos	Producción (*), kbd	Usos
Gasolinas	60,0	Consumo vehículos automotores
Diesel	65,0	Consumo vehículos automotores, transporte público y pesado
Jet	23,0	Combustible para aviones turbina o tipo jet
Fuel Oil	20,0	Combustible industrial y marinos
GLP	6,0	Combustible en estufas domésticas y equipo industrial
PGR	3,5	Precursor industria petroquímica (polipropileno)
Disolventes Aromáticos (kb/m)	90	Solventes
Bases lubricantes (kb/m)	55	Aceites lubricantes para motor
Parafinas (ton/m)	3500	Velas, betún, papel parafinado
Disolventes Alifáticos (kb/m)	28	Pinturas, precursores petroquímicos
Polietileno (ton/m)	4000	Plásticos de baja densidad
Azufre (ton/m)	2400	Fertilizantes
Acido (ton/m)	2100	Industria, fertilizantes

Fuente Portal web Ecopetrol SA. Tomado de: <https://nuevoportal.ecopetrol.com.co/wps/portal/es/ecopetrol-web/nuestra-empresa/quienes-somos/lo-que-hacemos/refinacion/complejo-barrancabermeja> Citado (15 de julio 2022)

La Tabla 2 muestra las inversiones de mega obras ejecutadas por Ecopetrol SA, para la actualización tecnológica, repotenciación y automatización de sus activos, en la Refinería de Barrancabermeja, cumpliendo con estándares nacionales e internacionales y así abastecer el mercado nacional produciendo diésel, jet, GLP, gasolina, propileno, asfalto, aromáticos, bases lubricantes, etc; y con el mercado internacional, exportando petroquímicos y excedentes de fuel oil.

**Tabla 2.** Inversiones de los proyectos en la Refinería de Barrancabermeja

AÑO	OBRAS	AÑOS	OBRAS
30 a 50	Construcción de las primeras Plantas industriales de destilación Combinada, primera unidad de Cracking catalítico, (modelo IV) Y primera unidad visco reducción	2000	Construcción de la nueva unidad de planta de ácido sulfúrico, reemplazo hornos de la U250, modernización Unidad filtración parafinas, inicio construcción plantas inicio construcción plantas hidrot ratamiento de Diesel y gasolina, nueva planta de Hidrógeno y nueva planta de de Azufre IV
60	Construcción de 2 unidades crudo y vacío (U250 y U2100), así como la segunda de Cracking (unidad Orthoflow) y primera unidad de Polietileno	2010	Puesta servicio Unidades de Hidrot ratamiento, automatización total de la Refinería, montaje del centro de optimización, modernización servicios industriales U5100, repotenciación unidad de Balance, montaje nuevas unidades de comprensión de aire industrial, nueva planta Ultra-Filtración, recuperación de la unidad de turbo U250
70	Montaje del tren de lubricantes y Parafinas, la planta de aromáticos y El conjunto de unidad de balance	2007 y 2017	Unidades de hidrot ratamiento de combustibles, recuperación de la unidad de turbo expande almacén gas licuado de petróleo y ampliación de capacidad de Producción de asfalto.

*Nota.* Esta Tabla muestra las inversiones efectuadas en la Refinería de Barrancabermeja.

Tomado de: <https://nuevoportal.ecopetrol.com.co/wps/portal/es/ecopetrol-web/nuestra-empresa/quienes-somos/lo-que-hacemos/refinacion/complejo-barrancabermeja> Citado (15 de julio 2022)

### 3.2 Problema de Investigación

El proceso de Refinación y Petroquímica juega un papel relevante en la cadena de valor para Ecopetrol SA y Colombia, al transformar el petróleo crudo en diversos productos. En ese sentido, la (Refinería Barrancabermeja, 2018), a través del departamento de refinación de crudos, cuenta con cinco plantas Topping y dos torres enfriadoras, en la ilustración 4 se muestra la unidad de destilación U-150, que posee un esquema único, debido a que su diseño puede producir bases parafínicas a partir del crudo liviano, con una capacidad de 38 KBD.

**Figura 2.** Unidad de destilación U-150



Fuente. Fotografía tomada Néstor Hernández en Refinería Barrancabermeja agosto 2 de 2018

La U-150 es una de las plantas del departamento, en donde su proceso consiste en generar las bases lubricantes y otros productos, además, aporta en la producción de nafta pesada, nafta liviana, jet y ACPM; hacen parte en la operación de esta unidad de refinación el Horno atmosférico H150, cuyo objetivo es aumentar la temperatura del crudo al punto de flasheo, 680° F aproximadamente y de esta manera generar un buen despojo de vapores en la torre atmosférica.

El tiempo de corrida de este equipo es de 5 a 6 años, antes de realizar el mantenimiento mayor de la unidad, programado en esos mismos intervalos de tiempo. No obstante, en los últimos años, ha disminuido considerablemente este tiempo de corrida, debido a los cambios en las

composiciones del crudo y la variación en la calidad del gas combustible, generando en la tubería interna (serpentines) la formación de coque en las paredes de los mismos, reduciendo su diámetro, aumentando la temperatura de piel de tubos y produciendo la deflexión y rotura de los mismos.

Enmarcados en este contexto, surge la pregunta de investigación ¿Cómo detectar e Identificar los modos de falla, empleando un algoritmo basado en el análisis de componentes (PCA) y una red neuronal del horno H150? ¿Qué modos de fallas se pueden identificar empleando un algoritmo basado en el análisis de componentes (PCA) y una red neuronal, del horno H150 de la Refinería de Barrancabermeja?

### **3. 3 Definición del Problema**

El horno H150 ha presentado desde el año 2013 coquizamiento prematuro, aumentando las temperaturas de piel de tubo  $> 1300^{\circ}\text{F}$  por encima de la referencia de integridad en la zona de choque y convección, requiriendo limpieza mecánica con pigging con una frecuencia de 2 a 3 años durante el ciclo de corrida. Por lo que, el departamento de refinación de crudos de la GRB (Gerencia Refinería Barrancabermeja) ha visto la necesidad de realizar de coquizado de los serpentines cada 3 años, impactando la carga del departamento, la operación de refinación, los costos de producción, afectando los indicadores de esta por los días de parada no programada, y mantenimiento, incumpliendo en los planes de producción en los productos blancos.

El último de coquizado efectuado fue en la parada de septiembre de 2022 donde se encontró fragmentos de coque de forma generalizada de hasta 10 cm como se muestra en la Figura 3.

**Figura 3.** Muestra Fragmento Coque Serpentina H150

Fuente. Adoptado de formato Recomendación técnica (GRB-GTE-F-001) Ecopetrol SA

Estos modos de falla se han visto reflejados en la disminución del flujo de carga en el horno, en el aumento del porcentaje de apertura de las válvulas de control de manera automática, mayor consumo de gas en el horno para mantener la temperatura en el hogar, rompimiento o deflexión de los tubos de los serpentines y afectación en el proceso de la planta de Parafinas. En consecuencia, la afectación en la producción requerida de los productos blancos y los destilados parafínicos, riesgos en la calidad y seguridad, en la rentabilidad y optimización de la confiabilidad operacional en Ecopetrol SA.

### **3.0 Hipótesis**

El uso de la inteligencia artificial hace posible predecir los modos de falla de los serpentines de crudo en el horno H-150 en el departamento de refinación de crudos en la refinería de Barrancabermeja, aprovechando la cantidad de data recopilada de las variables de proceso que intervienen en la operación del equipo.

### 3.4 Justificación

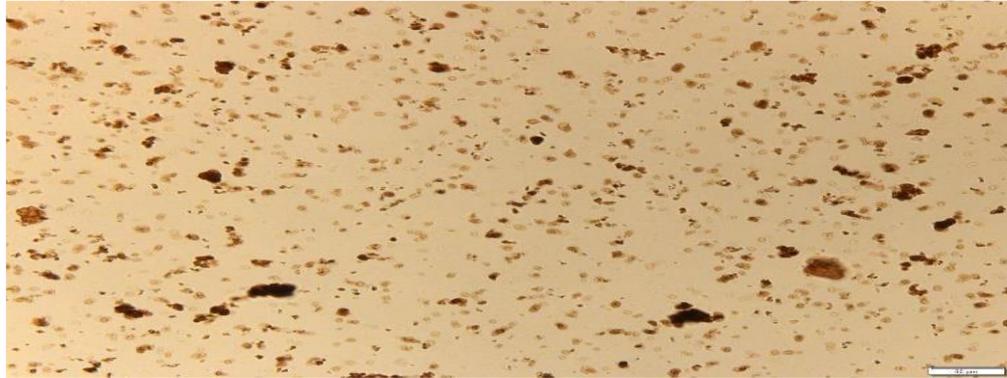
Para atender el abastecimiento de la demanda de consumo nacional e internacional y a estándares internacionales, la Gerencia Refinería de Barrancabermeja cuenta con instalaciones, plantas, equipos, sistemas, procesos normalizados y funciones, que dan soporte a las operaciones industrial del sector de hidrocarburos, a la cual pertenece (exploración, perforación, producción, refinación, almacenamiento, transporte, distribución y comercialización). El mal funcionamiento de los equipos y plantas, que dan soporte a estas operaciones industriales, puede provocar pérdidas económicas, incumplimiento en la producción, situaciones de riesgos, peligro para los operarios e inconvenientes para los consumidores, entre otros; evidenciando que los sistemas de seguridad para estos procesos y confiabilidad de los activos se hacen cada vez más oportunos y necesarios, por su complejidad, tal como lo afirma (Amad, 2021).

Lo anterior, respalda el desarrollo de esta investigación que busca anticipar e identificar los modos de fallo, mediante el seguimiento de las diferentes variables de proceso, revisando su comportamiento en el tiempo e identificando las variaciones de dichas variables que puedan llevar a una falla de acuerdo a sus históricos en los últimos 5 años; y el uso de en un algoritmo de inteligencia artificial, para los cambios en la operación, que viene presentando el horno atmosférico H150, y así, lograr reducir o minimizar a 0 los riesgos de fallas, evitando un deterioro continuo al interior del horno. También lo avala, el estudio técnico realizado en julio del 2019 por la Gerencia Técnica (GTE) de la Gerencia Refinería Barrancabermeja GRB, plasmado en el Formato de recomendación técnica (GRB-GTE-F-001) dirigido al Departamento de Mantenimiento.

Allí, estudios en laboratorio asocian la causa del coquizamiento prematuro a la precipitación de asfáltenos por inestabilidad coloidal de la mezcla liviana parafínica cargada a la

unidad, como lo muestra la Figura 4, vista a través de microscopio para el crudo cargado en la U150 donde se evidencia la alta precipitación de asfáltenos.

**Figura 4.** Precipitación alta de asfáltenos U150



Fuente. Adoptado de formato Recomendación técnica (GRB-GTE-F-001) Ecopetrol SA

Y para evaluar el grado de coquización de un horno, se valen del uso de índices de coquización (IC), cuya fórmula es  $IC = 106 \Delta P / Q^2$

Dónde:

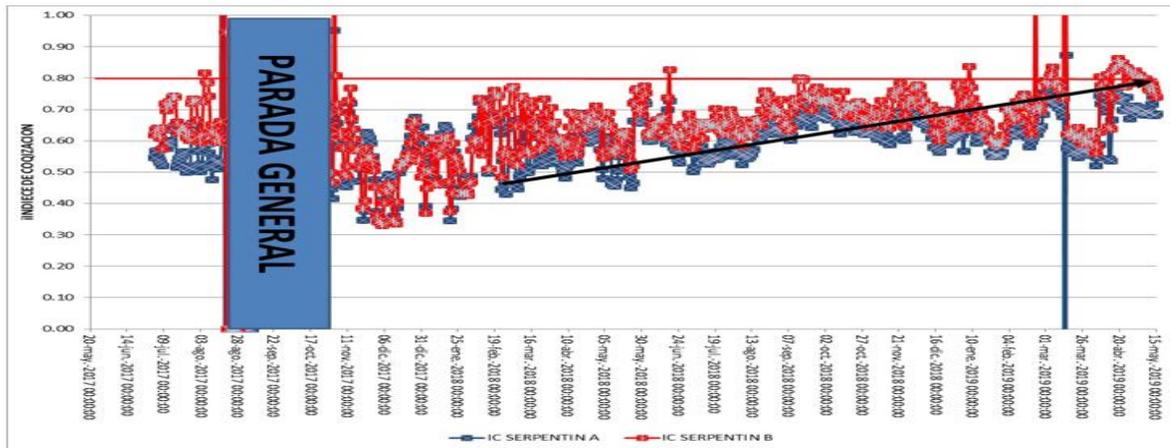
IC = Índice de Coquización [adimensional]

$\Delta P$  = Caída de presión en el serpentín o sección de serpentín a evaluar [psi]

Q = Flujo estándar a través del serpentín o sección de serpentín a evaluar [BPD@60°F]

La Figura 5, muestra la tendencia del índice de coquización del H150 desde su puesta en servicio luego de la reparación en septiembre de 2017. El índice (IC) es la prueba realizada a un horno para evaluar el grado de coquización, cuya fórmula esta descrita anteriormente, para esta prueba, alcanzó un valor de 0.8, que corresponde al valor máximo de referencia para este horno (IC eje vertical, eje horizontal: serpentín A, serpentín B).

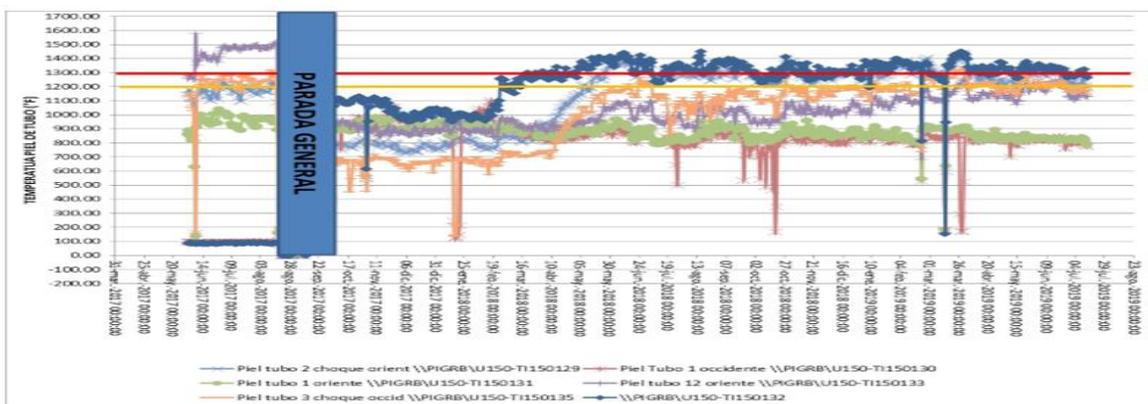
**Figura 5.** Tendencia IC Serpentina H150 2017 a 2019.



