



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



Universidad de Buenos Aires Facultad de Ciencias Económicas Escuela de Estudios de Posgrado

MAESTRÍA EN GESTIÓN ESTRATÉGICA DE SISTEMAS Y TECNOLOGÍA DE LA INFORMACIÓN

TRABAJO FINAL DE MAESTRÍA

“Machine Learning” como herramienta de mejora
continua en la gestión por procesos

AUTOR: ANDRÉS LIONEL MYSZOV

DIRECTOR: ERNESTO CHINKES

OCTUBRE -2024



Resumen

Este trabajo incorporó un modelo de “machine learning” como recurso para la mejora continua en la gestión del proceso de búsqueda sistemática de fugas en una empresa distribuidora de gas natural en Argentina.

Los procesos pueden clasificarse en estratégicos, operativos o de soporte. El proceso abordado en este estudio es de tipo operativo y se utilizó el ciclo de mejora continua como herramienta de gestión. El objetivo de este procedimiento fue monitorear y mejorar las actividades mediante la medición de ciertas variables que describen su funcionamiento.

Se emplearon conocimientos de estadística de negocios y algoritmos de “machine learning” para identificar patrones y predecir situaciones futuras. Los datos del proceso se utilizaron en modelos de “machine learning” para evaluar y predecir el comportamiento de la variable estudiada.

Dentro del proceso de búsqueda sistemática de fugas, se analizó la cantidad de fugas de grado 1 y su distribución en el tiempo con el fin de mejorar ciertos indicadores relacionados con esta variable. Se enumeraron las ventajas de aplicar un algoritmo de predicción al proceso, lo cual permite optimizar determinadas actividades dentro del mismo.

Palabras Clave

Procesos, gestión de procesos, ciclo de mejora continua, machine learning, series de tiempo, predicciones, indicadores, arima y sarima.

Clasificación JEL

M15, C10, C22, C5, C88, O31.



1 Contenido

Resumen	2
Palabras Clave	2
Clasificación JEL.....	2
1 Introducción.....	8
1.1 Justificación / Fundamentación.....	8
1.2 Planteamiento del tema	9
2 Objetivos.....	10
2.1 Objetivo general.....	10
2.2 Objetivos específicos	10
3 Metodología y técnicas para utilizar.....	11
4 Marco teórico.....	12
4.1 Gestión por procesos.....	12
4.1.1 Los procesos como base de la gestión en las organizaciones.....	13
4.1.2 Sistema de gestión basado en procesos	14
4.2 Ciclo de mejora continua	26
4.3 Series de tiempo.....	29
4.3.1 Tratamiento de las series de tiempo	29
4.3.2 El modelo ARIMA	30
4.3.3 Metodología Box-Jenkins No estacional.....	32
4.3.4 Modelos puramente estacionales: SARIMA.	46
4.3.5 Modelo multiplicativo.	46
4.4 Machine Learning	47
4.4.1 Diferentes algoritmos según su clasificación	48
4.4.2 Algoritmos no supervisados	51
4.4.3 Algoritmos supervisados	51
4.4.4 Implementación de algoritmos para predicciones de series de tiempo	55
4.5 Machine learning en la gestión de procesos	55
5 Desarrollo	58
5.1 Reseña de la empresa.....	58
5.1.1 Marco regulatorio	58
5.1.2 Servicio Ofrecido.....	59
5.2 Enfoque en procesos	60
5.3 Gestión de procesos operativos.....	63
5.4 Ciclo de mejora continua	64



5.4.1	Planificación global de procesos	64
5.4.2	Planificación individual de procesos	68
5.4.3	Ejecución del proceso	75
5.4.4	Actividad búsqueda sistemática: ubicación de la fuga	77
5.4.5	Verificación del proceso	80
5.4.6	Actuación para la mejora del proceso.....	84
5.5	Uso de “machine learning” para optimizar el proceso búsqueda sistemática de fugas media presión	84
5.5.1	Escenario inicial	85
5.5.2	De lo general a lo particular con “machine learning”	86
5.6	Como trabajar el modelo a futuro	104
5.7	Ventajas para la organización	105
5.7.1	Mejora la planificación de los recursos	105
5.7.2	Mejora en la calidad del servicio	106
5.7.3	Reducción de costo	106
6	Conclusiones.....	107
7	Bibliografía.....	108
8	Anexos.....	113
8.1	Anexo A.....	113
8.1.1	La Gestión basada en procesos: el enfoque ISO 9001.....	113
8.1.2	La Gestión basada en procesos: el modelo EFQM.....	114
8.2	Anexo B	115
8.2.1	Algoritmo desarrollado en Python.....	115



Índice de figuras

Figura 1 Procesos, procedimientos e instrucciones de trabajo	13
Figura 2 El sistema de gestión basado en procesos	14
Figura 3 Principales factores para la identificación y selección de procesos.....	15
Figura 4 Mapa de procesos: esquema general	16
Figura 5 Esquema de descripción de procesos a través de diagramas y fichas.....	17
Figura 6 Ejemplo de ficha para un proceso de revisión de requisitos del producto.....	18
Figura 7 Diagrama de proceso de admisión alumno posgrado	19
Figura 8 Tabla de indicadores y rendimiento	22
Figura 9 Tabla de indicadores, rendimientos e importancia.....	22
Figura 10 Ejemplo genérico de matriz pareada.....	23
Figura 11 Priorización de indicadores del proceso "Preparación y distribución de pedidos"	24
Figura 12 Método sistemático de mejora de procesos.....	25
Figura 13 Ciclo de mejora continua	27
Figura 14 Gestión siguiendo la diagonal perversa	28
Figura 15 Metodología BOX – Jenkins.....	33
Figura 16 Serie de tiempo estacionaria y no estacionaria	34
Figura 17 Función de autocorrelación se trunca.....	35
Figura 18 Obtención de las primeras diferencias en una serie	36
Figura 19 Transformación de la serie de tiempo mediante las primeras diferencias	36
Figura 20 Transformación de la serie de tiempo mediante las primeras diferencias	38
Figura 21 Datos de Eviews para análisis de serie.....	39
Figura 22 Evaluación de estacionalidad	39
Figura 23 Características FAS y FAP teóricas	40
Figura 24 Categorización de niveles de contraste	45
Figura 26 Tipos de “machine learning” clásico, clasificación de sus algoritmos según su uso y aplicaciones.....	48
Figura 27 Tipos de aprendizaje automático según el tipo de datos.....	49
Figura 28 Tipos de aprendizaje automático según la estructura de los datos.....	50
Figura 29 Tipos de aprendizaje automático según los algoritmos utilizados.....	51



Figura 31	Instrucciones del plan de trabajo MAN-VEH.....	61
Figura 32	Cadena de valor de la empresa.....	64
Figura 33	Mapa de procesos operativos	65
Figura 34	Desglose del mapa de procesos de la gerencia de operaciones.....	67
Figura 35	Proceso búsqueda sistemática de fugas media presión	69
Figura 36	Ficha del proceso Búsqueda Sistemática de Fugas Media Presión.....	71
Figura 37	Ficha del indicador Fugas grado 1 por km de cañería en MP	72
Figura 38	Cantidad de fugas grado 1 reparadas	73
Figura 39	Costo medio por fugas grado 1 reparadas por terceros.....	74
Figura 40	Cumplimiento de frecuencia de mantenimientos preventivos previstos por el ENARGAS	75
Figura 41	Sondeo para fuga en servicio	78
Figura 42	Sondeo para fuga en cañería principal	79
Figura 43	Importancia de indicadores del proceso.....	81
Figura 44	Resumen rendimiento de indicadores	83
Figura 45	Rendimiento del proceso “Búsqueda sistemática de fugas en MP”	83
Figura 46	Modelo de proyecto para mantenimiento predictivo	87
Figura 47	Serie de tiempo de fugas grado 1	89
Figura 48	Serie de tiempo resumida en cada mes.....	90
Figura 49	Serie de tiempo resumida en cada mes a través de un gráfico de cajas.....	91
Figura 50	Serie de tiempo resumida en cada mes.....	91
Figura 51	Fugas grado 1 por mes evaluadas a través de los años	92
Figura 52	Histograma de densidad de fugas grado 1	93
Figura 53	Estadístico Q de Ljung-Box.....	94
Figura 54	Test de Dickey - Fuller para la base de datos	95
Figura 55	Test de Dickey - Fuller para la base de datos diferenciada una vez	96
Figura 56	Gráfico de autocorrelación simple de la serie diferenciada una vez.....	97
Figura 57	Gráfico de autocorrelación parcial de la serie diferenciada una vez.....	97
Figura 58	Serie de datos dividida en entrenamiento y test.....	98
Figura 59	Salida de autoarima.....	98
Figura 60	Métricas del modelo autoarima.....	99



Figura 61 Modelo Arima - Sarima observando los gráficos de autocorrelación simple y parcial diferenciada una vez	99
Figura 62 Métricas del segundo modelo	99
Figura 63 Segunda interpretación de los gráficos de autocorrelación.....	100
Figura 64 Métricas del tercer modelo.....	100
Figura 65 Resumen de métricas de los modelos	100
Figura 66 Ajuste del modelo elegido	101
Figura 67 Gráfico de autocorrelación de los residuos	102
Figura 68 Ajuste en test y pronóstico	102
Figura 69 Valores de la predicción del año siguiente.....	103
Figura 70 Modelo de un sistema de gestión de la calidad basado en procesos (según ISO 9001:2000).....	113
Figura 71 Estructura del modelo de excelencia de la EFQM.....	114



1 Introducción

1.1 Justificación / Fundamentación

Con el objetivo de obtener valor, las empresas entregan servicios o productos a un segmento de mercado denominado clientes finales. Para generar valor, proponen ejecutar determinadas actividades de forma eficiente dentro del rubro en el que trabajan. Esto lleva a las organizaciones a conocer sus actividades, aunque no siempre entienden cómo se agrupan en procesos.

En primer lugar, los procesos agrupan actividades específicas que la organización realiza. Identificar y gestionar estos procesos permite a la empresa obtener un mayor valor agregado a sus actividades. Aunque este tema está ampliamente difundido, no todas las empresas conocen su cadena de valor ni trabajan para mejorar sus procesos. Dada la importancia de los procesos en las empresas, su gestión es esencial.

Es importante destacar lo que mencionan Pérez Grau y Zabaleta de Armas (2015) cuando relacionan el gerenciamiento de una organización con la gestión por procesos: “Las organizaciones que no tienen una gestión por procesos atienden sus necesidades a través de sectores especializados, lo que genera una cadena de mando y una estructura organizacional del tipo vertical” (Pérez Grau y Zabaleta de Armas, 2015).

En segundo lugar, el avance de la tecnología y las aplicaciones obliga a cada empresa a adaptarse a nuevos desafíos en el mercado para no perder competitividad. La digitalización, mediante la incorporación de sistemas aplicados a sus actividades, hace necesario conocer con precisión los distintos procesos con los que la organización opera.

Por último, para comenzar a gestionar los procesos, primero se debe conocer el estado actual y categorizarlos de acuerdo con su nivel de madurez. En una segunda instancia, es necesario detectar los puntos críticos sobre los cuales se desea trabajar y planificar las mejoras.

La implementación de machine learning proporciona una herramienta que facilita la consecución de los objetivos, mejorando la gestión del negocio y haciendo los procesos más eficaces.

Por lo tanto, es necesario llevar a cabo una gestión eficaz de los procesos. Para ello, podemos utilizar el ciclo de mejora continua, conocido por las siglas PHVA. Este ciclo consta de cuatro fases que se realizan de forma cíclica y permite el control y la mejora de



dichos procesos.

Incluir técnicas de machine learning dentro del ciclo de mejora continua, específicamente en las etapas de verificar y actuar, permitirá a la organización adoptar una herramienta tecnológica para aprovechar los datos disponibles. Esto no solo facilitará la elaboración de informes para la toma de decisiones, sino también la integración de inteligencia en procesos, productos y servicios, lo que permitirá enfocar las nuevas estrategias de negocio en datos.

1.2 Planteamiento del tema

Para enfrentar los desafíos de las nuevas arquitecturas empresariales y su enfoque en la transformación digital, es necesario identificar las actividades que requieren un análisis minucioso de los datos para generar inteligencia de negocio. Detectar los mecanismos para mejorar los procesos es un aporte significativo en estos cambios.

En este contexto, surge una herramienta derivada de la inteligencia artificial (IA), conocida como machine learning. Esta tecnología permite, entre otras cosas, predecir estados futuros de los datos y generar inteligencia en procesos, productos y servicios.

En concordancia con estas ideas, este trabajo final se enfoca en el proceso denominado “búsqueda sistemática de fugas”. Para gestionar este proceso, puede utilizarse el ciclo de mejora continua.

Por tal motivo, se indaga en la siguiente pregunta: ¿cómo demostrar la ventaja de incorporar modelos de machine learning para la mejora continua en la gestión del proceso de “búsqueda sistemática de fugas en media presión”?