Fuente. Adoptado de formato Recomendación técnica (GRB-GTE-F-001) Ecopetrol SA

De igual manera para validar el diagnóstico la Figura 6, muestra la tendencia de las temperaturas de piel de tubo donde se observa que la TIT150129 y TIT150133 están por encima de la ventana de integridad definida < 1300°F.

**Figura 6.** Tendencia Temperatura Piel de Tubo H150.

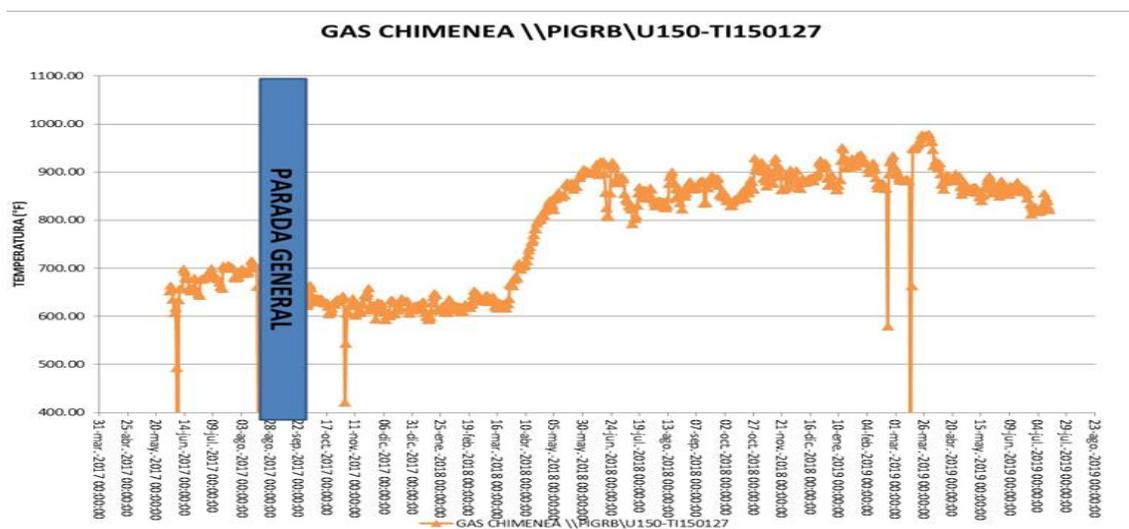


Fuente. Adoptado de formato Recomendación técnica (GRB-GTE-F-001) Ecopetrol SA

Para ello un ingeniero estático del área, empleó las termografías para validar las temperaturas de piel de tubo, utilizando una cámara termo gráfica, y así confirmar los datos arrojados por el DCS y detectar puntos calientes a lo largo de los serpentines.

La tendencia del aumento en la temperatura de gases de chimenea  $>800^{\circ}\text{F}$  indicando baja transferencia de calor en la zona de convección, confirmando así la pérdida de eficiencia del horno, se evidencia en la siguiente ilustración.

**Figura 7.** Tendencia Temperatura Gases de Chimenea H150



Fuente. Adoptado de formato Recomendación técnica (GRB-GTE-F-001) Ecopetrol SA

El desarrollo de la investigación busca la anticipación e identificación de los modos de fallo mediante el seguimiento de las diferentes variables de proceso, revisando su comportamiento en el tiempo e identificando las variaciones de dichas variables que puedan llevar a una falla de acuerdo a sus históricos en los últimos 5 años; de esta manera se podrán realizar ajustes operacionales, buscando alargar el ciclo de trabajo del activo.

Por otra parte, es importante conocer la operación en el tiempo de dicho equipo para prevenir y planear el recurso, los planes de producción y la estrategia de cargas en la refinería, y

evaluar en que parte del ciclo de operación es pertinente la ejecución de un mantenimiento sin incurrir en pérdidas mayores por la generación de un daño mayor, como también el aprovechamiento del tiempo de parada para la realización de actividades pendientes de mantenimiento también en la planta de parafinas la cual es afectada directamente si la unidad 150 sale de servicio.

El sistema de control de proceso (DCS), junto con PI DATA LINK, almacena toda la información de las señales de proceso de datos históricos que se maneja en los equipos y permite descargarla en un archivo de Excel, gracias a esto podemos utilizar la ayuda tecnológica, utilizar esta información y predecir de acuerdo con los históricos de falla cuando el equipo está entrando en una condición que requiera de mayor atención.

## 4. Marco Teórico

A continuación, se realiza una recopilación de las teorías que constituyen el soporte teórico para desarrollar la investigación de un modelo de detección e identificación de fallas basadas en un algoritmo de inteligencia artificial.

### 4.1 Análisis de componentes principales (PCA)

Algunos autores coinciden que el análisis en componentes principales (PCA por sus siglas en inglés, Principal Component Análisis) es una técnica estadística multivariada de gran uso en el monitoreo de procesos industriales, (Rodrigo, 2017).

Una justificación del uso de PCA en el monitoreo de procesos industriales, se da por la globalización y normatividad medioambientales que trajo las últimas décadas, causando competencia en el sector industrial y desvelos por el control, la calidad y la perfección de la producción en sus procesos industriales; ya que una falla en el funcionamiento de un proceso debe ser identificada en el momento cuando la falla todavía es pequeña y su impacto no es relevante al proceso, concluye (Rozo, 2008).

Según Rodrigo “el análisis de componentes principales (PCA) es una técnica estadística que nos permite describir un gran número o conjunto de datos en nuevas variables no correlacionadas, reduciendo la dimensionalidad de dicho conjunto de datos. Si tenemos una muestra con  $n$  individuos cada uno con  $p$  variables ( $X_1, X_2, X_p$ ), entonces, el espacio muestral tiene  $p$  dimensiones. Aplicar PCA brinda la posibilidad de convertir un conjunto de datos de variables de proceso generalmente correlacionadas en un nuevo conjunto de datos de dichas variables sin correlación lineal llamadas componentes principales.”. (Rodrigo, 2017)

PCA es una de las técnicas del aprendizaje no supervisado que tiene como objetivo obtener información usando predictores, buscando así identificar subgrupos. Unos de los inconvenientes en el uso de las técnicas de aprendizaje no supervisado es la validación de los resultados del análisis, puesto que no se tienen variables de respuesta que permita verificarlos.

Los métodos de supervisión están clasificados en dos categorías principales: los basados en modelos físicos del proceso, y aquellos creados con los datos históricos. Las técnicas basadas en datos con frecuencia usan la reducción de la medida en la data original. Con el fin de tener la mayor cantidad de información original del proceso se requiere eliminar información atípica en los valores arrojados por las variables.

Cuando se tiene un conjunto de datos de variables nutrido en cantidad y calidad se utiliza el PCA, buscando determinar las causas de fallas en el proceso, la prevención de dichas fallas, mejorar el control de calidad, la identificación y modelización de sistemas, la implementación de reglas y patrones de comportamiento, realizar la búsqueda de relaciones entre variables, etc. concluye Rozo en su estudio desarrollado de detección y diagnóstico de fallas en procesos industriales utilizando análisis de componentes principales (PCA).

#### **4.1.1 Finalidad del Análisis de componentes principales (PCA)**

El objetivo principal de PCA es modificar un conjunto de variables correlacionadas en un conjunto nuevo reducido de variables no correlacionadas llamadas componentes principales. Sea  $X$  ( $N \times M$ ) una matriz de orden  $p$ -dimensional conformada por  $p$  variables con  $N$  observaciones por variable, perteneciente a la organización de los datos originales.

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1M} \\ X_{21} & X_{22} & \dots & X_{2M} \\ \dots & \dots & \dots & \dots \\ X_{N1} & X_{N2} & \dots & X_{NM} \end{bmatrix}$$

PCA logra encontrar el número de componentes principales como número de variables originales haya, y dentro del resultado del análisis selecciona un nuevo número de variables (llamadas variables latentes) expresadas en un matriz resultante, disminuyendo la menor variabilidad posible existente en el conjunto de datos original. El nuevo conjunto de variables define un espacio k-dimensional mucho más reducido que el original donde  $k < M$ , describe Florelva.

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{1k} \\ Z_{21} & Z_{22} & \dots & Z_{2k} \\ \dots & \dots & \dots & \dots \\ Z_{N1} & Z_{N2} & \dots & Z_{Nk} \end{bmatrix}$$

**Las siguientes son las variables latentes que aparecen:**

- la componente principal  $t_1$  es la primera de las variables latentes y esta, es el resultado de la combinación lineal de las variables originales.

$t_1 = Xp_1$  (3) Que tiene mayor varianza, ¿esta se expresa para el conjunto total de individuos como  $t_1 = Xp_1$  (4)

- **Cargas (loads)** Donde  $p_1$  es el vector propio correspondiente al mayor valor propio  $p_1$  de la matriz de covariancias de  $X$ ,  $S_x$ , o de la matriz de correlación  $R$  de  $X$ ; y define la dirección de mayor variabilidad en el espacio de las  $M$  variables. Las componentes del

vector **p1** son llamadas cargas (loads), definen la orientación del hiperplano obtenido, con respecto al espacio de las variables originales, igualmente indican la magnitud y sentido de la correlación de la componente considerada con cada una de las variables originales.

- **Puntuaciones (scores)** El vector t1 de scores antes definido, representa la proyección de cada observación sobre la dirección definida por p1. ¿Además, la variancia del score t1 es igual a p1.
- **Segunda componente** Combinación lineal de las variables originales que se obtiene de forma tal que sea ortonormal a la primera, con la mayor varianza posible, inferior a la de t1. Con los mismos criterios hallamos las restantes componentes principales.

$t_2 = Xp_2$  (5) Si se extraen todas las componentes principales,  $T = XP$  (6)

Donde T es la matriz (NxM), cuyas columnas son los p vectores de scores y P es la matriz (MxM) cuyas columnas son los p vectores propios que constituye una base ortonormal. Post.multiplicando por P' se tiene que:  $X = TP'$

De acuerdo a esto el método de PCA permite reducir la dimensionalidad de la información adquirida de las variables del proceso y expresar dicha información en unas pocas componentes. Su aplicación junto con diferentes técnicas estadísticas como el clustering y la regresión ha sido relevante teniendo en cuenta la importancia de la cantidad y calidad de las variables originales para calcular las componentes.

Con base en estos planteamientos teóricos, se desarrolló el PCA utilizando las variables identificadas en la operación del horno 150, a través de la herramienta Excel, la cual permite calcular el porcentaje de información en cada una de las componentes.

## 4.2 Algoritmo de inteligencia artificial

Según Ponce y sus amigos, la inteligencia artificial es un área multidisciplinaria de la ciencia donde se construyen sistemas cuyo objetivo es el de realizar tareas y resolver problemas como lo realiza un ser humano, simulando de manera artificial la manera de actuar del cerebro para tomar decisiones. (Ponce & Torres Soto, Aurora; Quezada Aguilera, Fátima Sayu, 2014). Y aunque ha sido imposible abarcar todas y cada una de las acciones que el ser humano puede realizar, se ha avanzado notablemente en aplicaciones, dispositivos, y aparatos presentes en la vida cotidiana.

Las empresas industriales, interesadas en la innovación tecnológica; utilizan equipos robotizados y maquinas automatizadas para el desempeño de las tareas para las cuales fueron diseñadas. De esta manera, la inteligencia artificial posee infinitas posibilidades en dicho sector, mostrando el enorme potencial en la transformación digital, Haciendo posible la implementación de tecnologías de procesamiento de datos, software cada vez más avanzados y sensores diversos.

Por lo anterior, la Inteligencia artificial es indispensable, su aporte permite mejorar la productividad, optimizando tiempos y recursos, mejorando la calidad de la producción, logrando un importante incremento en el crecimiento del negocio, mediante el procesamiento de grandes cantidades de datos y el apoyo de algoritmos inteligentes, buscando el aprendizaje del software, para detectar tendencias o patrones que logran la eficiencia en los procesos.

Dentro de las tareas que pueden realizar los equipos gracias a la inteligencia artificial se tienen:

**Monitorización:** la inteligencia artificial permite la observación continua del proceso en línea, detectando cambios en el mismo. Toda la información recopilada es luego analizada

permitiendo su optimización. La recopilación de datos del entorno de trabajo es posible a través del uso de sensores inteligentes y la automatización industrial.

**Visión artificial:** dentro de la “Inteligencia Artificial” la visión artificial permite que los sistemas adquieran información a través de imágenes o videos, u otras entradas digitales, para luego tomar acciones o realizar recomendaciones de acuerdo a dicha información.

La aplicación de la visión artificial permite:

- Automatizar ciertas actividades de inspección que son repetitivas y que son realizadas por operadores.
- Realizar control de calidad en el proceso cuando dicha actividad se hace compleja por métodos tradicionales
- La inspección de objetos sin tener contacto físico con los mismos.
- Mejorar la inspección de la producción optimizando la velocidad del proceso.
- Optimizar el tiempo de un ciclo de trabajo en los procesos industriales automatizados.
- mejorar las inspecciones en los procesos donde existe diversidad de piezas las cuales sufren diversos cambios en la producción.

**Las principales aplicaciones de la visión artificial en la industria actual son:**

- Permite la reducción de riesgos mejorando el nivel de seguridad
- Ayuda e mejorar la precisión en el proceso para el diagnóstico de errores en el mismo.
- Realiza una disminución en la cantidad de desperdicio producido por dicho proceso
- Realiza supervisión en la trazabilidad de un producto durante todo el proceso.

### **Mediciones tridimensionales.**

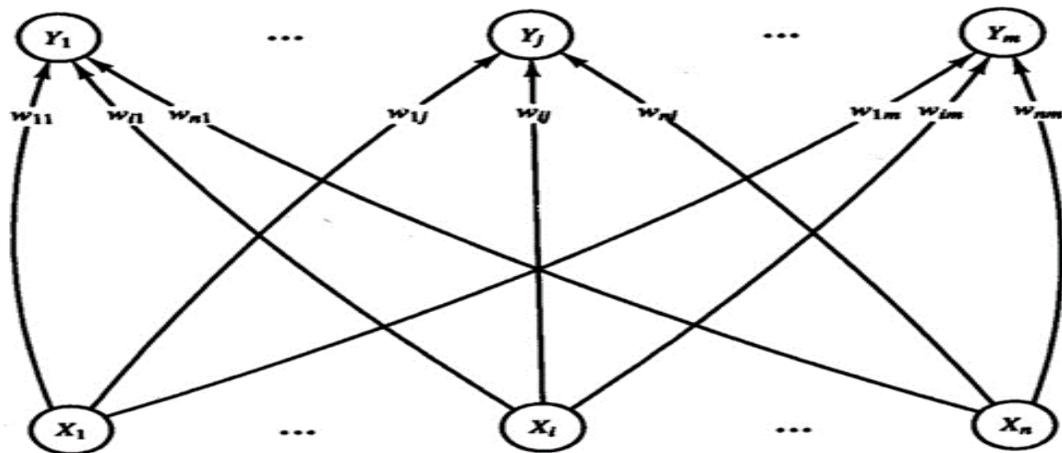
Algunos retos de la inteligencia Artificial:

Dentro de los retos que tiene la inteligencia artificial se encuentra en primera medida la adaptación a los requisitos de ley, los cuales exigen la confianza y transparencia, la solidez técnica en la obtención de los datos y las medidas de seguridad, la gestión de la privacidad y protección de los mismos; además, este nivel de innovación exige una mejora continua, en el desarrollo de hardware y software más potentes. Así mismo un gran avance en la computación y transmisión de datos, requiriendo de igual manera la formación académica especializada este actualizada de acuerdo a la era digital, ayudando a superar la resistencia a la transformación digital, mostrando el potencial que ofrece la tecnología.

#### **4.2.1 Red de mapas auto-organizados (SOM)**

Los mapas auto organizados fueron inventados por el profesor de la academia de Finlandia Teuvo Kohonen en 1982. Son un tipo de red neuronal no supervisada, con la cual uno de los objetivos es el clustering, agrupar los datos del conjunto de entrada con diferentes criterios. Otra característica importante es que son de 2 capas y manejan un lenguaje competitivo, ósea dependiendo de los datos de entrada se activan o no ciertas neuronas, entregando una representación discreta del espacio de las muestras de entrada, llamado mapa, estos, se diferencian de otras redes neuronales artificiales porque usan la función de vecindad para ayudar a preservar la topología de la red. Este proceso de reducir la dimensionalidad de vectores es una técnica de compresión de datos conocida como Cuantización Vectorial.

**Figura 8.** Mapa auto-organizado de Kohonen (SOM)



**Fuente:** FAUSETT, L. Fundamentals of Neural Networks. New York: Prentice Hall, 1994. p170.

La Figura 8 muestra la estructura de un mapa autoorganizado, se observa tanto las neuronas de entrada, como las neuronas de salida, la interconexión entre cada una de ellas y los pesos asociados a los datos de información externos, estos pesos son las características de cada uno de los datos de entrada. El algoritmo tiene la capacidad de tomar la información o cada de uno de los datos originales de entrada y colocarlo en un punto del mapa dimensional que ha creado. Cada uno de estos puntos se llama neurona o nodo. Con este algoritmo se consigue conservar la topología de los datos originales en el espacio de salida, ósea que los datos originales con características similares que estuvieran juntos, se van a ubicar en la misma neurona o en neuronas cercanas. Es importante resaltar que pueden conservar mucha información sobre el espacio original aun después de representarlos sobre el mapa. Como en la mayoría de las redes neuronales artificiales, los SOMs operan de dos formas: entrenamiento y mapeo. En el entrenamiento se construye el mapa usando los datos originales del proceso mientras que en el mapeo se clasifica una nueva entrada.

El funcionamiento de la red SOM se realiza en dos fases, la primera se conoce como fase competitiva o aprendizaje competitivo, puesto que las neuronas buscan quedaren en el

espacio de entrada, luego pasa a la fase cooperativa, donde las neuronas sufren un cambio de sus pesos para que sean semejantes a las coordenadas de los datos que se le ha asignado. Estos pesos son los parámetros modificables del modelo.

Al momento de programar una red SOM es importante definir la cantidad de neuronas de la red, el tamaño del mapa, su estructura y su forma, pues de ellos depende la distribución de los datos de entrada en los nuevos clusters. El número de neuronas puede ser tan grande como se quiera, inclusive la cantidad de neuronas podría exceder la cantidad de vectores de entrada sin afectar la calidad del mapa de la red, siempre que el tamaño de la vecindad sea seleccionado adecuadamente. Sin embargo, a medida que el tamaño del mapa aumenta, la fase de entrenamiento se vuelve impráctica desde el punto de vista computacional para la mayoría de las aplicaciones.

En términos generales, se recomienda el uso de la estructura hexagonal para la red, porque entonces 6 vecinos de una neurona se encontrarán a la misma distancia; de esta manera el mapa tiene una visibilidad más agradable y organizada.

### **4.3 Matlab**

Matlab es un lenguaje de programación de alto nivel que facilita expresar conceptos matemáticos a través de matrices, los cuales se han vuelto fundamentales en toda aplicación hoy en día. Dentro de su estructura se encuentra una amplia librería de herramientas y funciones para diferentes áreas como sistemas de control, comunicaciones inalámbricas, redes neuronales o inteligencia artificial, finanzas computacionales, etc. Está diseñado para resolver problemas de forma intuitiva, incluyendo interfaces de usuario y aplicaciones, como también la posibilidad de crear las propias aplicaciones interactivas. Está disponible para las plataformas Unix, Windows,

macOS y GNU/Linux (Wikipedia, 2020). Para (Moreno, 2009), Matlab es usado hoy en día por instituciones académicas como también por el sector industrial permitiendo el análisis de datos, y ayudando en el desarrollo de algoritmos y creación de modelos, gracias a esto, Permite resolver problemas complejos, que generalmente aplicando técnicas tradicionales no se puede llegar a dar solución. Debido a esto Matlab se ha convertido en una herramienta poderosa a nivel mundial.

Existen experiencias de proyectos como el de Alfonso Rodríguez (2009) que ha desarrollado una interfaz gráfica de redes neuronales usando Matlab, dirigido a usuarios que no han desarrollado experiencias con redes neuronales, con el objetivo que estos puedan realizar entrenamiento y prueba de redes neuronales backpropagation y de base radial de una manera fácil y sencilla; utilizando herramientas de Matlab, tanto las de la toolbox de redes neuronales, como las herramientas para la programación de ventanas gráficas que dispone (GUI). También el objetivo del proyecto es brindar una herramienta docente para la asignatura “Técnicas Avanzadas de Diseño en Ingeniería Mecánica”, (Moreno, 2009) puesto que las redes neuronales son técnicas no paramétricas muy utilizadas en el campo de la ciencia e ingeniería porque posibilitan resolver problemas complejos; además, ayudan para obtener un modelo no explícito que relaciona un conjunto de variables salida con un conjunto de variables entrada, facilitando el valor de salida, dados unos valores de entrada del modelo, disponiendo de un conjunto de observaciones de las variables.

## 5. Metodología del Proyecto

Esta investigación se aborda desde un enfoque cuantitativo, que se puede concebir según Sampieri (2010), como una recolección de datos para probar hipótesis con base en la medición numérica y el análisis estadístico para establecer patrones de comportamiento, identificar tendencias y promedios, realizar predicciones, y comprobar relaciones entre variables.

La investigación cuantitativa en las empresas aporta a la mejora de productos y servicios, apoya en la toma de decisiones en la gestión de los activos ayudando a conseguir los objetivos establecidos.

Para el desarrollo de este trabajo de investigación se recopilaron los datos de las mediciones numéricas de las variables de presión, flujo y temperatura dentro de la operación del horno H150 ubicado en la planta Topping de la refinería de Barrancabermeja; posteriormente se analizan los datos a través de un método de correlación de variables con el fin de establecer un patrón de comportamiento entre variables. A partir de esta definición, el trabajo desarrollado fue bajo un enfoque analítico Técnicas o parámetros usados en el desarrollo del proyecto.

### 5.1 Diseño de la investigación

Esta investigación se enfoca en el diseño cuantitativo de tipo no experimental, el cual, lo define Hernández y Baptista como “Estudios que se realizan sin la manipulación deliberada de variables y en los que sólo se observan los fenómenos en su ambiente natural para después analizarlos”. (Hernández S, Baptista L, & Fernández Collado, 2014) Estudio desarrollado en tres etapas, la recolección de los datos obtenidos del sistema donde se presenta la falla en un determinado lapso de tiempo, la medición de dichos datos en el mismo periodo de tiempo y el

análisis del comportamiento de las diferentes variables del sistema, buscando de esta manera realizar un seguimiento a dicho comportamiento prediciendo e identificando el modo de falla en el equipo de estudio.

Esta información es descargada mediante la herramienta PI, la cual esta enlazada al DCS (sistema de control distribuido), donde se muestran y almacenan los datos en línea de cada uno de los sensores de equipo. Posteriormente con el apoyo de herramientas como el PCA (análisis de componentes principales) se observa la relación de las diferentes variables del sistema de acuerdo con su comportamiento en el tiempo, para determinar con base en esto, las variables que arrojen la misma información, de esta manera se simplifica la cantidad de variables.

Con las variables definidas como importantes para el estudio se realiza el diseño de la red neuronal de acuerdo con un análisis previo de las clases de redes neuronales que se pueden utilizar, para entrenarla con la información obtenida de las variables importantes del sistema

En la anterior teoría, su aplicación en este trabajo de investigación se desarrolla en un ambiente natural como es el funcionamiento del Horno 150 dentro de un proceso de refinación de crudo dentro de la planta Topping de la refinería de Barrancabermeja. De este modo, se toman los datos de la medición de las variables de flujo, temperatura y presión que registran los sistemas de información a los cuales están anclados estos elementos. Los datos obtenidos se llevan a una etapa de análisis con el fin de encontrar una relación entre ellos.

Este “diseño correlacional predictivo, determina la varianza de una o más variables con base a la varianza de otra(s) variable(s), así, como con los modelos experimentales, las variables del estudio son clasificadas como independientes (predictivas) y dependientes (resultado). No

obstante, estas variables no son manipuladas, pues se dan de forma natural”. (Valmi, Diessnack, & Mendes, 2007)

## **5.2 Muestra**

El objeto de este estudio ha sido, el horno H-150 de la unidad 150 de la refinería de Barrancabermeja, aunque se cuenta con dos equipos similares en dicha unidad, se planteó la investigación en este equipo por la importancia en el proceso para la estructura del departamento pero ha tenido eventos repetitivos de coquizamiento con la necesidad de realizar intervenciones mecánicas que han generado horas de parada no programadas y pérdidas en la producción.

## **5.3 Técnicas de recolección de Datos**

Para la operación de dicho equipo se cuenta con medidores de presión manométricos de una sola cámara, utilizados para medir la presión en determinado punto. De presión diferencial utilizada para medir flujo o se tiene también los medidores de diferencial en filtros. Según el fabricante existen de diferentes tipos.

Medidores de presión resistivos, capacitivos, de presión magnética, de Presión Piezoeléctricos y Medidores de presión de esfuerzo. Para la medición de temperatura, el equipo cuenta con termocuplas de tipo inserción las cuales se adaptan al proceso con un termo pozo y las termocuplas de piel de tubo, las cuales facilitan la medición de la temperatura teniendo contacto directo con la superficie del tubo por donde circula el producto del proceso.

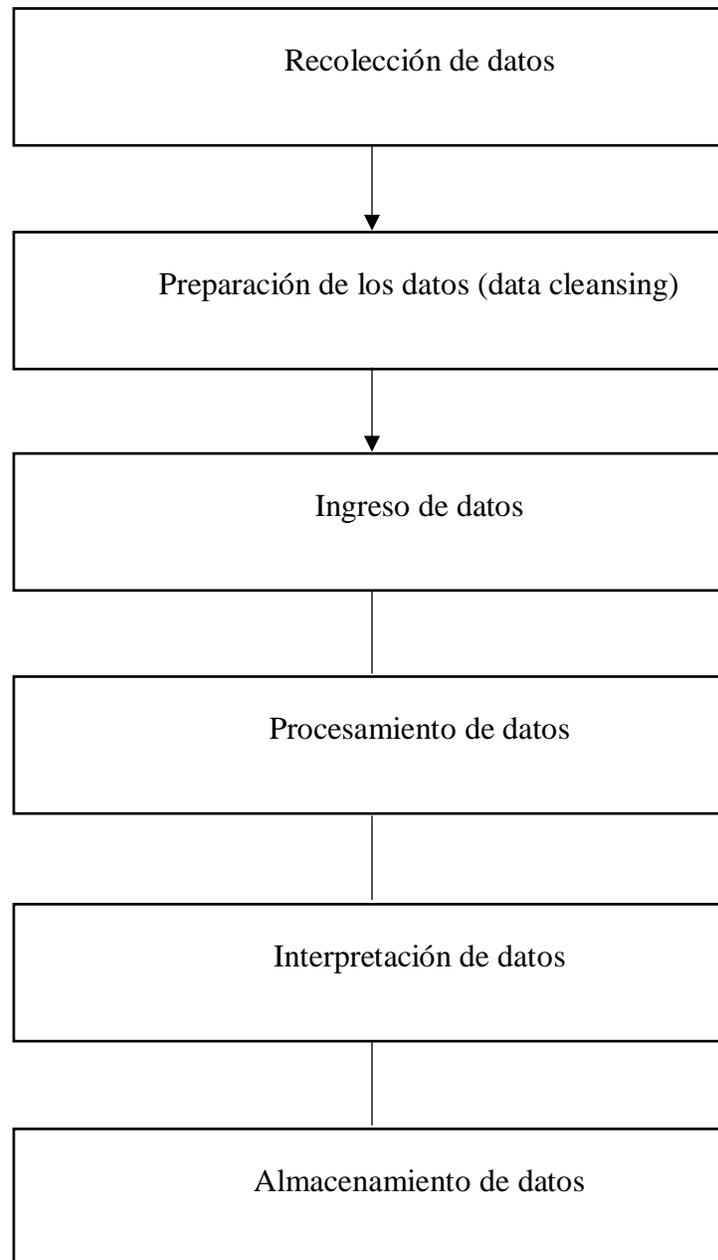
Estos instrumentos se encuentran instalados en puntos estratégicos a lo largo de los serpentines, en la cámara de combustión, en el hogar del horno, en la chimenea, etc. manejando información independiente para los dos serpentines que hacen parte de su diseño. Para esta investigación se tomaron valores en cada uno de los medidores desde la última intervención de

mantenimiento realizada al equipo hasta la fecha donde nuevamente se evidencia la falla por los cambios en las variables de operación por fuera de las ventanas operativas, cabe resaltar que se tienen en cuenta los datos de operación normal, paradas programadas, no programadas y fallas, la información asociada a los instrumentos de medición de cada una de las variables debe ser confiable, todo esto, buscando caracterizar la mayor información posible.