Así, surgen las siguientes preguntas: ¿cuál es el “estado del arte” en la gestión por procesos, el ciclo de mejora continua, la estadística de negocios aplicada y el machine learning? ¿Cómo se realiza y qué actividades implica ejecutar el proceso de “búsqueda sistemática de fugas en media presión”? ¿Cómo se aplica un ciclo de mejora continua y la tecnología de machine learning en el proceso seleccionado?

Esto conduce al siguiente planteamiento: ¿cómo diseñar un algoritmo para mejorar los resultados del proceso seleccionado y cuáles son sus ventajas?



2 Objetivos

2.1 Objetivo general

Incorporar un modelo de machine learning para mejorar la gestión del proceso “Búsqueda Sistemática de Fugas en Media Presión” en una empresa distribuidora de gas natural.

2.2 Objetivos específicos

- Describir los conceptos de gestión por procesos, ciclo de mejora continua, estadística de negocios aplicada y “machine learning”.
- Describir el proceso particular al cual se aplicará un modelo de “machine learning”, en este caso el proceso de “búsqueda sistemática de fugas en media presión”.
- Establecer el ciclo de mejora continua e identificar debilidades del proceso seleccionado.
- Diseñar un modelo predictivo utilizando “machine learning” que permita optimizar el proceso seleccionado.
- Enumerar las ventajas de aplicar dicho modelo en la gestión del proceso de “búsqueda sistemática de fugas en media presión”.



3 Metodología y técnicas para utilizar

Es un estudio exploratorio, explicativo y correlacional que expone cómo las herramientas de machine learning aportan una ventaja competitiva significativa en la gestión de procesos.

La investigación se basa en una selección de fuentes que incluye material bibliográfico, artículos relacionados con temas específicos y textos que sirven como marco teórico para esta tesis.

Además, se trabaja con datos secundarios extraídos de sistemas de información y fuentes abiertas de recolección. Se utilizan técnicas cuantitativas para el análisis del caso, realizando diversas operaciones con las variables.

En este trabajo, la unidad de análisis es una empresa distribuidora de gas natural en Argentina, centrándose en la gerencia de operaciones y en el proceso de detección y reparación de fugas de gas de grado 1.



4 Marco teórico

Se abordan los conceptos generales y específicos relacionados con los procesos como marco de gestión de negocios, así como la importancia del ciclo de mejora continua dentro del enfoque de análisis. Además, se revisan temas específicos de estadística vinculados al análisis de series de tiempo, que son útiles para el presente trabajo. También se analizan los tipos y los principales algoritmos de “machine learning” para el análisis de datos y su aplicación. Se profundiza en una de las herramientas descritas para implementar mejoras en el proceso seleccionado, con el objetivo de optimizar su rendimiento. Por último, se examina el estado del arte en relación con la aplicación de modelos de inteligencia artificial, en particular el “machine learning”, en la mejora de procesos.

4.1 Gestión por procesos

Los procesos, procedimientos e instrucciones de trabajo y sus diferencias las menciona Robledo (2017) de la siguiente manera:

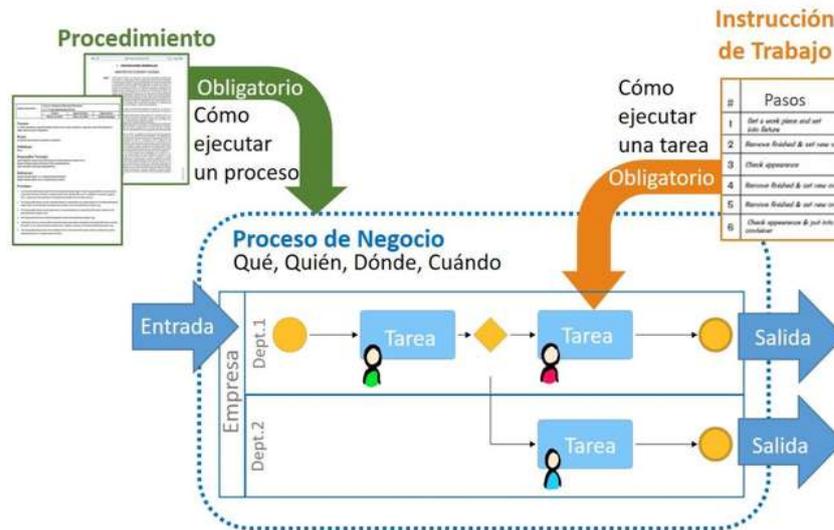
Proceso: cualquier actividad, o conjunto de actividades, que utiliza recursos para transformar elementos de entrada en resultados.

Procedimientos: la forma específica de llevar a cabo una actividad o un proceso.

Instrucciones de trabajo: son documentos que describen de manera clara y precisa la manera correcta de realizar determinadas tareas.

Figura 1

Procesos, procedimientos e instrucciones de trabajo



Fuente: Recuperado de Robledo (2017).

4.1.1 Los procesos como base de la gestión en las organizaciones

Según Club BPM (2011), el término BPM (Business Process Management) corresponde a la gestión por procesos y va más allá del aspecto tecnológico. El mismo autor menciona que “es un sistema de gestión enfocado en perseguir la mejora continua del funcionamiento de las actividades empresariales mediante la identificación y selección de procesos, así como la descripción, documentación y mejora de estos”. Además, agrega que “BPM debe estar alineado con la estrategia, con la gestión de recursos humanos, con la gestión financiera, con la gestión de la información, con la gestión de la calidad y con las disciplinas tradicionales de gestión” (Club BPM, 2011). Por último, el autor comenta que “la gestión por procesos es impulsada y hecha realidad por un conjunto de tecnologías totalmente maduras que permiten alcanzar unos resultados excelentes”.

Otros autores mencionan que se puede considerar la gestión por procesos como “la forma de gerenciar los procesos empresariales en sustitución de la gestión tradicional basada en funciones. Busca lograr la alineación de estos con la estrategia, misión y objetivos” (Medina et al., 2019).

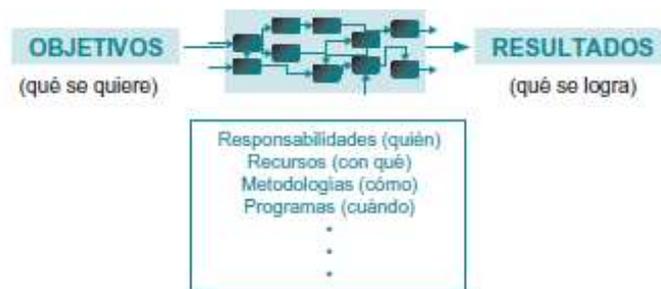
En cuanto a los procesos como base de gestión en las organizaciones, se puede agregar que “los mercados actuales seguirán exigiendo a las empresas continuas

innovaciones de productos, así como reorganizaciones estructurales. La forma más eficiente de abordar estas innovaciones es a través de reestructuraciones de los procesos clave y estratégicos de la empresa” (Zaratiegui, 1999).

Los sistemas de gestión son muy importantes en las organizaciones. Sanz et al. (2009) mencionan que establecer metodologías ayuda a obtener buenos resultados en los objetivos planteados.

Figura 2

El sistema de gestión basado en procesos



Fuente: Recuperado de Sanz et al. (2009).

4.1.2 Sistema de gestión basado en procesos

Para considerar que una organización tiene un sistema de gestión basado en procesos plantea Sanz et al. (2009) que sobre cada uno de estos se aplica una metodología de trabajo como la siguiente:

- ✚ La identificación y secuencia de los procesos.
- ✚ La descripción de cada uno de los procesos.
- ✚ El seguimiento y la medición para conocer los resultados que se obtienen.
- ✚ La mejora de los procesos con base en el seguimiento y medición realizada.

El anexo A muestra dos enfoques muy conocidos de gestión de procesos.



4.1.2.1 La identificación y secuencia de los procesos.

En primer lugar, comenta Sanz et al. (2009) que “los procesos ya existen dentro de una organización de manera que el esfuerzo se debería centrar en identificarlos y gestionarlos de manera apropiada”.

En concordancia con esta idea Zaratiegui (1999) menciona que los procesos se pueden agrupar de la siguiente manera:

Estratégicos: procesos relacionados con las metas de la empresa, sus políticas y estrategias. Dependen de la alta dirección.

Operativos: son los que convierten las estrategias en acciones para dar servicio a los clientes. De estos procesos se encargan los directores funcionales.

De apoyo o soporte: son aquellos que responden directamente a las políticas pero que su concreción ayuda a los procesos operativos.

En segundo lugar, Sanz et al. (2009) menciona algunos factores a tener en cuenta para la identificación y selección de procesos:

Figura 3

Principales factores para la identificación y selección de procesos

- Influencia en la satisfacción del cliente.
- Los efectos en la calidad del producto/servicio.
- Influencia en Factores Clave de Éxito (FCE).
- Influencia en la misión y estrategia.
- Cumplimiento de requisitos legales o reglamentarios.
- Los riesgos económicos y de insatisfacción.
- Utilización intensiva de recursos.

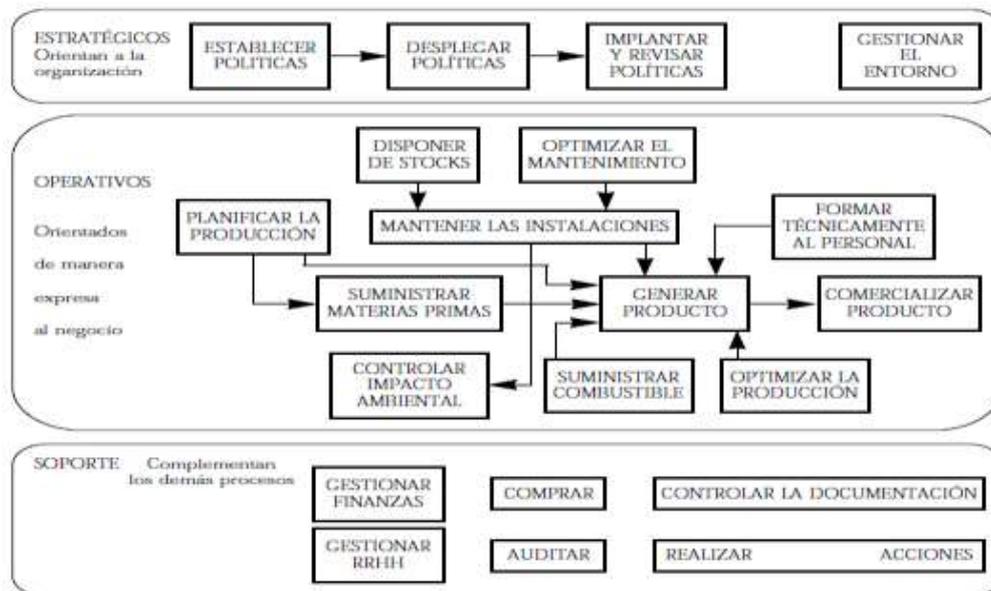
Fuente: Recuperado de Sanz et al. (2009).

Los mismos autores señalan que, una vez identificados los procesos, surge la necesidad de definir una estructura que facilite la determinación e interpretación de las interrelaciones existentes entre ellos. Esta actividad se realiza mediante un mapa de procesos.

Podemos observar que el mapa de procesos tiene una estructura que abarca las definiciones previamente vistas de procesos estratégicos, operativos y de apoyo, los cuales se relacionan en secuencias ordenadas.

Figura 4

Mapa de procesos: esquema general



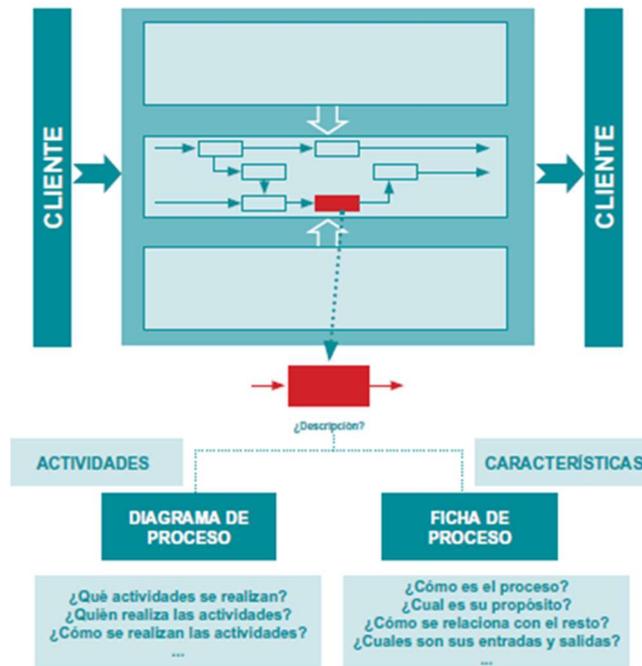
Fuente: Recuperado de Zaratiegui (1999).