Los datos se tomaron en intervalos de una hora entre valor y valor, pues al analizar la información se observó que en intervalos de tiempo más cortos las mediciones no mostraban alteraciones importantes en el proceso. Se recogieron aproximadamente 26000 datos por cada variable para un total de 234000 datos, tomados desde el 1 de julio de 2019 hasta el 31 de diciembre de 2021. La información obtenida por dichos sensores es llevada a la pantalla del DCS, esta información histórica de las diferentes variables a lo largo del tiempo se puede descargar a un archivo de Excel mediante la herramienta PI DATA LINK.

## **6. Descripción de la implementación del algoritmo de inteligencia artificial**

El modelo de detección e identificación de fallas basadas en un algoritmo de inteligencia artificial fue diseñado para visualizar la condición de operación acercándose al modo de falla en el horno H-150 de la refinería de Barrancabermeja, de acuerdo al comportamiento de las variables del equipo. En la tabla 3 se muestra las distintas etapas del procesamiento de los datos recolectados para el desarrollo de esta investigación, dentro de esa etapa tenemos, la recolección, la preparación, el ingreso, el procesamiento, la interpretación y el posterior almacenamiento de dichos datos.

**Tabla 3. Etapas del procesamiento de datos**

A continuación, se enuncian las etapas en las cuales se desarrolló este modelo.

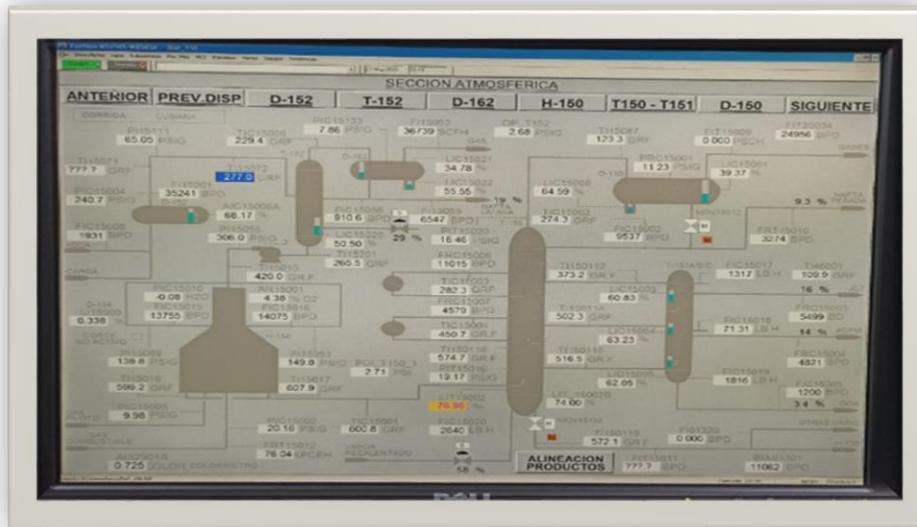
### **6.1 Fase 1: Recolección de datos**

El horno H-150 de la refinería de Barrancabermeja es un equipo diseñado internamente por dos serpentines (A y B) por donde circula el crudo. Para el desarrollo de esta investigación se trabaja en el serpentín B, pues dicho serpentín ha presentado mayor índice de coquizamiento en las últimas intervenciones. Para su operación cuenta con el respaldo de un sistema robusto de control e instrumentación de las señales de flujo, temperatura y presión, que ayuda a monitorear permanentemente todas las variables de operación identificando las posibles desviaciones.

Las señales están conectadas al DCS (sistema de control de proceso) y dicha información es descargada en Excel por medio de la herramienta PI DATA LINK. Para esta investigación se descargó la información desde el primero de julio de 2019 hasta el 31 de marzo de 2023, pues en ese periodo de tiempo ha sido necesario varias intervenciones a dicho equipo, aproximadamente una por año.

Para la recolección de la data fue necesario, inicialmente, la revisión y el aseguramiento del correcto funcionamiento de los sensores en cada una de las variables a medir. Se organizó con el líder de mantenimiento la programación de tareas para dicha revisión.

**Figura 20.** Pantalla de control del DCS



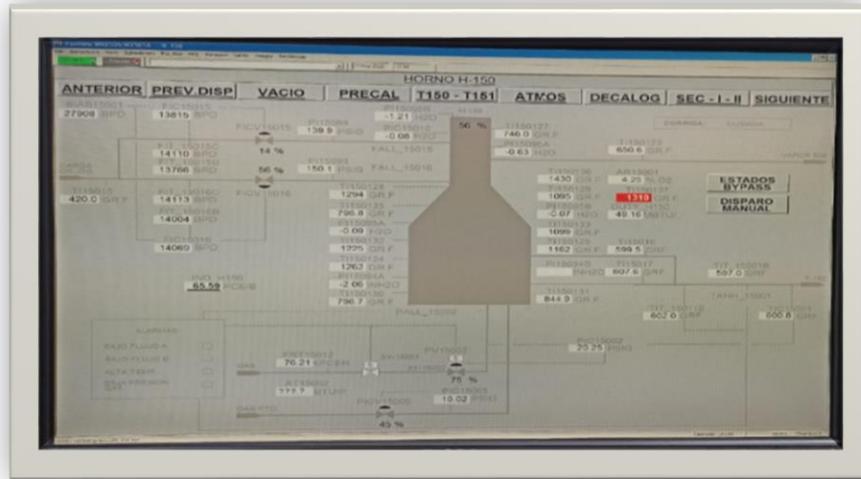
**Nota.** La Figura 20 muestra la interfaz del usuario del sistema en donde se observa los valores de la medición en línea de cada uno de los sensores de proceso. Tomado del DCS (sistema de control de proceso)

**Tabla 4.** Variables de proceso del Horno H-150

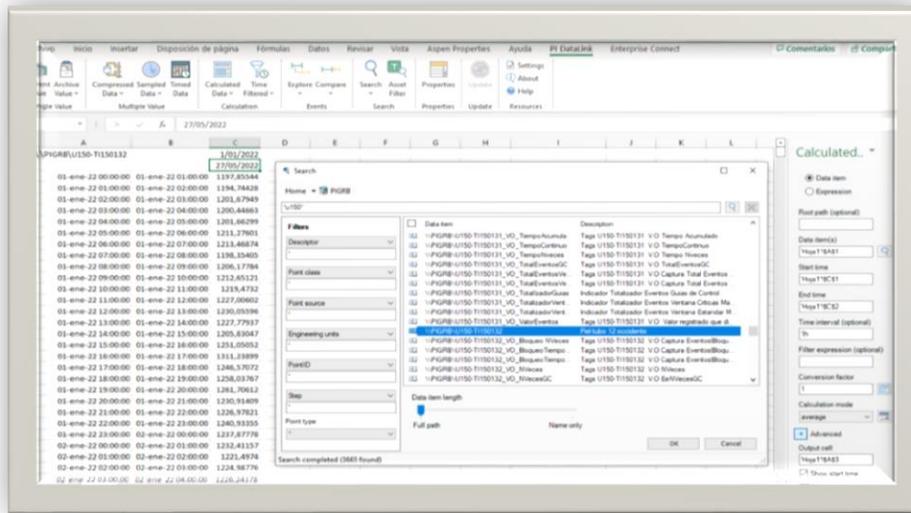
SERPENTIN A:	SERPENTIN B:
FLUJO DE CARGA SERP A	FLUJO DE CARGA SERP B
FLUJO DE CARGA SERP A % VALVULA	FLUJO DE CARGA SERP B % VALVULA
FLUJO DE GAS A QUEMADORES	FLUJO DE GAS A QUEMADORES
PRES ENTRADA SERP A	PRES ENTRADA SERP B
TEMP SALIDA SERP A	TEMP SALIDA SERP B
TEMP PIEL TUBO 1 SERP A	TEMP PIEL TUBO 1 SERP B
TEMP PIEL TUBO 12 SERP A	TEMP PIEL TUBO 12 SERP B
TEMP CHOQUE SERP A	TECHO INCLINADO RADIACION B
TECHO INCLINADO RADIACION A	TEMP CHOQUE SERP B
CAM COMBUST SERP A	CAM COMBUST SERP B
TEMP GASES DE CHIMENEA	TEMP GASES DE CHIMENEA
PRESION HOGAR H-150	PRESION HOGAR H-150

*Nota.* Las variables por cada uno de los serpentines se visualizan en la Tabla 4 y el proyecto se desarrolla con las variables del serpentín B.

**Figura 21.** Control DCS del H-150



*Nota.* Una vez descargada la información, se organizó de acuerdo a los tags de las variables del equipo. Tomado del DCS (sistema de control de proceso). La Figura 21 muestra la forma como se ve el proceso en línea y las variaciones del mismo. El comportamiento de las variables dependientes y los ajustes realizados a las variables independientes.

**Figura 22.** Información del PI DATA LINK

Nota. Se observa en la Figura 22 el archivo de Excel al descargar la información de PI DATA LINK, permitiendo ubicar toda la instrumentación de la planta por el número de tag que tiene cada instrumento, y el historial de datos que se requiera y en que intervalos de tiempo entre dato y dato. Tomado del complemento de MS Excel PI DATA LINK

El objetivo de esta fase era el de recopilar la información de todas las variables posibles que afectan el funcionamiento del horno en los últimos 4 años. Ubicando durante este periodo de tiempo los valores de las variables que mostraban la falla en el equipo. Se tomaron valores en cada uno de los medidores con intervalos de una hora entre valor y valor, pues después de analizar la información se observó que en intervalos de tiempo más cortos las mediciones no mostraban alteraciones importantes en el proceso.

Se recogieron aproximadamente 32000 datos por cada variable para un total de 384000 datos, tomados desde el 1 de julio de 2019 hasta el 31 de marzo de 2023. Teniendo en cuenta las 12 de variables de proceso por cada serpentín mostradas en la Tabla 3.

### 6.2 Fase 2: Preparación de los datos.

Para realizar el procedimiento de entrenamiento del sistema propuesto, se requirió del análisis y depuración de los datos arrojados en el lapso de tiempo relacionado, se descargaron los datos arrojados de proceso en dos etapas, antes y después de la intervención mecánica realizada al equipo. Durante esta fase, luego de tener el archivo de Excel donde se descarga la información del DCS, se organizó la información arrojada por cada una de las variables en el tiempo seleccionado de acuerdo a las fechas donde el equipo empezó a mostrar la falla hasta tenerlo nuevamente en operación. Mediante el proceso de Data Cleansing, se realizó el análisis y se identificaron datos equivocados y mal procesados, datos atípicos con errores de señal que alteraban el cálculo, datos no numéricos que no permitían realizar el proceso.

**Tabla 5. Datos atípicos**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
32326	0	-7,90337822	1,88669907	94,0951742	79,825796	85,5704891	84,2639227	86,0246694	90,754313	-0,00263096	85,325894					
32327	0	[-11059]	No	1,91543812	92,014521	79,9949496	85,5242857	84,3046614	85,9821779	90,2383218	-0,00562403	84,2740172				
32328	0	[-11059]	No	2,48597687	89,9338678	80,5774491	86,2780822	85,3434001	85,9396864	90,3836835	-0,00861709	83,3763242				
32329	0	[-11059]	No	2,5838502	88,2639788	81,7704229	86,5630762	85,7793857	86,0389105	90,7828178	-0,01101926	83,8448611				
32330	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No Good Data For Calculation
32331	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No Good Data For Calculation
32332	0	-7,9765625	2,5533968	87,745491	82,1385956	86,7020035	85,8022614	86,194725	90,7866974	-0,01279541	84,1113358					
32333	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No Good Data For Calculation
32334	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No	[-11059]	No Good Data For Calculation
32335	0	[-11059]	No	1,38140688	84,334873	80,9959242	83,4770609	81,8802766	84,1464704	88,2419558	-0,00131413	82,8855471				
32336	0	[-11059]	No	1,55738168	83,9909676	80,8833143	83,3016681	81,6347257	83,3533385	87,8177725	-0,00156266	82,6166113				
32337	0	[-11059]	No	1,8624369	83,5760684	80,918752	83,0900683	81,3384848	83,7207789	87,6248559	-0,0018625	82,2921579				
32338	0	[-11059]	No	1,80770801	83,1611891	80,7299807	82,8784684	81,0422438	83,4880392	87,3759949	-0,00216234	81,9677045				
32339	0	[-11059]	No	1,78329233	82,7462699	80,4848221	82,6668686	80,7460029	83,2552596	87,3095395	-0,00246218	81,6432512				
32340	0	[-11059]	No	1,67679342	82,3313706	80,0317795	82,4552687	80,4497619	83,0224999	86,7559008	-0,00276201	81,3187978				
32341	0	[-11059]	No	1,49827205	81,9164714	79,5502232	82,2436689	80,153521	82,7897402	86,2193712	-0,00305185	80,9943444				
32342	0	[-11059]	No	0,88362795	81,5015721	79,1402021	82,032069	79,8572801	82,5509806	85,5741214	-0,00380109	80,6698911				
32343	0	[-11059]	No	0,24443429	81,3406486	78,810695	82,0090912	78,8432699	82,4872587	85,39403	-0,00368273	80,5111994				
32344	0	[-11059]	No	0,33531601	81,4801461	76,5237131	82,5991354	80,7465099	82,9474097	84,9791694	-0,00407268	80,8912333				
32345	0	[-11059]	No	0,84610231	81,7581375	78,1465222	83,236325	81,7203076	83,4483202	84,5862269	-0,00446792	81,3127076				

Nota. La Tabla 5 enseña la cantidad de datos eliminada, alcanzo casi el 1% de la totalidad de los datos, siendo una cantidad despreciable a la hora del análisis. Tomado de la Información de sensores del equipo DSC.

### 6.3 Fase 3: Ingreso de datos

Al tener la información organizada y con los datos correctos en todas las variables, teniendo 3 variables independientes y 9 variables dependientes, Se realizó un análisis de correlación entre las variables independientes y cada una de las variables dependientes, para determinar qué variables podían proporcionar información semejante que pudiera generar exceso de la misma. Donde se observa que ninguna de las variables dependientes alcanza índices de correlación relevante o que entreguen la misma información.