4.1.2.2 Descripción de los procesos.

Según Sanz et al. (2009) “la descripción de un proceso tiene como finalidad determinar los criterios y métodos para asegurar que las actividades que comprende se llevan a cabo de manera eficaz al igual que el control de este”.

Figura 5

Esquema de descripción de procesos a través de diagramas y fichas



Fuente: Recuperado de Sanz et al. (2009).

Cada proceso tiene un nombre que representa su actividad. Además, a través de la ficha técnica, se conocen otras características, como su objetivo, variables de entrada y salida, indicadores relacionados y el responsable, entre otros datos. Un ejemplo de ficha de proceso es el siguiente:



Figura 6

Ejemplo de ficha para un proceso de revisión de requisitos del producto

INTESA		REVISIÓN DE REQUISITOS DEL PRODUCTO		FP-722
PROCESO: REVISIÓN DE LOS REQUISITOS DEL PRODUCTO			PROPIETARIO: DTOR COMERCIAL	
MISIÓN: Asegurar que los requisitos aplicables a los productos para los clientes están correctamente definidos en ofertas, pedidos y contratos, aclarados y que se tiene capacidad para cumplirlos.			DOCUMENTACIÓN: PC-722	
ALCANCE	• Empieza: Cuando empezamos cualquier relación comercial.			
	• Incluye: Ofertas, pedidos y contratos. Recogida de información para asegurar la capacidad.			
	• Termina: Con la elaboración de una oferta, aceptación de un pedido o modificación del mismo.			
ENTRADAS: Necesidades del cliente. Información sobre capacidad de producción y stock.		PROVEEDORES: Cliente. Producción. Logística.		
SALIDAS: Ofertas. Pedidos aceptados. Contratos firmados. Modificaciones a los anteriores		CLIENTES: Cliente externo.		
INSPECCIONES: Inspección mensual de las ofertas y pedidos		REGISTROS: Reclamaciones, devoluciones, FORM 722.1		
VARIABLES DE CONTROL: <ul style="list-style-type: none"> • Inmovilizado de producto final. • Capacidad de producción. • Plazo de entrega estándar. • Catálogo de productos. • Política comercial. 		INDICADORES: <ul style="list-style-type: none"> • I722.1 = % de ofertas aceptadas • I722.2 = % ofertas/pedidos/contratos no conformes • I722.3 = % modificaciones de requisitos por causa propia 		
				Revisión: 02 Fecha: 2001/02/05

Fuente: Recuperado de Sanz et al. (2009).

Existen varias técnicas de modelado de procesos. Hitpass (2012) afirma que modelar “los procesos dentro de la organización permite conocer las áreas problemáticas y susceptibles de mejora, los niveles y la delegación de autoridad, las áreas de alto riesgo, el volumen de sus operaciones y el ciclo de vida de sus procesos”.

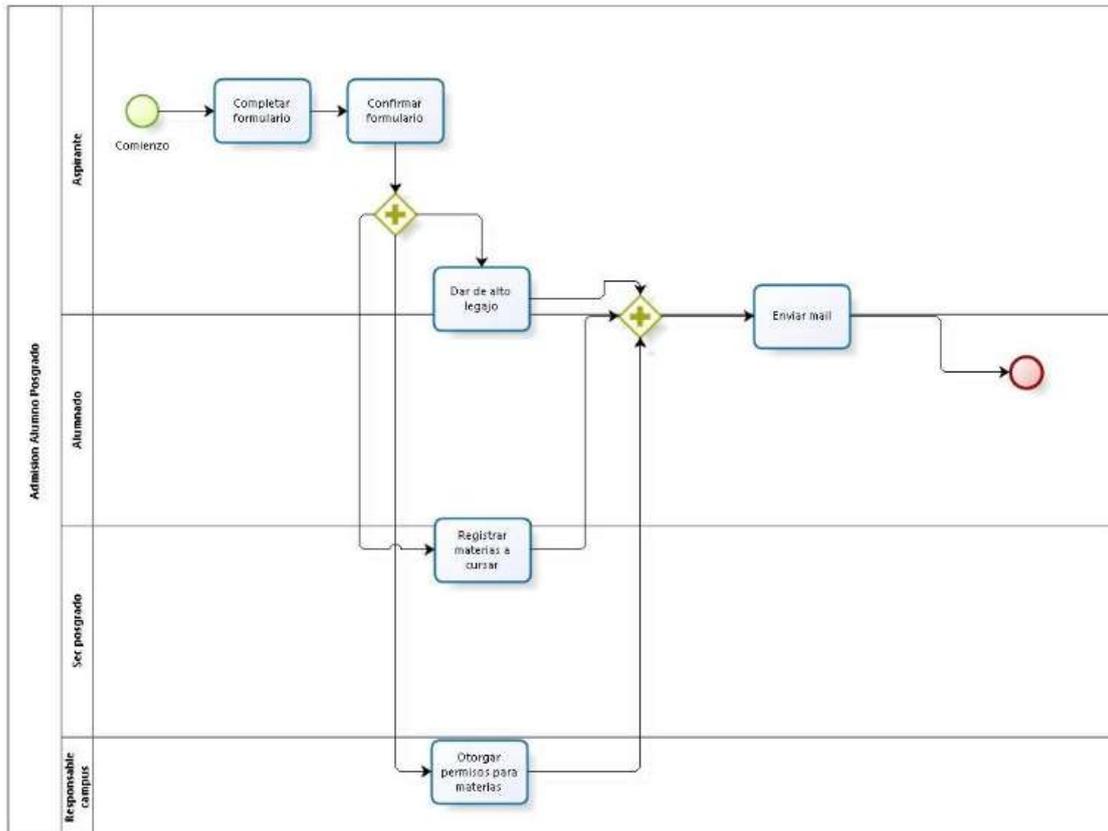
BPMN (Business Process Model and Notation) se define como un “estándar mundialmente reconocido, aceptado y aplicado en el mercado para la diagramación y especificación de procesos de negocio, desde la modelización de procesos conceptuales y lógicos hasta el diseño de procesos orientados a tecnología Workflow” (Club BPM, 2011).

En BPMN, “los procesos de negocio involucran la captura de una secuencia ordenada de actividades e información de apoyo. Modelar un proceso de negocio implica representar cómo una empresa realiza sus objetivos centrales” (Hitpass, 2012).

A continuación, se muestra, a modo de ejemplo y mediante el uso del software Bizagi, el diagrama de un proceso muy simple. El proceso se llama “admisión alumno posgrado”:

Figura 7

Diagrama de proceso de admisión alumno posgrado



Fuente: Diseño propio.

4.1.2.3 El seguimiento y la medición de los procesos.

Sanz et al. (2009) indican que un sistema de gestión tiene un enfoque basado en procesos si se preocupa por conocer sus resultados. El seguimiento y la medición constituyen la base para saber cómo se está desarrollando el proceso, si se cumplen los resultados deseados y hacia dónde se deben orientar las mejoras.

Los indicadores permiten establecer, en el marco de un proceso, qué es necesario medir para conocer la capacidad y eficacia de este, según el autor mencionado. Un indicador, como su nombre indica, debe señalar algo en relación con el objetivo.



En función de los valores que adopte un indicador y su evolución a lo largo del tiempo, la organización podrá determinar si debe actuar sobre las variables de control para cambiar el comportamiento del proceso según convenga (Sanz et al., 2009).

Estos son los pasos para seguir el establecimiento de indicadores de un proceso, según Sanz et al. (2009):

- ✚ Reflexionar sobre la misión del proceso.
- ✚ Determinar la tipología de resultados a obtener y las magnitudes a medir.
- ✚ Determinar los indicadores representativos de las magnitudes a medir.
- ✚ Establecer los resultados que se desean alcanzar para cada indicador definido.
- ✚ Formalizar los indicadores con los resultados que se desean alcanzar (objetivos) a través de una ficha de indicador.

4.1.2.4 Cálculo del rendimiento de un indicador.

Se define el “rendimiento como la relación existente entre el valor obtenido al medir un indicador y el valor límite de cumplimiento asociado a ese indicador, expresado en porcentaje” (Álvarez, 2013). Dice el mismo autor que la fórmula para su cálculo será diferente. Dependiendo del comportamiento del indicador tendremos dos casos:

1. Si un incremento del valor del indicador supone una situación favorable, el rendimiento se calculará mediante la relación:

$$\text{Rto.} = (\text{Valor del indicador} / \text{Valor límite del cumplimiento}) * 100$$

Es, por ejemplo, el caso del indicador “Porcentaje de pedidos servidos en menos de 48 horas”. Si el valor del indicador que medimos crece de una semana para otra, por ejemplo, del 96% al 98%, significa que lo estamos haciendo mejor y por tanto la situación es favorable. Para el último dato medido (98%) el rendimiento sería la relación entre este dato y el valor límite de cumplimiento. Tomando como valor de cumplimiento 95%:



$$\text{Rto. Indicador} = (98 / 95) \times 100 = 103\%$$

2. Si un incremento del valor del indicador supone una situación desfavorable, el rendimiento se calculará mediante la relación:

$$\text{Rto.} = (\text{Valor límite del cumplimiento} / \text{Valor del indicador}) * 100$$

Es el caso del indicador “número de reclamaciones por rotura o desperfecto de la mercancía”. Cualquier incremento de este indicador supone tener más reclamaciones y, por tanto, una situación desfavorable. En este caso, para un supuesto valor medido de 12 reclamaciones y un valor de cumplimiento de 10, el rendimiento se calcula como:

$$\text{Rto. Indicador} = (10 / 12) \times 100 = 83\%$$

Un rendimiento del 100% o superior significa que estamos alcanzando o superando el valor límite de cumplimiento marcado, y por tanto podríamos calificar la situación de adecuada. Si el rendimiento está por debajo del 100%, el proceso necesitará ser revisado introduciendo los ajustes y mejoras pertinentes.

La utilización de valores de rendimiento permite comparar cualquier tipo de indicador, independientemente de la unidad de medida utilizada para su cálculo.

El rendimiento se mide en porcentajes, que es una medida relativa y por tanto comparable.

4.1.2.5 Cálculo del rendimiento de un proceso.

El cálculo del rendimiento de un proceso se puede “realizar a través del rendimiento de sus indicadores. Para ello calcularemos la media aritmética o ponderada de los valores de rendimiento ofrecidos por cada indicador” (Álvarez, 2013).



Figura 8

Tabla de indicadores y rendimiento

Proceso: preparación y distribución de pedidos	
Indicadores	Rendimientos
Porcentaje de pedidos servidos en menos de 48 horas	103%
N.º de reclamaciones por rotura o desperfecto de la mercancía	83%
Coste medio de entrega por pedido	108%

Fuente: Recuperado de Álvarez (2013).

Dice el mismo autor que si consideramos la misma importancia para los tres indicadores, el valor del rendimiento del proceso “preparación y distribución de pedidos” sería:

$$\text{Rto. Proceso} = (103 + 83 + 108) / 3 = 98\%$$

Por otro lado, agrega Álvarez (2013) que si queremos realizar la suma ponderada es necesario priorizar por importancia los indicadores del proceso.

Figura 9

Tabla de indicadores, rendimientos e importancia

Proceso: preparación y distribución de pedidos		
Indicadores	Rendimientos	Importancia
Porcentaje de pedidos servidos en menos de 48 horas	103%	0,5%
N.º de reclamaciones por rotura o desperfecto de la mercancía	83%	0,2%
Coste medio de entrega por pedido	108%	0,3%

Fuente: Recuperado de Álvarez (2013).

Con estos valores, el rendimiento del proceso resultaría:



$$\text{Rto. Proceso} = (103 \times 0,5) + (83 \times 0,2) + (108 \times 0,3) = 101\%$$

También menciona Álvarez (2013) que la asignación de importancias a los indicadores puede realizarse de manera subjetiva o empleando la matriz pareada. Esta matriz es una herramienta utilizada para priorizar y asignar pesos de forma más objetiva.

Figura 10

Ejemplo genérico de matriz pareada

	Ítem 1	Ítem 2	...	Ítem N	Total	Importancia
Ítem 1						
Ítem 2						
...						
Ítem N						

Fuente: Recuperado de Álvarez (2013).

Según Álvarez (2013) la matriz pareada es una tabla de doble entrada en la que los indicadores se sitúan en filas y columnas. En cada una de las celdas enfrentamos una sola vez a cada indicador con el resto (por eso la mitad de la tabla se anula), y el resultado del enfrentamiento se puntúa con los valores 0, 1, 2. A continuación el mismo autor describe como elegir la puntuación:

- Valor (0): Consideramos que los indicadores comparados tienen la misma importancia en el proceso.
- Valor (1): Este es el valor si el indicador elegido se considera un poco más importante que su rival.
- Valor (2): Estamos en esta situación si el indicador elegido es mucho más importante que su oponente.