Se dice que la función de coeficiente de correlación lineal es una fórmula de Excel que genera como resultado el coeficiente de correlación lineal entre dos o más variables. El coeficiente de correlación  $r$  es un valor sin unidades entre -1 y 1, entre más se aproxima  $r$  a cero, menor es la relación lineal. Es un concepto estadístico fácil de entender para medir la relación que existe entre dos conjuntos de datos, esto facilita el entendimiento de la información que se quiera analizar en cualquier aspecto.

En ocasiones basta con un simple gráfico para conocer la correlación que existe entre dos o más variables, pero con el coeficiente de correlación además de afirmarlo da a conocer el grado de relación entre las mismas.

**Tabla 6.** Cálculo de correlación con la presión de entrada serpentín B.

Resumen									
Estadísticas de la regresión									
Coefficiente de correlación múltiple	0,866387896	PRES ENTRADA SERP B							
Coefficiente de determinación R <sup>2</sup>	0,750627987								
R <sup>2</sup> ajustado	0,750604736								
Error típico	12,26861328								
Observaciones	32180								
ANÁLISIS DE VARIANZA									
	Grados de libertad	Suma de cuadrados	dfo de los cua	F	Valor crítico de F				
Regresión	3	14578070,61	4859356,87	32284,0373	0				
Residuos	32176	4843095,221	150,518872						
Total	32179	19421165,83							
	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%	
Intercepción	-1,057754423	0,424774289	-2,4901564	0,01277369	-1,8903281	-0,2251808	-1,8903281	-0,2251808	
Variable X1	0,009281423	3,13155E-05	296,383883	0	0,00922004	0,0093428	0,00922004	0,0093428	
Variable X2	-0,000241036	4,37607E-05	-5,5080512	3,656E-08	-0,0003268	-0,0001553	-0,0003268	-0,0001553	
Variable X3	-9,428520315	2,36669822	-3,9838287	6,7963E-05	-14,067338	-4,7897025	-14,067338	-4,7897025	

Nota. La Tabla 6 muestra un coeficiente de correlación de las variables independientes con la variable dependiente de presión de entrada del serpentín B. Se observa un valor de 0,8663 indicando que es una correlación aceptable entre dichas variables. Tomado de MS Excel.

Además muestra los análisis de varianza, en los cuales está la regresión y los residuales. Los residuales indican la diferencia entre el valor observado de la variable dependiente y el valor esperado de la misma.

Los grados de libertad son valores de un conjunto de datos con la posibilidad de variar después de haber impuesto alguna restricción a los mismos.

El coeficiente de determinación es un valor importante dentro del análisis estadístico, su resultado oscila entre 0 y 1. Cuanto más cerca de 1 se encuentre su valor, mayor será el ajuste del modelo a la variable en estudio. Y de manera inversa, cuanto más cerca de cero, menos ajustado estará el modelo y, por tanto, menos fiable será.

**Tabla 7.** Cálculo de correlación con la variable de flujo de carga por el serpentín B

Resumen								
Estadísticas de la regresión								
Coefficiente de correlación múltiple	0,746976596	FLUJO DE CARGA SERP B % VALVULA						
Coefficiente de determinación R <sup>2</sup>	0,557974035							
R <sup>2</sup> ajustado	0,557932822							
Error típico	8,924353247							
Observaciones	32180							
ANÁLISIS DE VARIANZA								
	Grados de libertad	Suma de cuadrados	Medio de los cua	F	Valor crítico de F			
Regresión	3	3234832,266	1078277,42	13538,7013	0			
Residuos	32176	2562627,946	79,6440809					
Total	32179	5797460,212						
	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%
Intercepción	-6,651254489	0,308986494	-21,526036	4,669E-102	-7,2568797	-6,0456293	-7,2568797	-6,0456293
Variable X 1	0,004256838	2,27793E-05	186,872708	0	0,00421219	0,00430149	0,00421219	0,00430149
Variable X 2	0,000350045	3,18321E-05	10,9966071	4,4523E-28	0,00028765	0,00041244	0,00028765	0,00041244
Variable X 3	-25,73084712	1,721567911	-14,94617	2,4351E-50	-29,105185	-22,356509	-29,105185	-22,356509

Nota. La Tabla 7 enseña un coeficiente de correlación de las variables independientes con la variable dependiente de flujo de carga por el serpentín B, mostrando el porcentaje de apertura de la válvula de control y un valor de 0,7469 indicando que es una correlación regular entre dichas variables. Tomado de MS Excel.

**Tabla 8.** Coeficiente de correlación con la variable de temperatura de salida serpentín B

Resumen								
Estadísticas de la regresión								
Coefficiente de correlación múltiple	0,725004907	TEMP SALIDA SERP B						
Coefficiente de determinación R <sup>2</sup>	0,525632116							
R <sup>2</sup> ajustado	0,525587887							
Error típico	53,23189591							
Observaciones	32180							
ANÁLISIS DE VARIANZA								
	Grados de libertad	Suma de cuadrados	Medio de los cua	F	Valor crítico de F			
Regresión	3	101028181,5	33676060,5	11884,4042	0			
Residuos	32176	91175031,48	2833,63474					
Total	32179	192203213						
	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%
Intercepción	252,0013187	1,843039652	136,73136	0	248,388891	255,613746	248,388891	255,613746
Variable X 1	0,022762692	0,000135874	167,527911	0	0,02249637	0,02302901	0,02249637	0,02302901
Variable X 2	-0,001418488	0,000189872	-7,4707699	8,1748E-14	-0,0017906	-0,0010463	-0,0017906	-0,0010463
Variable X 3	-354,773338	10,26879162	-34,548694	7,469E-257	-374,90056	-334,64612	-374,90056	-334,64612

Nota. Observamos en la Tabla 8, un coeficiente de correlación de las variables independientes con la variable dependiente de temperatura de salida del serpentín B y un valor de 0,7250 indicando que es una correlación regular entre dichas variables. Tomado del MS Excel

**Tabla 9.** Coeficiente de correlación con la variable de temperatura del tubo 1 serpentín B

Resumen								
<i>Estadísticas de la regresión</i>								
Coefficiente de correlación múltiple	0,809980057	TEMP PIEL TUBO 1 SERP B						
Coefficiente de determinación R <sup>2</sup>	0,656067693							
R <sup>2</sup> ajustado	0,656035626							
Error típico	68,55394022							
Observaciones	32180							
ANÁLISIS DE VARIANZA								
	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Medio de los cua</i>	<i>F</i>	<i>Valor crítico de F</i>			
Regresión	3	288451350,2	96150450,1	20459,0978	0			
Residuos	32176	151215704,2	4699,64272					
Total	32179	439667054,4						
	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>
Intercepción	263,5739898	2,373532409	111,047142	0	258,921777	268,226203	258,921777	268,226203
Variable X 1	0,039296043	0,000174983	224,570073	0	0,03895307	0,03963902	0,03895307	0,03963902
Variable X 2	0,000538099	0,000244524	2,20060062	0,02777139	5,8823E-05	0,00101737	5,8823E-05	0,00101737
Variable X 3	-446,7853693	13,22451728	-33,784626	6,655E-246	-472,70592	-420,86482	-472,70592	-420,86482

Nota. Se observa en la Tabla 9, un coeficiente de correlación de las variables independientes con la variable dependiente de temperatura de piel del tubo 1 del serpentín B y un valor de 0,8099 indicando que es una correlación aceptable entre dichas variables. Tomado del MS Excel

**Tabla 10.** Coeficiente de correlación con la variable de temperatura del tubo 12 serpentín B

Resumen								
<i>Estadísticas de la regresión</i>								
Coefficiente de correlación múltiple	0,798915115	TEMP PIEL TUBO 12 SERP B						
Coefficiente de determinación R <sup>2</sup>	0,638265362							
R <sup>2</sup> ajustado	0,638231635							
Error típico	98,03239454							
Observaciones	32180							
ANÁLISIS DE VARIANZA								
	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Medio de los cua</i>	<i>F</i>	<i>Valor crítico de F</i>			
Regresión	3	545610166,4	181870055	18924,3938	0			
Residuos	32176	309222633,8	9610,35038					
Total	32179	854832800,2						
	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>
Intercepción	300,8575011	3,394160347	88,6397431	0	294,204819	307,510183	294,204819	307,510183
Variable X 1	0,054746092	0,000250227	218,785732	0	0,05425564	0,05523655	0,05425564	0,05523655
Variable X 2	-0,001205341	0,00034967	-3,4470835	0,0005674	-0,0018907	-0,00052	-0,0018907	-0,00052
Variable X 3	-496,3189383	18,91110986	-26,244834	3,134E-150	-533,38543	-459,25245	-533,38543	-459,25245

Nota. La Tabla 10 enseña, un coeficiente de correlación de las variables independientes con la variable dependiente de temperatura de piel del tubo 12 del serpentín B y un valor de 0,7989 indicando que es una correlación regular entre dichas variables. Tomado del MS EXcel

**Tabla 11.** Coeficiente de correlación con la variable de temperatura de choque serpentín B

Resumen								
Estadísticas de la regresión								
Coefficiente de correlación múltiple	0,505323276	TEMP CHOQUE SERP B						
Coefficiente de determinación R <sup>2</sup>	0,255351613							
R <sup>2</sup> ajustado	0,255282184							
Error típico	173,0184521							
Observaciones	32180							
ANÁLISIS DE VARIANZA								
	Grados de libertad	Suma de cuadrados	Medio de los cuadrados	F	Valor crítico de F			
Regresión	3	330296712	110098904	3677,88505	0			
Residuos	32176	963200939,7	29935,3848					
Total	32179	1293497652						
	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%
Intercepción	345,3728006	5,990390952	57,6544675	0	333,631408	357,114193	333,631408	357,114193
Variable X 1	0,035857546	0,000441628	81,1939427	0	0,03499194	0,03672315	0,03499194	0,03672315
Variable X 2	-0,001625534	0,000617136	-2,6339965	0,00844267	-0,0028351	-0,0004159	-0,0028351	-0,0004159
Variable X 3	-1338,254708	33,37642593	-40,095806	0	-1403,6738	-1272,8357	-1403,6738	-1272,8357

Nota. Se observa en la Tabla 11 un coeficiente de correlación de las variables independientes con la variable dependiente de temperatura de choque del serpentín B y un valor de 0,5053 indicando que es una correlación mínima entre dichas variables. Tomado del MS Excel.

**Tabla 12.** Coeficiente de correlación con la variable de temperatura de la cámara de combustión serpentín

Resumen								
Estadísticas de la regresión								
Coefficiente de correlación múltiple	0,805809936	CAM COMBUST SERP B						
Coefficiente de determinación R <sup>2</sup>	0,649329652							
R <sup>2</sup> ajustado	0,649296957							
Error típico	101,2560377							
Observaciones	32180							
ANÁLISIS DE VARIANZA								
	Grados de libertad	Suma de cuadrados	Medio de los cuadrados	F	Valor crítico de F			
Regresión	3	610857770,7	203619257	19859,8969	0			
Residuos	32176	329893615,6	10252,7852					
Total	32179	940751386,2						
	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%
Intercepción	345,3774164	3,505772042	98,5167924	0	338,505971	352,248862	338,505971	352,248862
Variable X 1	0,060001563	0,000258455	232,154507	0	0,05949498	0,06050814	0,05949498	0,06050814
Variable X 2	-0,007335431	0,000361168	-20,310295	3,893E-91	-0,0080433	-0,0066275	-0,0080433	-0,0066275
Variable X 3	-80,72099705	19,53297236	-4,1325506	3,5967E-05	-119,00636	-42,435635	-119,00636	-42,435635

Nota. Se observa en la Tabla 12 un coeficiente de correlación de las variables independientes con la variable dependiente de temperatura de la cámara de combustión del serpentín B y un valor de 0,8058 indicando que es una correlación aceptable entre dichas variables. Tomado del MS Excel.

**Tabla 13.** Coeficiente de correlación con la variable de temperatura de gases de chimenea

Resumen								
Estadísticas de la regresión								
Coeficiente de correlación múltiple	0,669743691	TEMP GASES DE CHIMENEA						
Coeficiente de determinación R <sup>2</sup>	0,448556612							
R <sup>2</sup> ajustado	0,448505197							
Error típico	80,00986026							
Observaciones	32180							
ANÁLISIS DE VARIANZA								
	Grados de libertad	Suma de cuadrados	Medio de los cua	F	Valor crítico de F			
Regresión	3	167546517,9	55848839,3	8724,23043	0			
Residuos	32176	205977165,3	6401,57774					
Total	32179	373523683,2						
	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%
Intercepción	279,5311034	2,770168946	100,907601	0	274,101468	284,960739	274,101468	284,960739
Variable X 1	0,029430543	0,000204225	144,108715	0	0,02903025	0,02983083	0,02903025	0,02983083
Variable X 2	0,000916715	0,000285386	3,21219925	0,00131853	0,00035735	0,00147608	0,00035735	0,00147608
Variable X 3	-426,0259069	15,43444149	-27,602289	8,96E-166	-456,27799	-395,77382	-456,27799	-395,77382

Nota. La Tabla 13 muestra un coeficiente de correlación de las variables independientes con la variable dependiente de temperatura de gases de chimenea, se registra un valor de 0,6697 indicando que es una correlación regular entre dichas variables, Tabla 12

**Tabla 14.** Coeficiente de correlación con la variable de temperatura del techo inclinado

Estadísticas de la regresión								
Coeficiente de correlación múltiple	0,815711017	TECHO INCLINADO RADIACION B						
Coeficiente de determinación R <sup>2</sup>	0,665384463							
R <sup>2</sup> ajustado	0,665353265							
Error típico	116,904358							
Observaciones	32180							
ANÁLISIS DE VARIANZA								
	Grados de libertad	Suma de cuadrados	Medio de los cua	F	Valor crítico de F			
Regresión	3	874419853	291473284	21327,3724	0			
Residuos	32176	439737451,8	13666,6289					
Total	32179	1314157305						
	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%
Intercepción	361,7770217	4,047561401	89,3814783	0	353,843649	369,710395	353,843649	369,710395
Variable X 1	0,068218575	0,000298398	228,616427	0	0,06763371	0,06880345	0,06763371	0,06880345
Variable X 2	-0,000123265	0,000416984	-0,2956109	0,76752915	-0,0009406	0,00069404	-0,0009406	0,00069404
Variable X 3	-820,1642848	22,55163885	-36,368279	7,685E-284	-864,36635	-775,96222	-864,36635	-775,96222

Nota. Analizamos en la Tabla 14, el coeficiente de correlación de las variables independientes con la variable dependiente de temperatura del techo inclinado en la zona de radiación, con registro de un valor de 0,8157 indicando que es una correlación aceptable entre dichas variables. Tomado del MS Excel

El objetivo de esta fase era disminuir, con base en el índice de correlación de variables, la información excesiva, pero se observa que ninguna de las variables dependientes presenta alta

correlación con las variables independientes, y ninguna de las variables dependientes presenta un alto porcentaje de información similar entre ellas.

#### 6.4 Fase 4: Procesamiento de datos.

Para el desarrollo de esta fase, se toman en cuenta todas las variables dependientes e independientes del proceso. Después de tener 12 variables del serpentín B, al realizar la correlación de variables, los resultados muestran que no hay correlación entre ellas. Para el análisis de componentes principales se trabaja con todas las variables que intervienen en el proceso que son las presentadas en la Tabla 15.

**Tabla 15.** Variables de proceso para PCA

<b>FLUJO DE GAS A QUEMADORES</b>
FLUJO DE CARGA SERP B
FLUJO DE CARGA SERP B % VALVULA
FLUJO DE GAS A QUEMADORES
PRES ENTRADA SERP B
TEMP SALIDA SERP B
TEMP PIEL TUBO 1 SERP B
TEMP PIEL TUBO 12 SERP B
TECHO INCLINADO RADIACION B
TEMP CHOQUE SERP B
CAM COMBUST SERP B
TEMP GASES DE CHIMENEA
PRESION HOGAR H-150

**Nota.** Variables del serpentín B (12) del flujo de gas a quemadores. Tomado del Análisis de los Componentes Principales (ACP).

**Tabla 16.** Resultados del PCA datos en control

Factor Matrix (rotated Varimax)						
	1	2	3		Commun	Specific
22,83	0,31093838	-0,38855646	-0,67494633		0,70321134	0,29678866
30,7877044	0,85679105	-0,33667855	-0,07364124		0,85286639	0,14713361
-0,03332482	0,22985312	-0,08035256	0,87765298		0,82956374	0,17043626
39,5987101	0,21430335	-0,90740485	-0,15670261		0,8938652	0,1061348
571,803784	0,91637178	-0,20809048	0,05562622		0,88613317	0,11386683
704,904343	0,93943125	-0,2771187	0,00349675		0,95933809	0,04066191
613,801167	0,86295126	-0,45702556	0,07017599		0,95848191	0,04151809
843,615913	0,91599671	-0,34559717	-0,08516231		0,96574	0,03426
323,182947	0,42461926	-0,87774214	-0,0033112		0,95074375	0,04925625
530,133488	-0,00778076	-0,9234854	-0,21876091		0,90074216	0,09925784
1000,24383	0,85751698	-0,48928537	-0,04631125		0,97688028	0,02311972

**Nota.** Resultado del PCA, Tomado de los datos de control PCA

**Tabla 17.** Resultados del PCA datos por fuera de las ventanas operativas

Factor Matrix (rotated Varimax)						
	1	2	3	4	Commun	Specific
15009,3356	0,69899956	-0,40547511	-0,01156851	-0,37200787	0,79153414	0,20846586
63,371156	0,53152435	0,52240937	-0,17988421	-0,2780748	0,66511361	0,33488639
77,4505347	0,33699728	-0,18394112	0,19324054	-0,77440429	0,78444541	0,21555459
-0,06931562	0,06565066	-0,01980454	0,10342311	0,70636029	0,51434343	0,48565657
153,375406	0,88264797	-0,14706403	-0,019158	-0,24949449	0,8633098	0,1366902
600,626248	-0,87993069	-0,06033687	0,24454335	-0,21055763	0,88205452	0,11794548
858,785103	-0,39263963	0,07708586	0,78419849	-0,11629253	0,78859933	0,21140067
1138,54969	-0,08988881	0,29628996	0,83074735	0,2476318	0,84733041	0,15266959
1217,99567	0,22638712	-0,46887967	0,73282738	-0,13619292	0,82668375	0,17331625
730,760262	0,12105814	0,4698289	-0,01065334	-0,64043375	0,64566315	0,35433685
1117,64437	-0,15552039	0,86935139	-0,01520568	0,21064982	0,82456299	0,17543701
1450,57389	-0,08673732	0,83320518	0,17773095	-0,21217187	0,77835942	0,22164058
	2,70220318	2,48113656	2,01477628	2,01388394	9,21199996	2,78800004

**Nota.** Resultado del PCA, Tomado de los datos de control PCA

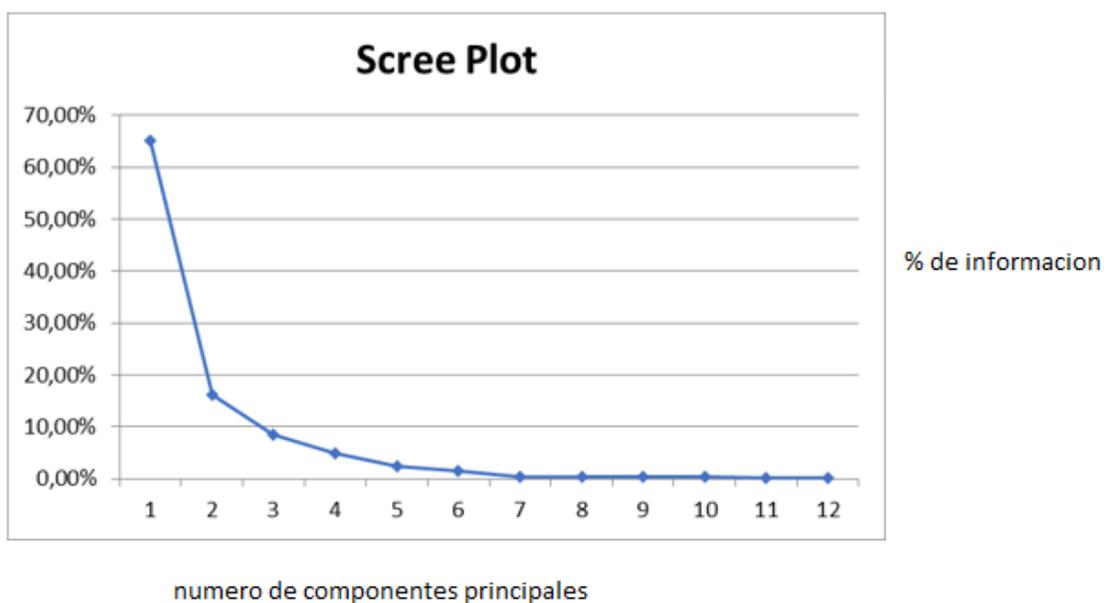
Este análisis de componentes principales se realiza a los datos del proceso antes y después de la intervención mecánica ejecutada en el equipo, para tener la referencia en el entrenamiento de la red y en el posterior test realizado a dicho entrenamiento, esto me va indicar el cambio en los valores de proceso y muestra la tendencia marcada hacia la falla. La Tabla 17, muestra que en la componente 1,2 y 3 se encuentra el 89.66% de la información, ubicada en 12 variables.

En las columnas de las variables 1, 2 y 3 por ejemplo, se observa el porcentaje de información que arroja cada una de las variables del proceso del 100% de información recolectada.

En la columna 'commun' registra la cantidad total de información arrojada por esas dos componentes en todas las variables y en la columna 'specific' muestra el valor más alto de porcentaje de información para cada una de las variables, en los 12 componentes principales

Como se muestra en la gráfica siguiente:

**Figura 23.** Componentes principales PCA



Nota. La Figura 23 enseña el volumen de información de estos dos componentes en todas las variables frente al porcentaje gradual de cada variable. Tomado de los datos de control PCA

Así mismo enseña, la reducción de variables a 4 componentes principales: 1 Componente guarda el 65,20% de la información, 2 componente guarda el 16,06% de la información, 3 componente guarda el 8,40% de la información y 4 componente guarda el 4,76% de la información, también lo refleja la siguiente Tabla.

**Tabla 18.** Porcentaje de información en cada componente  
Scree Plot

eValue	%	Cum %
7,8244132	65,20%	65,20%
1,92771388	16,06%	81,27%
1,00807258	8,40%	89,67%
0,57125071	4,76%	94,43%
0,28872784	2,41%	96,83%
0,17282182	1,44%	98,28%
0,05646319	0,47%	98,75%
0,0432549	0,36%	99,11%
0,04027114	0,34%	99,44%
0,0366405	0,31%	99,75%
0,01564417	0,13%	99,88%
0,01472606	0,12%	100,00%

12

Nota. La Tabla 18 refleja, la reducción del porcentaje de variables a 4 componentes principales. Tomado de los datos de control PCA

Adicional, arroja el valor de los porcentajes que reflejan la cantidad de información que se tiene por cada componente y en cada una de las variables. El mayor porcentaje de información se halla en las 4 primeras componentes, pero con la cantidad de información reflejada en las componentes 1, 2 y 3 es suficiente e importante para desarrollar los cálculos de esta investigación.

### 6.5 Fase 5: Interpretación de los datos y Entrenamiento de la Red SOM

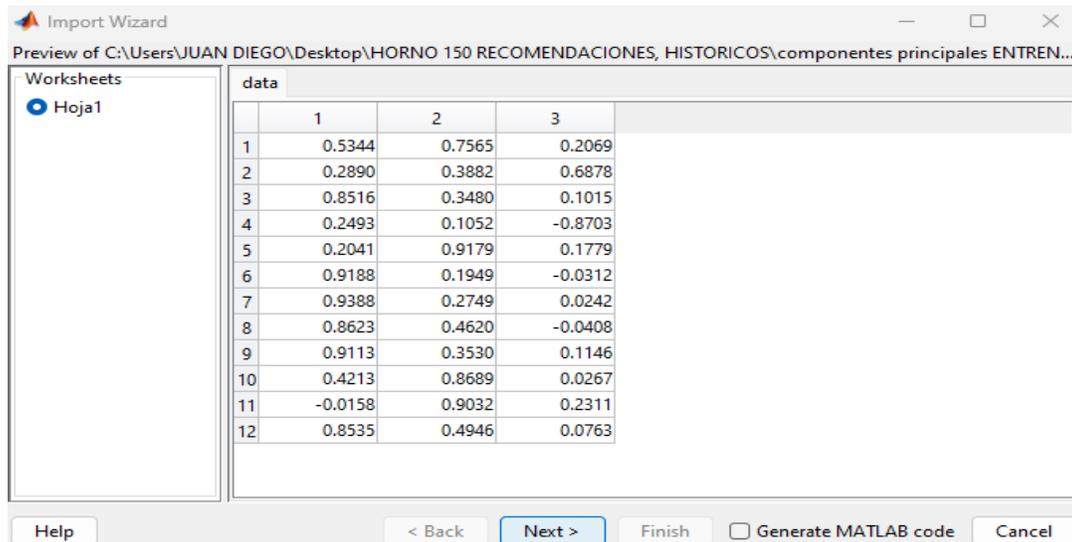
Después de realizado los análisis estadísticos de correlación de variables y de componentes principales, tenemos para el entrenamiento de la red SOM las componentes principales 1, 2 y 3 de PCA. Se utilizaron los algoritmos disponibles en la tool box SOM de Matlab. Es importante resaltar que una red SOM se caracteriza por tener la capacidad de descubrir características similares en un conjunto de datos, esto la convierte en una solución ideal para esta investigación.

**Tabla 19:** Componentes principales utilizadas para entrenamiento de la red

	1	2	3
FLUJO DE CARGA SERP B	0,53449481	0,75675153	0,20274976
FLUJO DE CARGA SERP B % VALVULA	0,29035377	0,38802289	0,68797441
FLUJO DE GAS A QUEMADORES	0,85241597	0,34618827	0,09745982
PRESION HOGAR H-150	0,25017293	0,10528503	-0,86897003
PRES ENTRADA SERP B	0,20192671	0,91891201	0,17432411
TEMP SALIDA SERP B	0,93125666	0,20161614	-0,02971128
TEMP PIEL TUBO 1 SERP B	0,94115483	0,26988115	0,01621581
TEMP PIEL TUBO 12 SERP B	0,86314193	0,45988775	-0,04800879
CAM COMBUST SERP B	0,91433478	0,34943182	0,10784243
TEMP GASES DE CHIMENEA	0,41863926	0,87021659	0,02132203
TEMP CHOQUE SERP B	-0,01720448	0,90309904	0,22786199
TECHO INCLINADO RADIACION B	0,85548006	0,4925105	0,06919853
	5,44139833	3,9610808	1,38165718

*Nota.* Tabla 19 que registra los componentes principales 1,2, y 3 para entrenamiento en red SOM. Tomado de tool box SOM de Matlab

Se utilizó en la herramienta de MATLAB, una red SOM de una capa con 10 neuronas. Una sola capa competitiva que puede agrupar un conjunto de datos en tantas clases como neuronas tenga la capa. Se carga en la red los datos de las componentes principales con las cuales se hará el entrenamiento.