Continúa explicando Álvarez (2013) que “en cada celda se coloca la puntuación resultante del “enfrentamiento” seguida del nombre del ítem considerado como ganador, excepto para el valor cero (empate técnico)”.

Luego se suman las puntuaciones de cada elemento en los distintos enfrentamientos y se colocan en la columna total. El total indica ya una priorización de cada elemento en relación con el resto dice Álvarez (2013).

Figura 11

Priorización de indicadores del proceso "Preparación y distribución de pedidos"

	IND. 1: % de pedidos servidos en <48 horas	IND. 2: N.º reclamaciones por rotura o desperfecto	IND. 3: Coste medio de entrega por pedido	Total	Importancia
IND. 1: % de pedidos servidos en <48 horas		2 (IND. 1)	1 (IND. 1)	3 (+1) = 4	4 / 7 = 0,57
IND. 2: N.º reclamaciones por rotura o defecto			1 (IND. 3)	0 (+1) = 1	1 / 7 = 0,14
IND. 3: Coste medio de entrega por pedido				1 (+1) = 2	2 / 7 = 0,29
				7	7 / 7 = 1

Fuente: Recuperado de Álvarez (2013).

El mismo autor finaliza expresando que:

Si al sumar los puntos alcanzados por cada elemento de comparación, alguno de los elementos resultase cero, sumaremos 1 a todos los totales. Esto se realiza para poder calcular una cifra de importancia para todos los elementos (si no lo hiciéramos así, el elemento que tuviera valor cero tendría cero también en importancia). Para calcular la importancia dividiremos el valor obtenido por cada elemento entre el total común. Con estos valores de importancia, el rendimiento del proceso en nuestro ejemplo resultaría:

$$\text{Rto. Proceso} = (103 \times 0,57) + (83 \times 0,14) + (108 \times 0,29) = 102\%$$



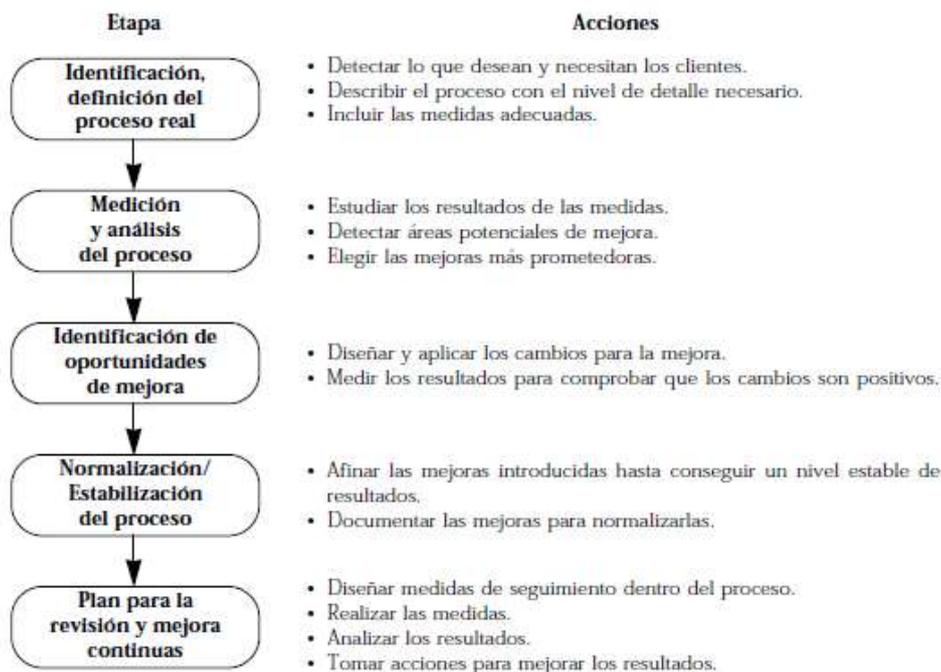
4.1.2.6 La mejora de los procesos.

Entre las técnicas para gestionar y mejorar los procesos se encuentran el modelo sistemático de mejora y la reingeniería aplicada a procesos concretos, según menciona Zaratiegui (1999).

El método sistemático de mejora de procesos consiste en etapas, donde cada una tiene un propósito y acciones específicas a realizar. Básicamente, implica identificar los procesos, medirlos a través de indicadores y proponer acciones para mejorarlos.

Figura 12

Método sistemático de mejora de procesos



Fuente: Recuperado de Zaratiegui (1999).

Dice Medina et al. (2019) que “la mejora de procesos centra su objetivo en lograr el enfoque al cliente, el alineamiento estratégico y la mejora continua”.

Amplia Zaratiegui (1999) cuando dice que “un principio fundamental de la calidad total es el papel de los clientes como árbitros absolutos de ésta. La calidad total busca la satisfacción de los clientes y, mediante ella, su fidelidad a la empresa proveedora”.



El mismo autor agrega que el diseño de los productos y servicios, su realización, el servicio de ventas y el servicio posventa, todos esos procesos deben ser gestionados con un enfoque en el cliente.

La reingeniería de procesos se puede definir como el rediseño radical de los procesos de la empresa para conseguir mejoras espectaculares en medidas críticas actuales, como los costos, la calidad, el servicio y la rapidez (Zaratiegui, 1999). Este método implica cambios profundos de mentalidad, por lo que, en la práctica, se modifican bloques enteros de procesos relacionados, todos los procesos básicos de una unidad de negocio o incluso todos los procesos clave de la organización. Según Zaratiegui (1999), cuando la reingeniería tiene éxito, “produce un salto cualitativo equivalente a varios años de progreso continuo y, a veces, va más allá de lo que se conseguiría con los métodos de mejora progresiva”.

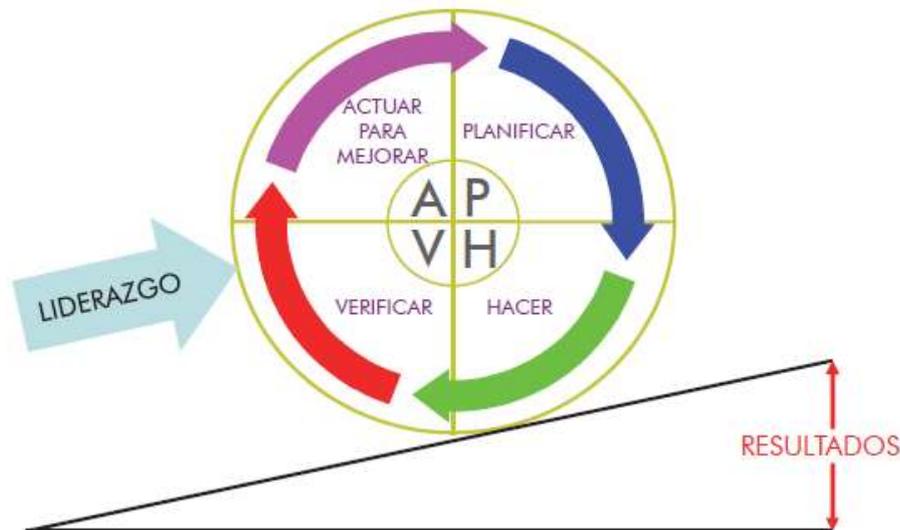
4.2 Ciclo de mejora continua

Cuando se habla de gestión de procesos es necesario incorporar el ciclo de mejora continua para lograr una excelencia.

Definimos la gestión de procesos como actividades articuladas para dirigir y coordinar una organización. Un mecanismo por el cual podemos gestionar de una manera efectiva es el ciclo PHVA que responde a Planificar – Hacer – Verificar – Actuar (Pardo Álvarez, 2017).

Figura 13

Ciclo de mejora continua



Fuente: Recuperado de Pardo Álvarez (2017).

Si queremos obtener buenos resultados, Pardo Álvarez (2017) sugiere planificar las actividades, realizar lo planificado, verificar si se ha conseguido lo deseado y actuar para corregir las desviaciones. “Estamos muy focalizados en la fase ‘Hacer’ del ciclo PDCA, y no reparamos en que las otras fases son tan importantes como esta o más”, menciona el mismo autor.

Pardo Álvarez (2017) agrega que “gestionar los procesos es aplicar el ciclo de mejora continua”. Para ello, el autor explica detalladamente cada paso del ciclo de mejora continua:

Planificar los procesos: esta planificación debe ser meditada. Tiene dos enfoques. El primero es global a través de un mapa de procesos y el segundo es individual donde generalmente se utiliza una herramienta de planificación (procedimientos, flujogramas, fichas de proceso).

Llevar a cabo los procesos: en esta fase se ejecuta el proceso siguiendo lo planificado para después poder aprobar o modificar la planificación.

Verificar los procesos: existen distintos mecanismos de verificación como indicadores, controles o auditorías con el fin de contrastar los resultados parciales o finales obtenidos. Si los resultados son positivos, podremos decir que nuestros



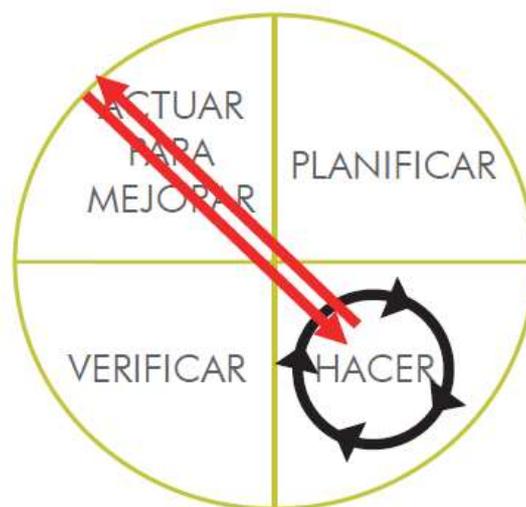
procesos están bajo control. Si surgen desviaciones deberán ser tratadas en la etapa del ciclo de mejora continua. Esta etapa es importante porque nos abre las puertas a la mejora de los procesos.

Actuar para mejorar los procesos: las desviaciones encontradas permiten poner en marcha acciones para mejorar el proceso. Esta fase observaremos si los métodos de trabajos actuales son adecuados y a su vez detectar iniciativas para mejorar el desempeño de este.

Además, Pardo Álvarez (2017) menciona que esta manera de gestionar los procesos es sencilla y práctica, pero no es muy habitual ponerla en práctica. A menudo se opta por la “diagonal perversa”, comenzando por la fase de hacer como prioritario y actuando para mejorar al tratar de solucionar las incidencias detectadas, sin pasar por las fases de planificar y verificar, según agrega el mismo autor.

Figura 14

Gestión siguiendo la diagonal perversa



Fuente: Recuperado de Pardo Álvarez (2017).

Dice Pardo Álvarez (2017) que las implicaciones negativas pueden ser, por ejemplo, deficiente calidad de los productos y servicios, baja satisfacción de los clientes, deterioro de la imagen y sobrecostos por reprocesados, devoluciones, penalizaciones, etc. La organización debe utilizar el ciclo completo en lo posible.



4.3 Series de tiempo

Una serie de tiempo se refiere a “datos estadísticos que se recopilan, observan y registran en intervalos regulares. Estos intervalos pueden ser días, semanas, meses, años u otra unidad de tiempo. Para analizar estos datos, es útil elaborar una gráfica de serie temporal, la cual es una representación visual del tiempo y las variables” (EBC Academia, 2017).

Según Kasmier y Díaz Mata (1996), una serie de tiempo “es un conjunto de valores observados, tales como datos de producción o ventas, organizados secuencialmente en periodos de tiempo. Algunos ejemplos de estos datos son las ventas de un producto determinado durante varios meses”.

El objetivo del análisis de series temporales es explicar las variaciones observadas en la serie en el pasado, tratando de determinar si responden a un determinado patrón de comportamiento. Si se define dicho patrón, se intentará predecir el comportamiento futuro de la serie (De la Fuente Fernández, 2023).

Así, la serie tiene una variable dependiente y una variable independiente o explicativa, que es el tiempo. “Es importante aclarar que el tiempo, en sí, no es una variable explicativa, sino simplemente un 'soporte' o escenario en el que se realiza o tiene lugar una serie temporal” (De la Fuente Fernández, 2023). Esta forma de abordar el estudio de una serie temporal se conoce como enfoque clásico.

4.3.1 Tratamiento de las series de tiempo

Dice UNAM (2022a) que las series de tiempo se analizan de manera visual, gráfica y con la observación de los parámetros estadísticos más importantes, aunque finalmente se requiere experiencia e interpretación para evitar decisiones equivocadas. Por estas razones, “se hace necesario y se propone algún tipo de tratamiento que ayude a extraer resultados y conclusiones útiles que complementen el análisis actual de la información, especialmente cuando se dispone de una gran cantidad de datos” (UNAM, 2022a).

Existen distintos tipos de análisis que han surgido a lo largo del tiempo. Dos de ellos destacan por su eficacia y su amplia divulgación: la modelación clásica y la modelación ARIMA. Cada uno de ellos “trabaja sobre enfoques distintos: el enfoque determinista y el enfoque estocástico, respectivamente” (UNAM, 2022a).



4.3.2 El modelo ARIMA

La metodología Box-Jenkins se utiliza como modelo para el análisis de series de tiempo de media móvil autorregresivo (ARIMA). Este método es apropiado “para series de tiempo de longitud media a larga, con al menos 50 observaciones. Dado que la metodología está destinada solo para la aplicación de modelos estacionarios, es necesario hacer algunos ajustes antes de modelar estas series no estacionarias” (Miranda Chinlli, 2021).

De acuerdo con De la Fuente Fernández (2022), Box y Jenkins consideran la dependencia existente entre los datos, ya que cada observación en un momento dado se modela en función de los valores anteriores.

Los modelos se conocen con el nombre genérico de ARIMA (AutoRegresive Integrated Moving Average), que deriva de sus tres componentes: AR (Autoregresivo), I (Integrado) y MA (Medias Móviles).

El modelo ARIMA “permite describir un valor como una función lineal de datos anteriores y errores debidos al azar. Además, puede incluir un componente cíclico o estacional. Box y Jenkins recomiendan como mínimo 50 observaciones en la serie temporal” (De la Fuente Fernández, 2022).

La modelación ARIMA toma la serie como un proceso estocástico. Dice UNAM (2022b) que “bajo este enfoque se concibe a la serie de tiempo como un conjunto de valores de tipo aleatorio generados a partir de un proceso totalmente desconocido, es decir, se concibe a la serie como un proceso estocástico”.

4.3.2.1 Modelos AR, MA y mixtos.

La principal característica de dichos modelos es que “no involucran a las variables independientes en su construcción, es decir a la variable del tiempo t , como se hace en el análisis clásico. En cambio, emplean la información que se encuentra en la serie misma para generar los pronósticos” (UNAM, 2022b).

Entonces, UNAM (2022b) explica como ARIMA está compuesto por varios modelos:

Modelo autorregresivo de orden p - AR (p): un modelo en el que una determinada observación es predecible a partir de la observación inmediatamente



anterior (modelo autorregresivo de primer orden) o a partir de las dos observaciones que le preceden (modelo autorregresivo de segundo orden). En este caso, la observación actual se define como la suma ponderada de una cantidad finita p de observaciones precedentes más un impulso aleatorio independiente.

Matemáticamente, un modelo autorregresivo tiene la forma siguiente:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi_3 Y_{t-3} + \dots + \varphi_p Y_{t-p} + a_t$$

En donde Y_t es la variable dependiente y $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ son las variables independientes.

En este caso, las variables independientes son valores de la misma variable (de aquí el nombre de auto), pero de periodos anteriores ($t-1, t-2, t-3, \dots, t-p$). Por último, a_t es el error, o término residual, que representa perturbaciones aleatorias que no pueden ser explicadas por el modelo.

Modelo de Medias Móviles de orden q - MA(q): una determinada observación está condicionada por los “impulsos aleatorios” de las observaciones anteriores. De esta forma la observación actual se define como la suma del impulso actual y de los impulsos aleatorios anteriores con un determinado peso.

Matemáticamente, un modelo de medias móviles tiene esta fórmula:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Donde a_t es el residuo o error en el periodo t y $a_{t-1}, a_{t-2}, a_{t-3}, \dots, a_{t-q}$ son los valores anteriores del error. La ecuación es semejante a la anterior, con la excepción de que implica que la variable dependiente Y_t depende de valores previos del término residual más que de la variable misma.

Modelo autorregresivo de medias móviles de orden p, q - ARMA(p, q): es la combinación de las estructuras anteriores; modelo autorregresivo y modelo de medias móviles. Así, una observación está determinada tanto por observaciones



anteriores así como por “impulsos aleatorios” o también llamados “errores” de observaciones pasadas.

Matemáticamente, un modelo autorregresivo de medias móviles tiene esta fórmula:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi_3 Y_{t-3} + \dots + \varphi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Modelo autorregresivo integrado de medias móviles de orden p,d,q - ARIMA(p,d,q): al igual que un modelo ARMA, es la combinación de los modelos autorregresivo y el de medias móviles, con la particularidad de incluir un proceso de restablecimiento (el cual se denomina integración) de inestabilidad original presente en una serie de tiempo.

$$Y'_t = \varphi_1 Y'_{t-1} + \varphi_2 Y'_{t-2} + \varphi_3 Y'_{t-3} + \dots + \varphi_p Y'_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Donde:

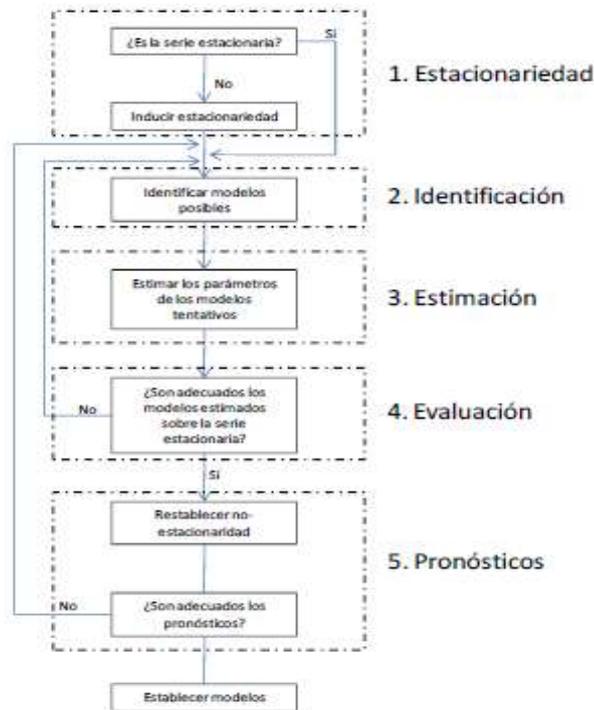
Y'_t: Es la serie inducida a la estabilidad

4.3.3 Metodología Box-Jenkins No estacional.

En la siguiente figura se observan las etapas de la metodología Box-Jenkins:

Figura 15

Metodología BOX – Jenkins



Fuente: Recuperado de UNAM (2022b).

4.3.3.1 Estacionariedad.

Los procesos estocásticos se clasifican en estacionarios y no estacionarios. Existen dos formas de conocer si una serie es estacionaria:

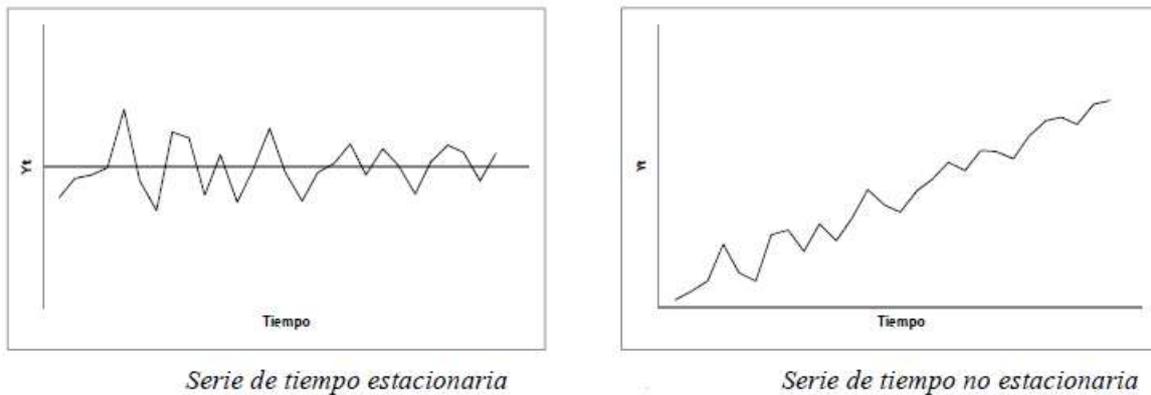
- por medio del gráfico de la serie.
- mediante la exploración de función de autocorrelación simple aclara el mismo autor.

En primer lugar, por medio de la visualización gráfica “se puede advertir si la serie es estacionaria si se detectan elevaciones o inclinaciones en los datos. Cualquier patrón de este tipo expresa que la serie es inestable” (UNAM, 2022b).



Figura 16

Serie de tiempo estacionaria y no estacionaria



Fuente: Recuperado de UNAM (2022b).

En segundo lugar, si visualmente no es factible determinar si la serie es estacionaria se recurre a la función de autocorrelación simple (FAS). La función de autocorrelación simple se puede cortar o truncar. Para explicar este término dice UNAM (2022b) que “existe una espiga en el desfaseamiento k en la FAS si r_k es estadísticamente grande. Se considera que existe una espiga en la función si el valor absoluto de”:

$$t_{r_k} = \frac{r_k}{s_{r_k}}$$

es mayor que 2. La función de autocorrelación se trunca después del desfaseamiento k si no hay espigas en los desfaseamientos mayores que k en la función. Por ejemplo, si:

$$|t_{r_1}| = 3.671$$

$$|t_{r_2}| = 2.873$$

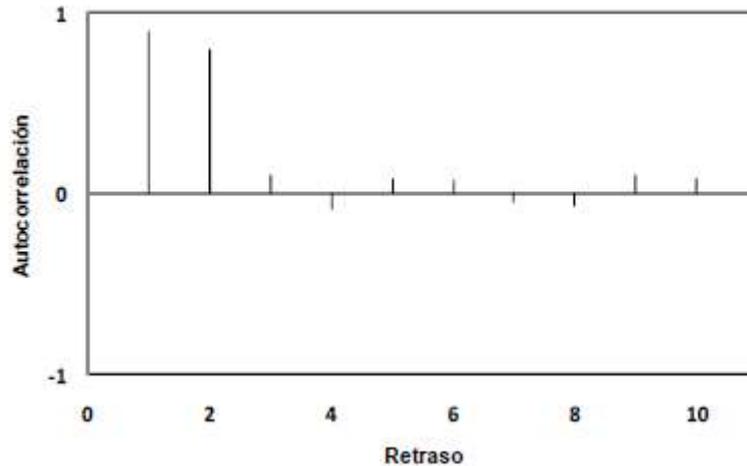
$$|t_{r_3}| = 0.517$$

Entonces la función de autocorrelación simple se trunca o se corta después del desfaseamiento 2. Esto se ilustra en la siguiente figura:



Figura 17

Función de autocorrelación se trunca



Fuente: Recuperado de UNAM (2022b).

Agrega el mismo autor que “la función de autocorrelación puede extinguirse en forma exponencial, de seno amortiguada o una combinación de ambas. Además, puede cortarse con gran rapidez o puede extinguirse con lentitud extrema”.

Por otro lado, dice UNAM (2022b) que la realidad es que “las series de tiempo normalmente no se comportan establemente, es decir, no son procesos estocásticos estacionarios. Cuando se presenta esa situación entonces es preciso aplicar una transformación a los valores de la serie para tener una serie temporal estacionaria”.

Existen distintas maneras de inducir la estacionariedad en una serie de tiempo, la más usada es el método llamado de construcción de diferencias.

Comenta UNAM (2022b) que el método consiste en obtener diferencias entre los mismos valores de la serie, para remover cualquier patrón de tendencia.

Agrega el mismo autor:

Las primeras diferencias de los valores de la serie de tiempo Y_1, Y_2, \dots, Y_n son”:

$$Y'_t = Y_t - Y_{t-1}$$

$$t = 2, 3, \dots, n$$



La siguiente figura muestra cómo se hace la diferenciación y cómo los datos con una tendencia lineal se convierten en estacionarios (horizontales) después de diferenciar.

Figura 18

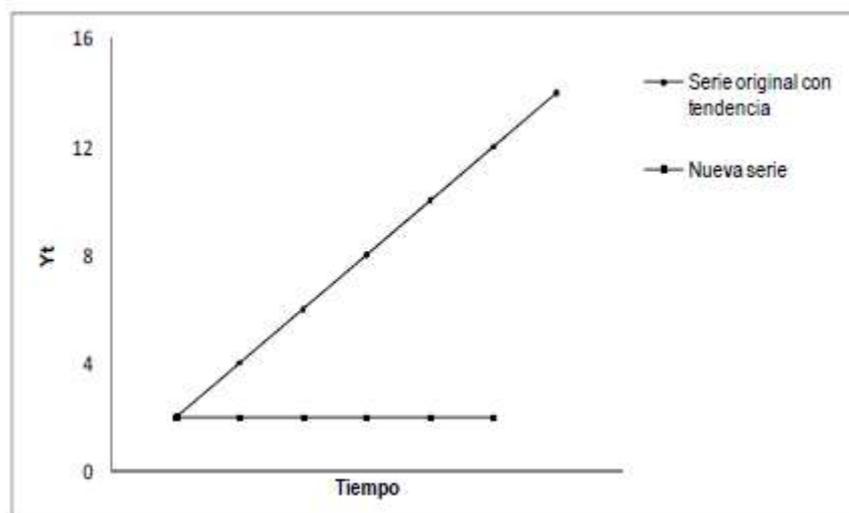
Obtención de las primeras diferencias en una serie

Serie de datos	Primeras diferencias	Nueva serie
2	$4-2 = 2$	2
4	$6-4 = 2$	2
6	$8-6 = 2$	2
8	$10-8 = 2$	2
10	$12-10 = 2$	2
12	$14-12 = 2$	2
14		-

Fuente: Recuperado de UNAM (2022b).