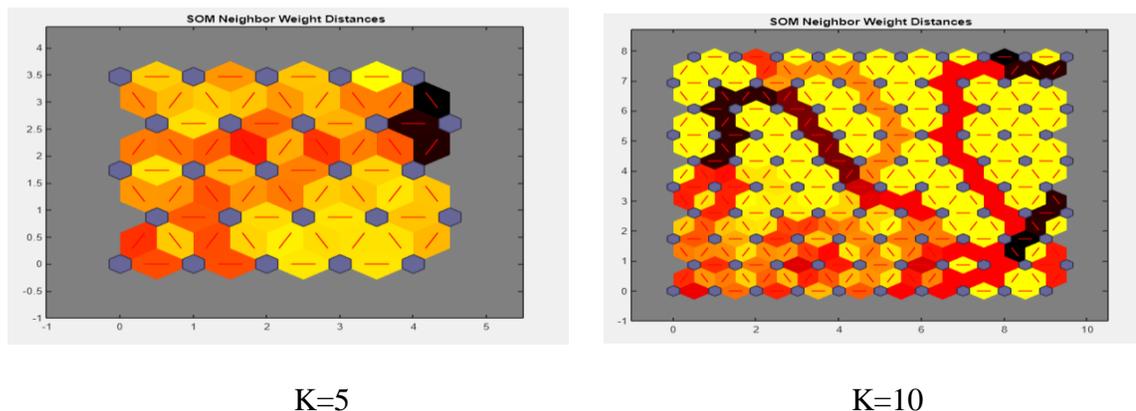
**Figura 24.** Cargue de los datos en la red SOM


	1	2	3
1	0.5344	0.7565	0.2069
2	0.2890	0.3882	0.6878
3	0.8516	0.3480	0.1015
4	0.2493	0.1052	-0.8703
5	0.2041	0.9179	0.1779
6	0.9188	0.1949	-0.0312
7	0.9388	0.2749	0.0242
8	0.8623	0.4620	-0.0408
9	0.9113	0.3530	0.1146
10	0.4213	0.8689	0.0267
11	-0.0158	0.9032	0.2311
12	0.8535	0.4946	0.0763

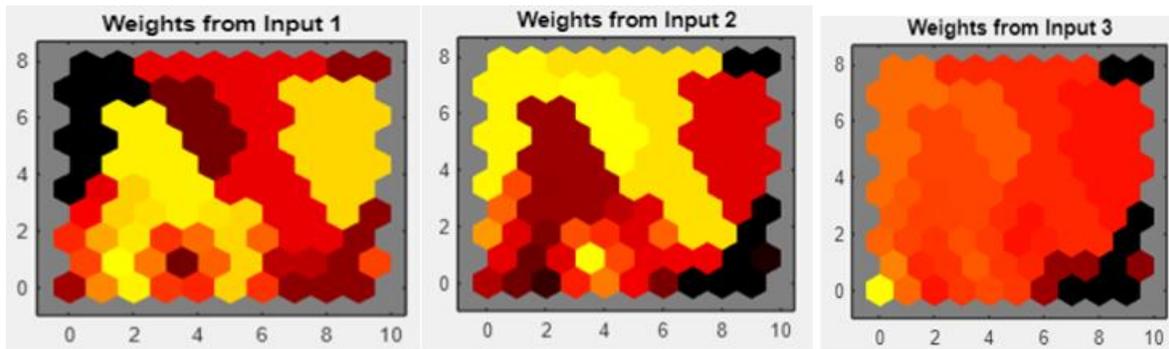
Nota. Tabla 24, muestra los Registros migrados a red SOM Tomado de tool box SOM de Matlab.

### 6.5.1 Explorar datos y resultados en el gráfico de respuesta

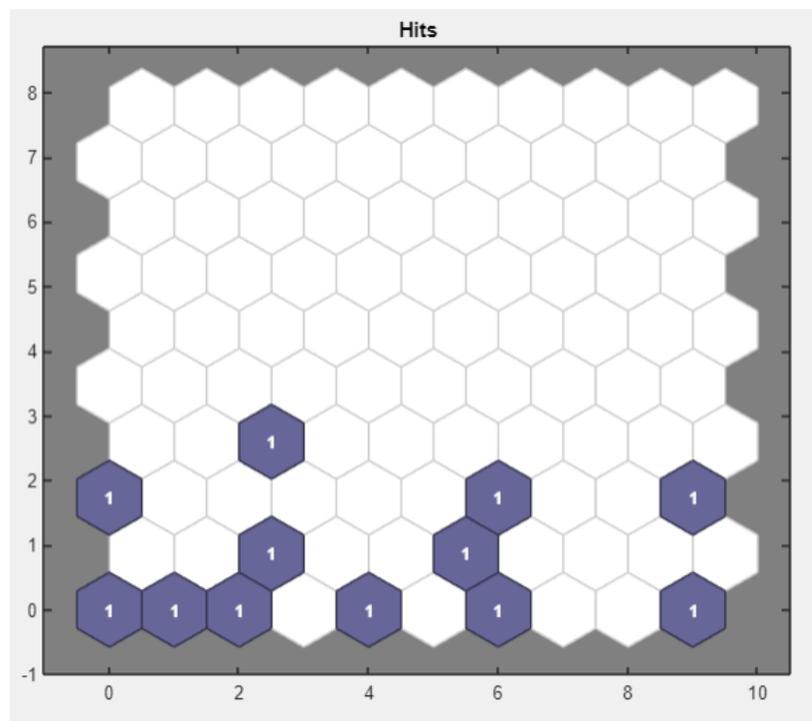
Después de correr el modelo muestra los resultados en 4 gráficas, La Figura 25 muestra los pesos de las neuronas cercanas, realizando el proceso de entrenamiento con 5 y 10 neuronas, donde la franja negra o más oscura representa mayor espacio entre las neuronas y los sectores claros o más claros un espacio menor entre ellas.

**Figura 25.** Pesos de las neuronas cercanas

Nota. Las imágenes mostradas en la Figura 26 enseñan el mayor porcentaje de información encontrados en los sectores más oscuros. Tomado de tool box SOM de Matlab

**Figura 26.** Pesos de cada una de las entradas

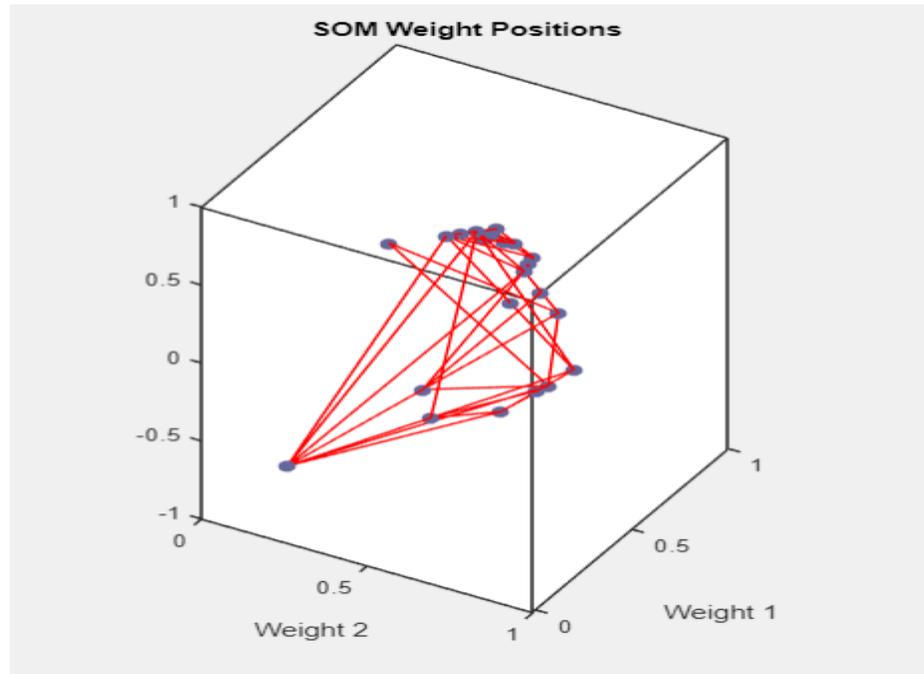
*Nota.* Observamos los pesos de cada una de las entradas, para nuestra investigación tenemos dos entradas con 12 observaciones por entrada. Las zonas más oscuras en cada una de las entradas representan las neuronas que se activan para cada una de esas columnas con respecto a todo el conjunto de datos de entrenamiento. Tomado de la red SOM

**Figura 27.** Evaluación del modelo según los residuos

Nota. La Figura 27 muestra las neuronas de color azul en donde se guarda toda la información y sus pesos por relevancia de acuerdo al porcentaje de la misma. Se observa la

organización de la información en 12 neuronas diferentes con una característica en cada una de ellas. Tomado de tool box SOM de Matlab.

**Figura 28.** Rotación posiciones de neuronas y pesos de la información en 3D.



*Nota.* La Figura 28 muestra la información de color verde, y las neuronas de color azul en las tres dimensiones. Las líneas rojas la conexión entre cada una de las neuronas, nótese que hay una agrupación de neuronas en la parte superior, en las cuales se concentra la mayor cantidad de información de acuerdo a los datos originales del proceso. Tomado de tool box SOM de Matlab.

En estas 4 gráficas se muestra el entrenamiento del mapa auto organizado con la información del proceso.

## 6.6 Fase 6: Almacenamiento de datos

Dentro de la fase de almacenamiento de datos, se busca el aseguramiento de esta información para uso futuro. Los datos guardados corresponden a los datos iniciales donde el equipo que hace parte de esta investigación opera en óptimas condiciones, Es importante este almacenamiento porque con dicha información se realiza el entrenamiento de la red.

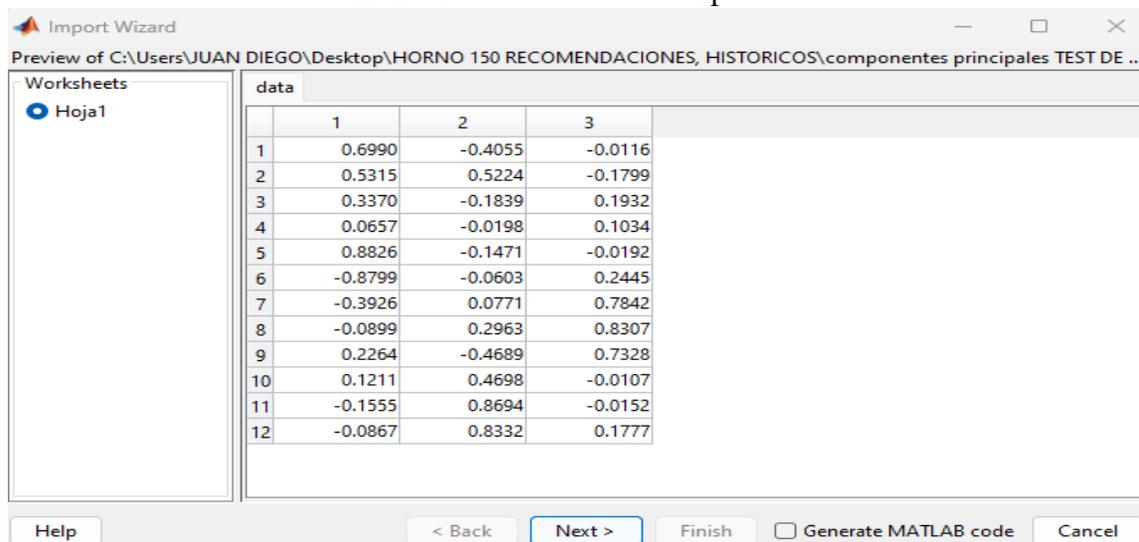
Es importante tener en cuenta lo importante que dicha información debe estar segura y actualizada dependiendo del proceso y operación del equipo de acuerdo a las futuras intervenciones.

## 7. Análisis del Desarrollo del Proyecto

Después de haber implementado y entrenado la red SOM con la ayuda de la herramienta de Matlab, el procedimiento siguiente era el de realizar las respectivas pruebas con la información antes de la intervención mecánica al horno, donde sus valores de proceso se encontraban por fuera de las ventanas operativas, buscando que dicha red mostrara las variaciones de acuerdo a esa nueva data, y mostrara la tendencia en el mapa.

Nuevamente se descarga la información del DCS con la aplicación de Excel PI DATA LINK desde el 12 de enero de 2022 a 20 de noviembre de 2022.

**Tabla 20.** Datos actualizados para Test

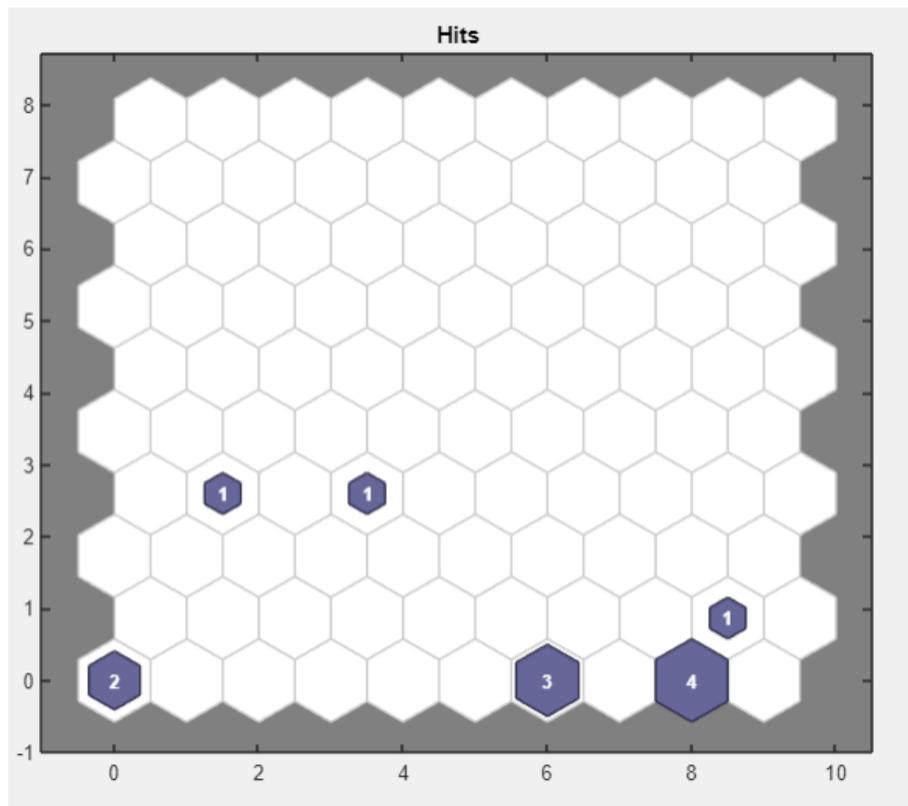


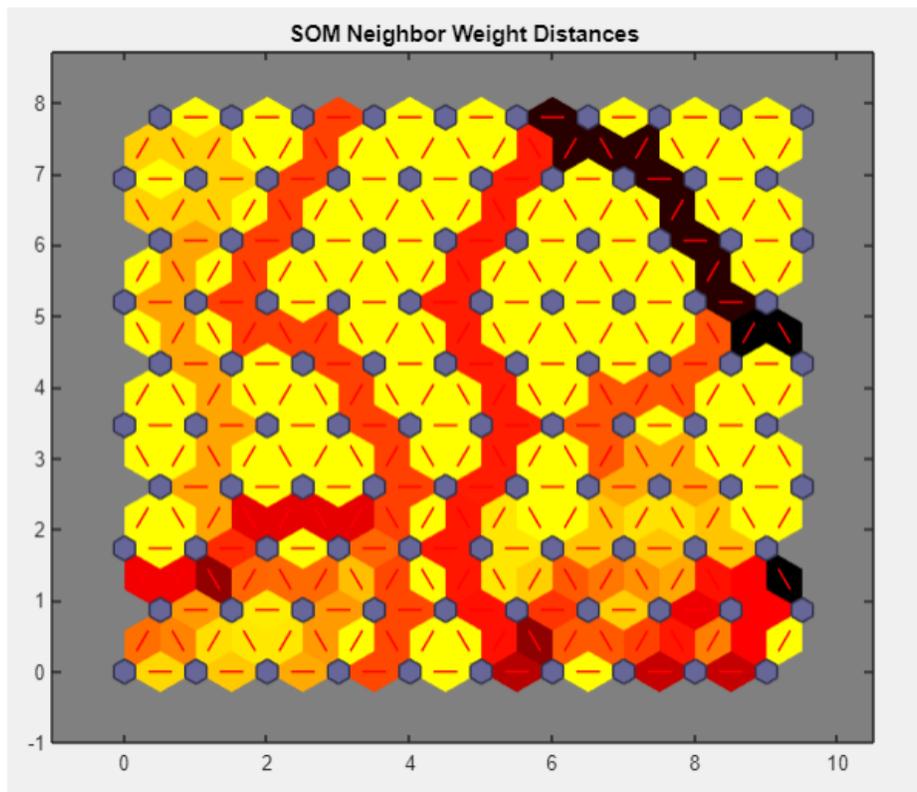
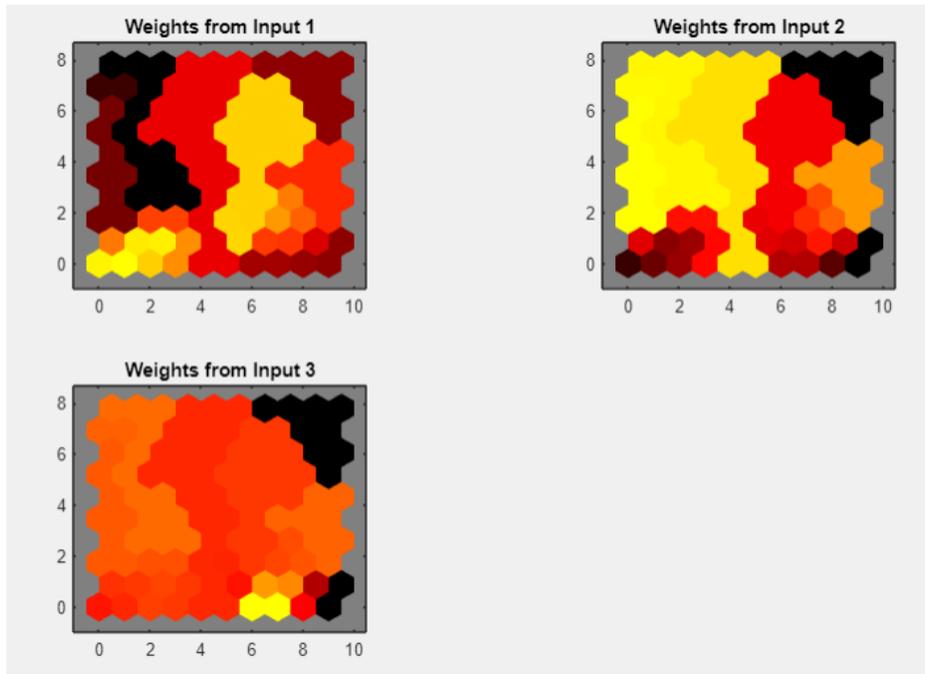
The screenshot shows an 'Import Wizard' window with a preview of data. The data is organized into a table with columns labeled 1, 2, and 3, and rows numbered 1 through 12. The values are as follows:

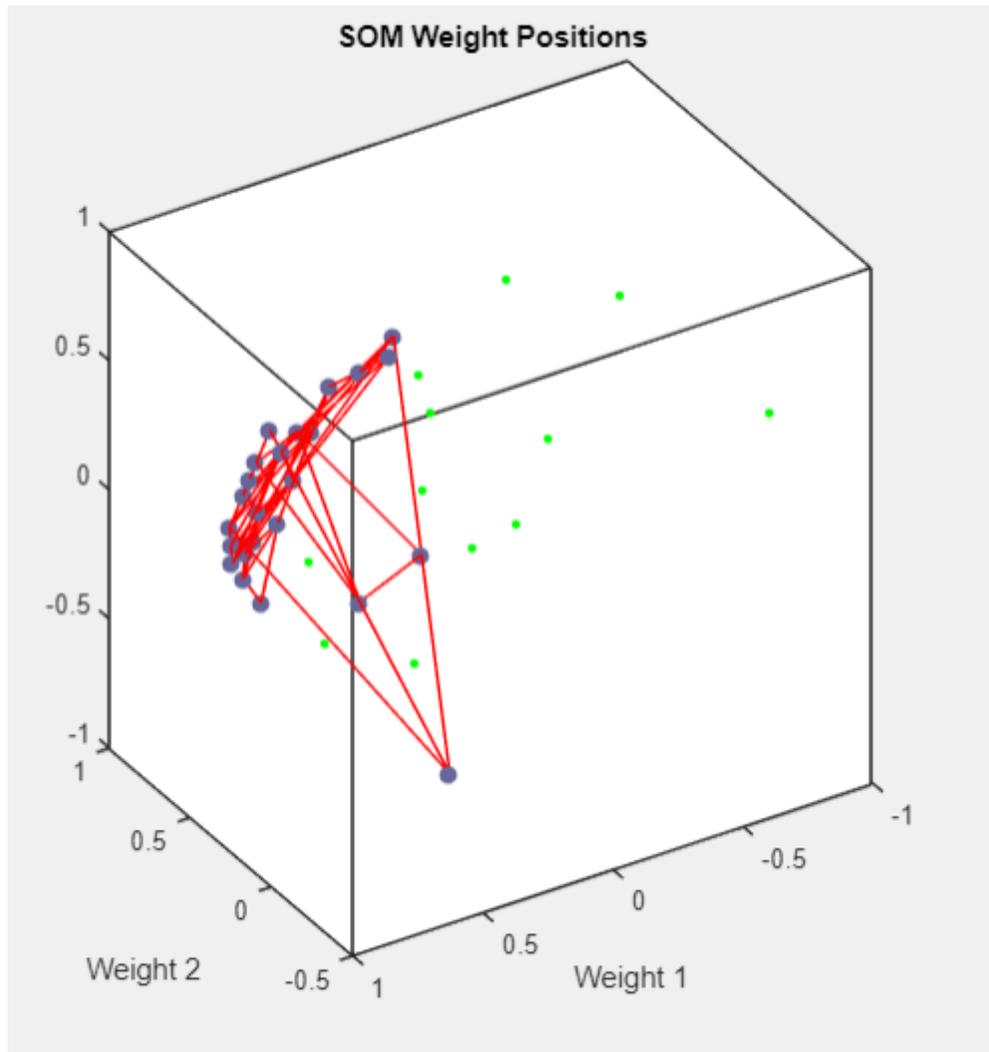
	1	2	3
1	0.6990	-0.4055	-0.0116
2	0.5315	0.5224	-0.1799
3	0.3370	-0.1839	0.1932
4	0.0657	-0.0198	0.1034
5	0.8826	-0.1471	-0.0192
6	-0.8799	-0.0603	0.2445
7	-0.3926	0.0771	0.7842
8	-0.0899	0.2963	0.8307
9	0.2264	-0.4689	0.7328
10	0.1211	0.4698	-0.0107
11	-0.1555	0.8694	-0.0152
12	-0.0867	0.8332	0.1777

Nota. La Tabla 20 refleja la información con los datos actualizados del archivo de Excel donde fueron descargados del sistema. Tomado de los datos de control PCA

La fig. 28, muestra las neuronas de color azul después de realizar el test, en donde se observa que esa nueva información con la cual se está evaluando la red, diferente a la figura 27, se organizó en 6 neuronas diferentes, en dos de ellas están la mayoría de características, 3 y 4 respectivamente y las demás características distribuidas en las otras cuatro neuronas. Tomado de tool box SOM de Matlab.





**Figura 29.** Datos y resultados en el gráfico de respuesta

Nota. La Figura 29 muestra la tendencia de los valores del proceso marcada con los puntos verdes, por fuera de los valores de control, de acuerdo al test realizado a la red. Tomado de tool box SOM de Matlab

También se nota una distancia considerable entre cada punto y la red. Esto indica valores demasiado altos en el proceso que indican que las condiciones de operación están fuera de los límites operacionales y se puede estar generando ya un coquizamiento en dichos serpentines.

## 8. Conclusiones

- El principal modo de falla que se puede determinar con la red neuronal de acuerdo al aumento en las temperaturas de piel de tubo por encima de 1300 F es el generado por la disminución del diámetro de la tubería de los serpentines debido a la formación de coque en las paredes de los mismos, este aumento en la temperatura se ve reflejado en la tendencia registrada por la red luego de realizar la prueba con los datos de proceso después de la intervención mecánica realizada al equipo.
- La red neuronal implementada en esta investigación para el horno H-150 es capaz de identificar el coquizamiento en los serpentines cada vez que los resultados arrojados de acuerdo al comportamiento de temperaturas y presiones estén por fuera de las guías y ventanas de operación.
- Se realizó un análisis estadístico de las variables de operación asociadas al Horno H-150 y se encontraron 12 variables relevantes para dicho análisis, donde se encontró que estas variables no tienen correlación entre ellas, todas y cada una de ellas aportan información importante para el estudio de investigación como insumo para el diseño de modelos de detección y clasificación de fallas.

## 9. Recomendaciones

- De acuerdo a la investigación realizada, se debe asegurar con operaciones la operación del equipo dentro de las ventanas operativas de diseño, para evitar la aceleración del modo de falla visto en la presente investigación.
- Tener en cuenta el comportamiento de la red y la información arrojada por la misma para asegurar la operación del H-150 mitigando los riesgos de una falla mayor y asegurando el proceso para la intervención por parte del departamento de mantenimiento, reduciendo los costos de la misma.
- Hacer uso de esta aplicación en el proceso, esto ayudara a realizar el seguimiento a la tendencia del mismo basado en el comportamiento de sus variables, así poder anticipar acciones para corregir desviaciones o mitigar riesgos de daño prematuro en el equipo.

## Referencias

- Amad, J. (10 de 05 de 2021). *Ciencias de datos*. Obtenido de <https://www.cienciadedatos.net/documentos/py35-redes-neuronales-python.html>: <https://www.cienciadedatos.net/documentos/py35-redes-neuronales-python.html>
- Burruecos, D. (2022). *Arquitectura redes neuronales*. Vasco: Euskal Herriko Unibertsitatea.
- Cadena, J., & Pérez, S. (2008). Aplicación de redes neuronales probabilísticas en la detección de fallas incipientes en transformadores. *Scientia et Technica* , 48-53.
- Corrales B., L. (2013). Clasificación de fallas con redes neuronales para grupos electrógenos. *Ingeniería Enérgica*, 34.
- Ecopetrol. (15 de Julio de 2022). *Portal web Ecopetrol SA*. Obtenido de <https://nuevoportal.ecopetrol.com.co/wps/portal/es/ecopetrol-web/nuestra-empresa/quienes-somos/lo-que-hacemos/refinacion/complejo-barrancabermeja> Citado (15 de julio 2022)
- Ferrer, M. (2004). Sistema de detección e Identificación de fallas aplicando inteligencia artificial. <http://virtual.urbe.edu/tesispub/0091020/cap02.pdf>.
- Freire, E. (2019). Redes Neuronales. *Bootcampai.medium.com/redes neuronales*.
- Freire, E., & Sarahi, S. (2019). Redes Neuronales. *Bootcampai.medium.com/redes neuronales*.
- Hernández S, R., Baptista L, M. d., & Fernández Collado, C. (2014). *Metodología de la Investigación*. México: Mac Graw Hill.
- Marvarez, R. (2006). Sistema de detección y diagnóstico de fallas basado en neuronas neodifusas y un filtro de detección. . *Doctoral Dissertation, Universidad Dr. Rafael Bello Chacín*.
- Matrich, D. (2001). *Redes Neuronales: Conceptos básicos y aplicaciones*. M+exico: Universidad Nacional Tecnológica de Mexico.
- Moreno, A. (2009). Desarrollo de una interfaz gráfica de redes neuronales usando Matlab. *e-archivo.uc3m.es*, 28-30.
- Ponce, J. C., & Torres Soto, Aurora; Quezada Aguilera, Fátima Sayu. (2014). *Inteligencia Artificial*. México: Letras Mexicanas.
- Refinería Barrancabermeja. (13 de 02 de 2018). <http://iris/contenido/contenido.aspx?catID=278&conID=48123>. Recuperado el 13 de 08 de 2022, de <http://iris/contenido/contenido.aspx?catID=278&conID=48123>.
- Rodrigo, J. (2017). Análisis de Componentes Principales (Principal Component Analysis, PCA). *SNE*, 431-433.
- Rovisora, A. (15 de septiembre de 2020). La Inteligencia Artificial, apuesta obligada para la industria del Petróleo. *Global Energy México*, pág. 2.

- Rozo, F. (2008). Detección y diagnóstico de fallas en procesos industriales usando análisis (PCA). *Revista Colombiana de Tecnologías Avanzadas*.
- Rozo, F. (2008). Detección y diagnóstico de fallas en procesos industriales utilizando análisis de componentes principales (PCA). *Revista Colombiana de Tecnologías Avanzadas*, 11-12.
- Rozo, F. (2008). Detección y diagnóstico de fallas en procesos industriales utilizando análisis de componentes principales (PCA). *Revista Colombiana de Tecnologías Avanzadas*.
- Rozo, F. (2008). Detección y diagnóstico de fallas en procesos industriales utilizando análisis de componentes principales (PCA). *Revista Colombiana de Tecnologías Avanzadas*.
- Salomon Associates. (2020). *Informe de competitividad internacional refinerías*. <https://www.portafolio.co/economia/barranca-una-refineria-competitiva-solomon-510930>.
- Valmi, S., Diessnack, M., & Mendes, I. (2007). Latino.am Enfermagem: Revisión de diseños de investigación. *USP Brasil*.
- Wikipedia. (19 de 03 de 2020). *Wikipedia.org*. Obtenido de <https://es.wikipedia.org/w/index.php?title=MATLAB&action=edit>: <https://es.wikipedia.org/w/index.php?title=MATLAB&action=edit>