Figura 19

Transformación de la serie de tiempo mediante las primeras diferencias



Fuente: Recuperado de UNAM (2022b).



La diferencia entre un modelo ARMA y un modelo ARIMA “se halla en que un modelo ARMA (p,q) es capaz de operar únicamente sobre series estacionarias, mientras que un modelo ARIMA (p,d,q) es capaz de operar tanto sobre series de tiempo no estacionarias como en series estacionarias” (UNAM, 2022b).

Existe un test que se utiliza para determinar si una serie es estacionaria llamado test de Dickey-Fuller. Para llegar a la definición Parra (2015) comienza explicando la prueba de raíz unitaria:

La ecuación de un modelo autorregresivo de primer orden:

$$Y_t = \phi Y_{t-1} + u_t, \quad -1 < \phi < 1$$

Donde u_t es un término de error de ruido blanco.

Si $\phi = 1$, es una serie no estacionaria en varianza. La no estacionariedad de este tipo de modelos también se denomina problema de raíz unitaria. Sin embargo, si $\phi < 1$, se demuestra que la serie de tiempo Y_t es estacionaria. En la práctica la manera de averiguar si una serie de tiempo tiene una raíz unitaria es a través de un test en el que la hipótesis que se contrasta es el valor unitario del coeficiente autorregresivo (ϕ) a través del análisis de la nulidad de $(\phi - 1)$. En definitiva se trata de transformar restando Y_{t-1} :

$$Y_t - Y_{t-1} = \phi Y_{t-1} - Y_{t-1} + u_t = (1 - \phi) Y_{t-1} + u_t$$

Que también se expresa como:

$$\Delta Y_t = \delta Y_{t-1} + u_t$$

Por tanto, obtenemos y probamos “la hipótesis (nula) de que $\delta = 0$, y la hipótesis alternativa es que $\delta < 0$. Si $\delta = 0$, entonces $\phi = 1$; tenemos una raíz unitaria, lo cual significa que la serie de tiempo es no estacionaria” (Parra, 2015).

En este contraste de hipótesis dice el mismo autor que “el estadístico $T = \left| \delta / \sigma_\delta \right|$ no sigue la distribución t ni siquiera en muestras grandes. Dickey y Fuller calcularon los



valores críticos del estadístico τ con base en simulaciones Monte Carlo”. El estadístico tau se conoce como prueba Dickey-Fuller.

La hipótesis nula, H_0 : no estacionariedad, se acepta si τ toma un valor situado a la derecha del valor crítico correspondiente al nivel de significación establecido. Los programas proporcionan los valores críticos, que son negativos, para niveles de significación del 1%, 5% y 10%, siendo el 5% el más utilizado en la práctica. Luego se rechaza si toma un valor menor que el valor crítico (Parra, 2015).

Por otro lado, la siguiente tabla indica la distribución del estadístico cuando incluimos o no la constante en el modelo. La decisión del contraste es rechazar no estacionariedad si $\tau \leq T_c$, donde el valor de T_c se obtiene de la tabla (Peña, 2010)

Figura 20

Transformación de la serie de tiempo mediante las primeras diferencias

T	Sin constante				Con constante			
	.01	.025	.05	.1	.01	.025	.05	.1
25	-2.66	-2.26	-1.95	-1.60	-3.75	-3.33	-3.00	-2.63
50	-2.62	-2.25	-1.95	-1.61	-3.58	-3.22	-2.93	-2.60
100	-2.60	-2.24	-1.95	-1.61	-3.51	-3.17	-2.89	-2.58
250	-2.58	-2.23	-1.95	-1.62	-3.46	-3.14	-2.88	-2.57
500	-2.58	-2.23	-1.95	-1.62	-3.44	-3.13	-2.87	-2.57
∞	-2.58	-2.23	-1.95	-1.62	-3.43	-3.12	-2.86	-2.57

Fuente: Recuperado de Peña (2010).

En la práctica dice Zarzosa Espina (2013) que a través de programas específicos como puede ser Eviews se puede calcular el test de Dickey y Fuller el cual arroja estos principales valores:

- El estadístico de Dickey y Fuller.
- El p valor.
- Los valores críticos al nivel de 1 %, 5 % y 10 %.

Usualmente se toma el valor crítico del 5 % para evaluar la estacionalidad de la serie.

Hay dos formas de proceder:

- Se evalúa el 5 % (0,05) con el p valor.
- Se evalúa el valor crítico del test al nivel del 5 % con el valor del estadístico de Dickey-Fuller.



Continúa la explicación el mismo autor a través del siguiente ejemplo:

Figura 21

Datos de Eviews para análisis de serie

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.660839	0.2591
Test critical values:		
1% level	-4.356068	
5% level	-3.595026	
10% level	-3.233456	

Fuente: Recuperado de Zarzosa Espina (2013).

Como datos tenemos:

- El estadístico de Dickey y Fuller en -2,66.
- El p valor en 0,25.
- El valor crítico al nivel de 5 % de -3,59.

Entonces se dice que la serie no es estacionaria dado que:

- El valor crítico al nivel de 5 % de -3,59 es menor que el estadístico de Dickey y Fuller en -2,66.
- También podemos interpretar el p valor 0,25, que lógicamente conduce a la misma conclusión, ya que es mayor que el nivel de significación prefijado (0,05).

Figura 22

Evaluación de estacionalidad



Fuente: Recuperado de Zarzosa Espina (2013).



4.3.3.2 Identificación.

Una vez que la serie es estable en el tiempo, el siguiente paso en la metodología Box-Jenkins es la identificación del modelo probable. Las ideas básicas de esta fase según UNAM (2022b) son las siguientes:

La serie de tiempo que se encuentre en proceso de estudio cuenta con sus respectivas funciones de autocorrelación simple y parcial (FAS y FAP) que se denominan prácticas o calculadas.

Por otra parte, cada una de las distintas configuraciones ARMA posee su FAS y FAP teóricas asociadas al modelo.

Si la FAS y FAP calculadas de la serie a la que deseamos ajustar un modelo se asemeja a alguna o varias FAS y FAP teóricas, entonces podemos decir que el modelo ARMA teórico es un modelo tentativo para la serie.

Entonces, la identificación del modelo probable se realiza mediante la comparación de las funciones de autocorrelación calculadas contra las teóricas, tanto simples como parciales (UNAM, 2022b).

Las funciones de autocorrelación calculadas solo se utilizan como guías para seleccionar uno o varios modelos ARMA que parezcan apropiados, ya que brindan una aproximación a la estructura más adecuada que debe considerarse, según el mismo autor.

Los profesores Box y Jenkins sugirieron una completa familia de modelos teóricos de los cuales se puede elegir: AR (1), AR (2), MA (1), MA (2), ARMA (1,1). Las principales características que se observan de las funciones de autocorrelación teóricas son:

Figura 23

Características FAS y FAP teóricas

Modelo	FAS	FAP
AR	Decae a cero	Se trunca o se corta (después del retraso p)
MA	Se trunca o se corta (después del retraso q)	Decae a cero
ARMA	Decae a cero	Decae a cero

Fuente: Recuperado de UNAM_2 (2022).



4.3.3.3 Estimación.

La estimación de los parámetros del modelo ARMA seleccionado se realiza por medio de minimizar la suma de los cuadrados de los residuales SSR (sum of squared residuals) dice (UNAM, 2022b). Según el mismo autor los residuales de cualquier modelo ARMA se define como:

$$\hat{a}_t = Y_t - \hat{Y}_t$$

Donde:

\hat{a}_t : residual de la serie

Y_t : serie original

\hat{Y}_t : serie calculada con los parámetros estimados

Entonces, la suma de los cuadrados de los residuales se define como:

$$SSR = \sum \hat{a}_t^2$$

Esta última función es la que se debe minimizar, a través de la búsqueda de los parámetros Φ y ϕ del modelo ARMA que se desea ajustar (UNAM, 2022b). Además, comenta el mismo autor que los parámetros Φ y ϕ del modelo ARMA propuesto, deben de cubrir en conjunto los siguientes requerimientos:

1. Se opta siempre por el modelo con el menor número de coeficientes.
2. La metodología Box-Jenkins requiere que el modelo que se utiliza para la descripción y pronóstico de una serie de tiempo sea tanto estacionario como inversible.
3. Los coeficientes del modelo deben ser estadísticamente significativos. Se debe determinar si los coeficientes son importantes en el modelo o se deben excluir del mismo. Se hace uso de los valores del error estándar y el valor t asociado a cada uno de los coeficientes del modelo. El valor de t se define como:



$$t = \frac{\text{coeficiente calculado}}{\text{error estándar asociado al coeficiente calculado}}$$

Ambos valores normalmente son calculados automáticamente por los programas estadísticos.

4. El modelo debe proporcionar un adecuado ajuste. Una medida útil que ayuda a conocer el grado de ajuste del modelo a la serie es la RMSE (root-meansquared error). Esta medida da a conocer la desviación estándar de los residuales \hat{a}_t , y se calcula como:

$$RMSE = \hat{\sigma}_a^2 = \frac{1}{n - m} \sum \hat{a}_t^2$$

Donde:

\hat{a}_t : residuales del modelo

n : número de residuales

m : número de coeficientes del modelo

Se utiliza el valor de RMSE para compararlo con el de otros modelos estimados para la misma serie. Se preferirá aquel modelo cuyo RMSE tienda a tener menor valor.

4.3.3.4 Evaluación del modelo.

Para evaluar el modelo comenta UNAM (2022b) que “en este paso se comprueba la eficiencia del modelo y se decide si es estadísticamente adecuado. Un modelo estadísticamente adecuado es aquel cuyos residuales son independientes entre sí. Es decir, si los residuales son completamente aleatorios”.

La función de autocorrelación simple de los residuales es el instrumento que se utiliza para determinar si el modelo es estadísticamente adecuado:



$$r_k(\hat{a}) = \frac{\sum_{t=1}^{n-k} (\hat{a}_t - \bar{a})(\hat{a}_{t+k} - \bar{a})}{\sum_{t=1}^n (\hat{a}_t - \bar{a})^2}$$

Donde:

$r_k(\hat{a})$: coeficiente de autocorrelación simple residual para un retraso de k periodos

\bar{a} : media de los residuales

\hat{a}_t : residual en el periodo t

\hat{a}_{t+k} : residual en el periodo con k retrasos

n : número total de residuales

Mediante el análisis de esta función se busca que los coeficientes de autocorrelación sean cero o muy cercanos a él. Específicamente se pretende que los coeficientes no sean significativos. La mayoría de los programas de computadora realiza automáticamente el cálculo.

Se puede detectar presencia de autocorrelación a través de algún contraste de hipótesis. “Éstos permiten, a través de una regla de decisión, considerar si con los datos de la muestra y con un nivel de significación concreto se debe o no rechazar la hipótesis nula” (Arranz & Zamora, 2002).

La forma general del contraste es:

- H_0 : No existe autocorrelación.
- H_1 : Existe autocorrelación.

Agrega el mismo autor que:

Esto es, en la hipótesis nula se considera que el término de perturbación correspondiente a una observación es independiente del correspondiente a cualquier otra observación. En la hipótesis alternativa se señala que el término de error de un modelo econométrico está autocorrelacionado a través del tiempo. Esta hipótesis alternativa, al considerar la existencia de un patrón de comportamiento para los residuos, se puede especificar con procesos autorregresivos —AR (p)—, de medias móviles —MA (q)— o mixtos —ARMA (p,q)— dependiendo del contraste que se vaya a utilizar.



El contraste desarrollado por Durbin y Watson es la prueba más frecuentemente empleada para detectar la presencia de autocorrelación en los modelos de regresión. “Este contraste permite verificar la hipótesis de no autocorrelación frente a la alternativa de autocorrelación de primer orden bajo un esquema autorregresivo — AR (1): $u_t = \rho * u_{t-1} + \epsilon_t$ ” (Arranz y Zamora, 2002).

Agrega el mismo autor que analíticamente el contraste se especifica del siguiente modo:

Formulación de las hipótesis:

$H_0: \rho = 0$; No existe autocorrelación AR (1)

$H_1: 0 < |\rho| < 1$; Existe autocorrelación AR (1)

Estadístico de prueba:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

Profundiza Arranz y Zamora (2002) comentando que:

Teniendo en cuenta los límites de variación del coeficiente de correlación empírico, $-1 < \rho < 1$, se puede deducir el rango de variación del estadístico de Durbin-Watson y el signo de la autocorrelación,

$\rho = -1$; $d \approx 4$ se considera que existe autocorrelación negativa

$\rho = 0$; $d \approx 2$ indica ausencia de autocorrelación

$\rho = 1$; $d \approx 0$ se puede admitir que existe autocorrelación positiva”

No obstante, estos valores (0, 2 y 4) son límites extremos que deben matizarse estableciendo regiones más amplias en las que pueda considerarse si existe o no autocorrelación y, en caso de detectarse, si ésta es positiva o negativa.

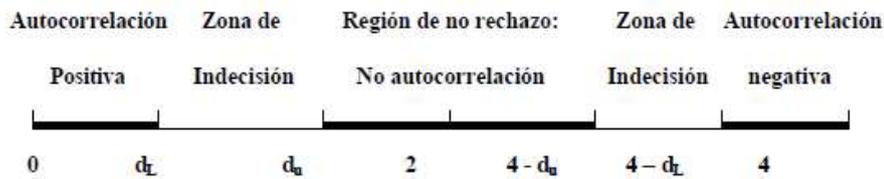
Durbin y Watson hallaron un límite superior (d_U) e inferior (d_L) que permiten tomar decisiones acerca de la presencia o ausencia de autocorrelación.

Entonces, gráficamente se pueden señalar las regiones del contraste en el siguiente segmento:”



Figura 24

Categorización de niveles de contraste



Esto es,

- $0 < d < d_L$ se rechaza H_0 , existe entonces autocorrelación positiva con un esquema AR(1)
- $4 - d_L < d < 4$ se rechaza H_0 , existe autocorrelación negativa con un esquema AR(1)
- $d_u < d < 4 - d_u$ no se rechaza H_0 , no existe autocorrelación
- $d_L < d < d_u$ el contraste no es concluyente
- $4 - d_u < d < 4 - d_L$ el contraste no es concluyente

Fuente: Recuperado de Arranz y Zamora (2002).

4.3.3.5 Pronóstico.

La fase final de la metodología Box-Jenkins es pronosticar valores futuros de la serie de tiempo. Se debe comprender cómo realizar predicciones puntuales a partir de un modelo ARIMA ya estimado.

El pronóstico de una serie de tiempo “se hace por medio de predicciones o estimaciones puntuales. Así, predicción puntual (\hat{y}_t) se define como el valor de la variable y en el tiempo t, calculado por medio del modelo ARIMA ajustado a la serie” cita (UNAM, 2022b).

4.3.3.6 Ventajas y desventajas de la modelación ARIMA.

Dice UNAM (2022b) que la modelación ARIMA presenta importantes ventajas frente a la modelación clásica:

- El gran grado de ajuste que proporciona a la mayoría de las series de tiempo mientras que la modelación clásica se ajusta una serie a un modelo matemático ya establecido, los modelos ARIMA se ajustan a una serie en particular.



- Los conceptos que se utilizan para la modelación ARIMA se derivan de sólidas teorías de la probabilidad clásica y de la estadística matemática.
- Los modelos ARIMA son una familia de modelos, no simplemente un único modelo.
- Box y Jenkins desarrollaron una completa estrategia que sirve de guía para escoger un apropiado modelo dentro de esta gama de modelos.
- Un apropiado modelo ARIMA produce óptimas predicciones.

Dentro de las dificultades que se pueden presentar en la modelación ARIMA, según el mismo autor son:

- El problema que supone para el analista, recién iniciado en la teoría y la ejecución de la metodología Box-Jenkins, la correcta elección de un modelo adecuado en los primeros intentos. Es por ello que la modelación de un proceso ARIMA ha sido referida por algunos autores como un arte.
- Aunque se utilizan programas de cómputo en el cálculo de los coeficientes de un modelo ARIMA probable, la aplicación de la metodología Box-Jenkins sigue siendo una labor manual, es decir, requiere ser ejecutada por el analista.

4.3.4 Modelos puramente estacionales: SARIMA.

Como hemos visto ARIMA se ocupa en modelar las series de tiempo enfocado en la estacionariedad de esta. Los modelos SARIMA “captan el comportamiento puramente estacional de una serie, en forma similar, como hemos visto, se realiza para la componente regular o no estacional.

Una serie con influencia solamente por la parte estacional puede ser descrito por un modelo SARIMA (P,D,Q)” (Quisbert, s.f.).

4.3.5 Modelo multiplicativo.

Explica Quisbert (s.f.) que en la práctica generalmente se presentan series afectadas por ambas componentes, tendencia regular y estacionalidad.

Box y Jenkins proponen un modelo denominado multiplicativo el cual puede explicar el comportamiento de una serie afectada por ambas componentes. Este modelo “es usado cuando la serie temporal tiene estacionalidad. La notación es similar a la del ARIMA” (Moreno, 2019):



Modelo multiplicativo= ARIMA (p,d,q) * SARIMA (P,D,Q)m

Fuente: adaptado de (Moreno, 2019)

Explica el mismo autor que “los parámetros p,d,q corresponden a AR, a la parte integrada y a MA. Los parámetros P,D,Q simplemente representan a (p,d,q) pero para la parte estacional de la serie temporal. El parámetro m indica el número de períodos en cada estacionalidad”.

4.4 Machine Learning

Dice Bagnato (2020) que el machine learning busca resolver cómo construir programas de computadora que mejoran automáticamente al adquirir experiencia. El mismo autor agrega que "el aprendizaje automático, relacionado con el campo estadístico, consiste en extraer y reconocer patrones y tendencias para comprender qué nos dicen los datos. Para ello, se vale de algoritmos que pueden procesar gigabytes y/o terabytes y obtener información útil".

También, Pedrero et al. (2021) explican que el machine learning (ML) tiene como propósito “el desarrollo de algoritmos para extraer información de los datos, ya sea para explicación, clasificación o predicción”. En cambio, la inteligencia artificial es una categoría aún más amplia que incluye técnicas para el análisis de datos estructurados y no estructurados, según el mismo autor.

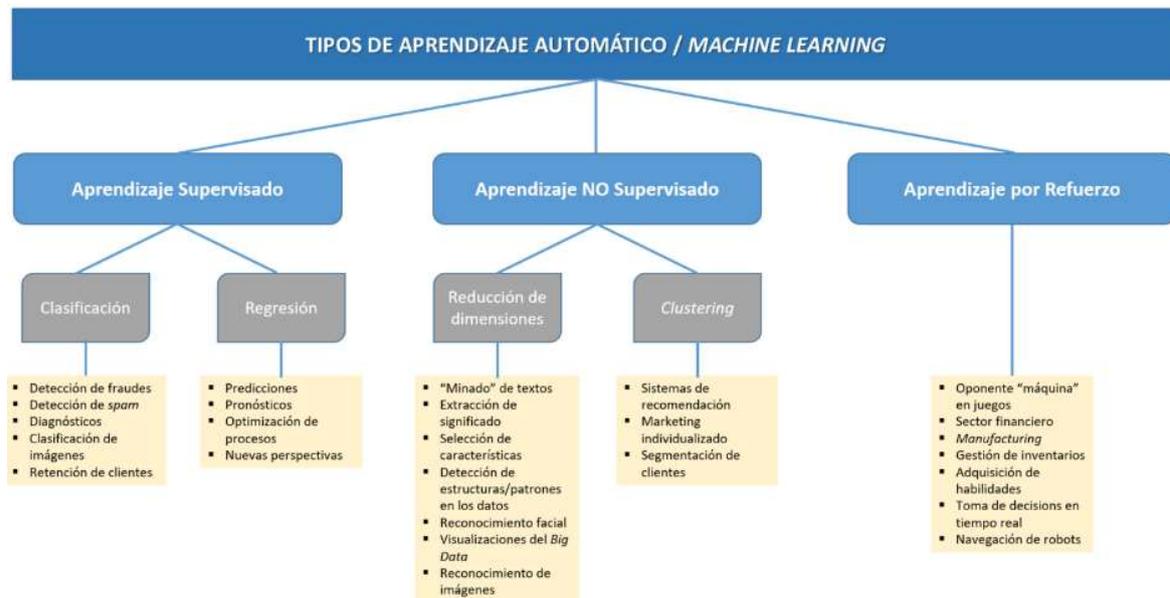
Por otro lado, ISA (2021) menciona para qué usan las empresas la inteligencia artificial:

- Automatizar un proceso o función que normalmente realizaría un ser humano. Por ejemplo, chatbots de atención a clientes.
- Optimizar la eficiencia de un proceso o función. Por ejemplo, mantenimiento predictivo.
- Mejorar la capacidad de las personas para realizar tareas o permitirles hacer algo que normalmente no podrían. Por ejemplo, diagnosticar averías en máquinas y prever la demanda.

Además, Chang (2022) señala que machine learning puede tener variadas aplicaciones, que se exponen en la siguiente figura:

Figura 25

Tipos de “machine learning” clásico, clasificación de sus algoritmos según su uso y aplicaciones



Fuente: Recuperado de ISA (2021).

4.4.1 Diferentes algoritmos según su clasificación

Cuando pretendemos clasificar los principales tipos de aprendizaje, surgen algunas cuestiones relacionadas con las bases de datos disponibles para el análisis.

En primer lugar, ISA (2021) sugiere que podemos comenzar exponiendo los siguientes grupos:

- Con datos simples y categorías bien definidos lo llama aprendizaje clásico.
- Cuando no hay suficientes datos, aprendizaje por refuerzo.
- Cuando la calidad de datos no es buena, más bien un problema, lo agrupa en métodos de ensamble.
- Por último, con datos complejos y categorías mal definidas los agrupa en aprendizaje profundo y redes neuronales.

En tal sentido, ISA (2021) expone:

Figura 26

Tipos de aprendizaje automático según el tipo de datos



Fuente: Recuperado de ISA (2021).

Nuestro proceso en estudio se encuadra en la categoría de aprendizaje automático clásico debido a que maneja datos simples y categorías bien definidas.

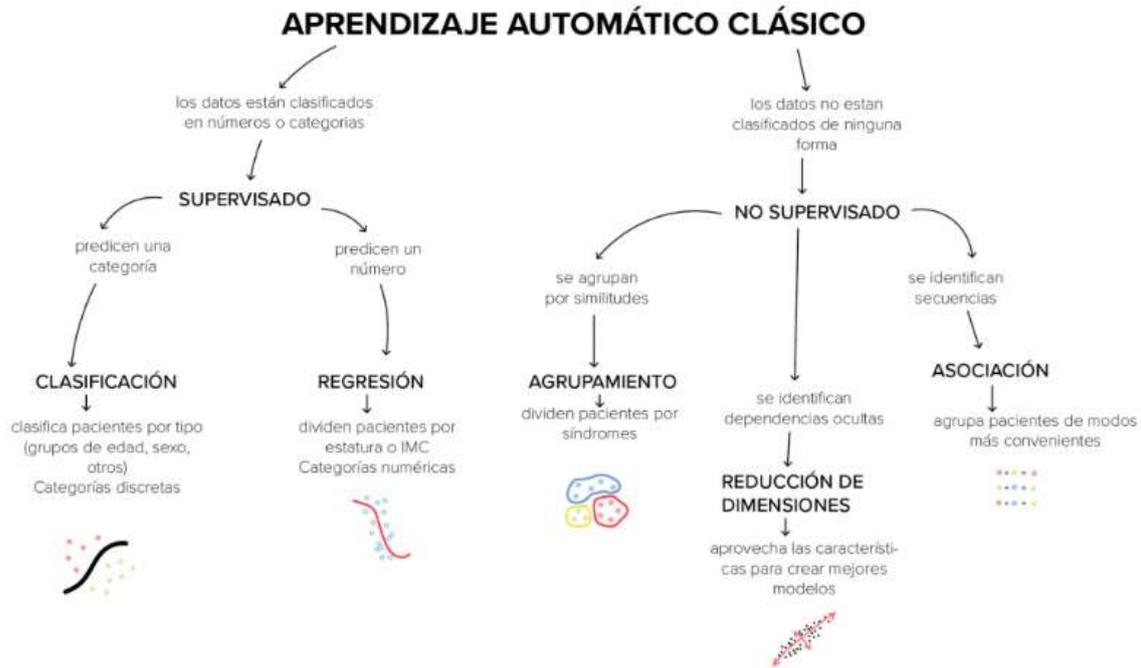
En segundo lugar, ISA (2021) menciona que los tipos de aprendizaje automático clásico pueden clasificarse en supervisados y no supervisados. Los algoritmos supervisados se utilizan cuando los datos están clasificados en números o categorías, mientras que los no supervisados se emplean cuando los datos no están clasificados de ninguna forma.

En este trabajo final utilizaremos los algoritmos supervisados, específicamente en la subcategoría de regresión, dado que necesitamos predecir valores o números.

A continuación, se expone una figura con las agrupaciones mencionadas y otros detalles que completan estas clasificaciones:

Figura 27

Tipos de aprendizaje automático según la estructura de los datos

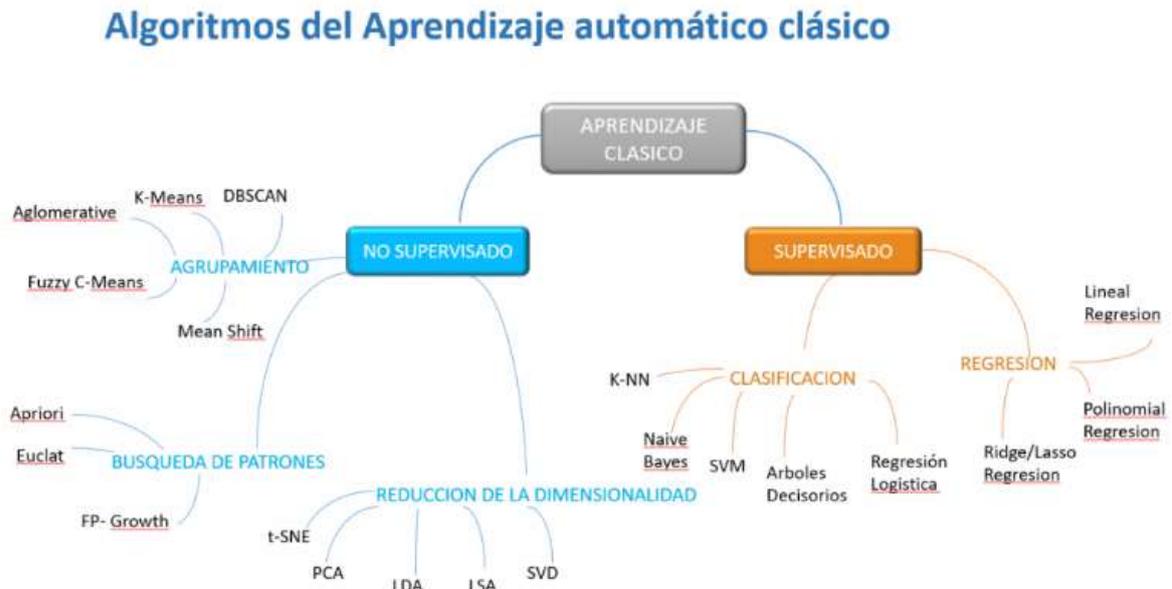


Fuente: Recuperado de ISA (2021).

Por último, dentro de los algoritmos supervisados y de regresión utilizaremos la metodología ARIMA – SARIMA que tratan las series de tiempo específicamente. No obstante, ISA (2021) nos muestra en la siguiente figura los algoritmos más recomendados:

Figura 28

Tipos de aprendizaje automático según los algoritmos utilizados



Fuente: Recuperado de ISA (2021).

4.4.2 Algoritmos no supervisados

Los algoritmos de aprendizaje no supervisado, según Yajure Ramírez (2022), se utilizan cuando no se conoce cuál es la salida correcta del sistema para un conjunto de datos de entrada. Son muy útiles cuando “se tienen datos no etiquetados y se desea explorar su estructura para extraer información significativa. No requieren datos de entrenamiento y se utilizan para descubrir patrones en datos para los cuales no hay una respuesta correcta” (Yajure Ramírez, 2022).

Además, Pedrero et al. (2021) mencionan que “los algoritmos no supervisados se caracterizan por su capacidad para extraer, clasificar y reducir la dimensionalidad. Este tipo de análisis es útil para la generación de nuevas hipótesis de trabajo”.

4.4.3 Algoritmos supervisados

En primer lugar, el objetivo de los algoritmos supervisados, según Pedrero et al. (2021), es la predicción. Para ello, el mismo autor señala que “el algoritmo es entrenado para relacionar un conjunto de variables con un resultado de interés. Una vez que el entrenamiento



se ha completado, el algoritmo puede hacer predicciones utilizando nuevos datos”. Por ejemplo, se podrían utilizar para predecir el riesgo de contraer alguna enfermedad a partir de un conjunto de variables asociadas a su aparición (Pedrero et al., 2021). El resultado de este tipo de análisis “puede ser una clasificación (algoritmos de clasificación), así como también una probabilidad o un puntaje (algoritmo de predicción)”, agrega el mismo autor.

En segundo lugar, Yajure Ramírez (2022) menciona que los algoritmos supervisados requieren que el sistema sea entrenado a partir de datos de entrada y el correspondiente conjunto de datos de salida para obtener el modelo de predicción o de clasificación.

El aprendizaje es supervisado “debido a que el conjunto de entrenamiento es establecido por una persona. Esta persona, ejerciendo el rol de ‘maestro’, le informa al sistema cuál es la salida deseada para un conjunto de elementos de entrada”, agrega el mismo autor.

El proceso de clasificar “las salidas del sistema para un conjunto dado de entradas se conoce como ‘etiquetamiento’, en el sentido de que se indican cuáles salidas del sistema son las correctas para cada conjunto de entradas” (Yajure Ramírez, 2022).

4.4.3.1 Lenguaje de programación: Python.

El lenguaje de programación Python “es un lenguaje de programación de alto nivel creado a finales de los 80 y principios de los 90 por Guido van Rossum. Sus instrucciones están muy cercanas al lenguaje natural en inglés y se hace hincapié en la legibilidad del código” (Quintero, 2023). Toma su nombre de los Monty Python, un grupo humorístico de los 60 que le gustaba mucho a Guido, agrega el mismo autor.

Menciona Quintero (2023) que Python es un lenguaje de programación cuyas principales características son:

- Es dinámico.
- Es interpretado.
- Es de alto nivel.
- Es multiplataforma.
- Es multiparadigma.
- Es de código abierto.



Sus ventajas son:

- Libre y gratuito (OpenSource).
- Fácil de leer, parecido a pseudocódigo.
- Aprendizaje relativamente fácil y rápido, claro e intuitivo.
- Alto nivel.
- Alta Productividad: simple y rápido.
- Tiende a producir un buen código: orden, limpieza, elegancia y flexibilidad.
- Multiplataforma. Portable.
- Multiparadigma: programación imperativa, orientada a objetos y funcional.
- Interactivo, modular y dinámico.
- Librerías extensivas.
- Gran cantidad de librerías de terceros.
- Extensible (C++, C) y embebible.
- Gran comunidad, amplio soporte.
- Interpretado.
- Tipado dinámico¹.
- Fuertemente tipado².

Agrega Quintero (2023) que “al ser un lenguaje de propósito general, podemos encontrar aplicaciones prácticamente en todos los campos científico-tecnológicos”:

- Análisis de datos.
- Aplicaciones de escritorio.
- Bases de datos relacionales / NoSQL.
- Buenas prácticas de programación / Patrones de diseño.
- Concurrencia.
- Criptomonedas / Blockchain.

¹ “Tipado dinámico significa que una variable puede cambiar de tipo durante el tiempo de vida de un programa. C es un lenguaje de tipado estático” (Quintero, 2023).

² “Fuertemente tipado significa que, de manera nativa, no podemos operar con dos variables de tipos distintos, a menos que realice una conversión explícita. Javascript es un lenguaje débilmente tipado” (Quintero, 2023)



- Desarrollo de aplicaciones multimedia.
- Desarrollo de juegos.
- Desarrollo en dispositivos embebidos.
- Desarrollo web.
- DevOps / Administración de sistemas / Scripts de automatización.
- Gráficos por ordenador.
- Inteligencia artificial.
- Internet de las cosas.
- Machine Learning.
- Programación de parsers / scrapers / crawlers.
- Programación de redes.
- Propósitos educativos.
- Prototipado de software.
- Seguridad.
- Tests automatizado.

Para utilizar Python, los programadores requieren una interfaz o entorno de desarrollo llamado IDE. Un IDE es “un entorno de programación que consiste en un editor de código, un compilador, un depurador y un constructor de interfaz gráfica. Facilita al programador el desarrollo de su aplicación” (Riabiz, 2022).

En este trabajo se utilizará Python versión 3.9 y, como IDE, se empleará un software llamado “Jupyter Notebook”. Esta aplicación, de código abierto, permite crear y compartir documentos que contienen código, ecuaciones, visualizaciones y texto narrativo. Puede usarse “para propósitos generales, aunque suele estar más enfocada en ciencia de datos: limpieza y transformación de datos, simulación numérica, modelado estadístico, visualización o ‘Machine Learning’” (Quintero, 2023). Para finalizar, podemos verlo como un intérprete de Python (contiene un “kernel³” que permite ejecutar código) con la

³ “Proceso específico para un lenguaje de programación que ejecuta instrucciones y actúa como interfaz de entrada/salida” (Quintero, 2023).



capacidad de incluir documentación en formato Markdown⁴ lo que potencia sus funcionalidades agrega el mismo autor.

4.4.4 Implementación de algoritmos para predicciones de series de tiempo

Como se ha mencionado anteriormente, ARIMA es un algoritmo para la predicción de series de tiempo. Se clasifica dentro de los algoritmos supervisados como un método de aprendizaje automático clásico de regresión. Este algoritmo se utilizará en el presente trabajo.

Para su implementación, se requiere una base de datos de al menos 5 años, proporcionada con datos que cumplan ciertas características, como precisión, validez e integridad, entre otras. Obtener una base de datos con estas características es un paso crucial en el desarrollo de este tipo de proyectos.

En concordancia con este trabajo, sería beneficioso que la organización tenga una gestión por procesos establecida, o, en su defecto, que se realice un análisis de gestión de procesos para todas las actividades relacionadas con la implementación de las mejoras deseadas.

Así, partiendo del ciclo de mejora continua, en sus etapas de verificar y actuar para mejorar, se introducirá el análisis y las acciones correspondientes derivadas de los hallazgos obtenidos a partir de la predicción.

4.5 Machine learning en la gestión de procesos

Las empresas utilizan ecosistemas de inteligencia de negocio, según Martínez Zabaleta y Luna (2023), con el fin de reducir los costos en sus operaciones productivas. A partir de estos ecosistemas, se deben tomar decisiones basadas en información que, como mencionan Quinto et al. (2021), consiste en identificar una problemática y encontrar una solución.

En 1956, según menciona García Herrera (2020), se dio inicio a la era de la inteligencia artificial cuando los científicos John McCarthy, Marvin Minsky y Claude Shannon dieron a

⁴ “Markdown también se considera un lenguaje que tiene la finalidad de permitir crear contenido de una manera sencilla de escribir, y que en todo momento mantenga un diseño legible, así que para simplificar puedes considerar Markdown como un método de escritura” (Cristobal, 2016).



conocer el término "inteligencia artificial" para referirse a "la ciencia e ingenio de hacer máquinas inteligentes, principalmente programas de cálculos inteligentes".

Actualmente, se piensa en la inteligencia artificial como la capacidad de las máquinas para realizar tareas que normalmente requerirían inteligencia humana, como el aprendizaje, la percepción, el razonamiento y la resolución de problemas, según menciona Rubín (2024). Nieves et al. (2024) señalan que la adopción de la inteligencia empresarial ha crecido, impulsada por los avances tecnológicos y la necesidad de competir que tienen las empresas. Esto ha permitido automatizar tareas, optimizar procesos y tomar decisiones basadas en datos de manera más eficiente y precisa. Rubín (2024) también menciona que la inteligencia artificial se ha vuelto esencial para crear productos y servicios más inteligentes, permitiendo a las organizaciones ser más competitivas, aumentar la eficiencia operativa y reducir costos. Además, según García Herrera (2020), la inteligencia artificial se ha convertido en un factor importante en las organizaciones y el mundo laboral, por lo que se considera que las organizaciones deben estar preparadas para utilizar esta tecnología en sus procesos.

Así pues, Rubín (2024) sostiene que la inteligencia artificial es un nuevo activo intangible que genera un valor agregado en cada componente de la organización empresarial.

Dentro de lo que se considera inteligencia artificial, ISA (2021) define "machine learning" como un subconjunto de esta que, mediante el uso de ciertas técnicas, permite mejorar el desempeño en algunas tareas a partir de la "experiencia".

Además, Quinto et al. (2021) mencionan que estas tecnologías pueden ayudar a los tomadores de decisiones humanos con análisis predictivos. Una de las funciones principales del análisis predictivo es generar nueva información y predicciones sobre clientes, activos y operaciones. Los mismos autores afirman que el uso en la toma de decisiones organizacionales sigue siendo una de sus principales aplicaciones.

En cuanto al campo de aplicación, García Herrera (2020) menciona que los sectores más destacados son la salud, la educación, la producción, el desarrollo empresarial, la industria automotriz, la vigilancia, la seguridad, el sistema bancario, la meteorología y la climatología, entre otros.

En esa línea, existen investigaciones y estudios sobre la aplicación de modelos de "machine learning" para la mejora de procesos en la gestión asistencial de la salud (Pedrero et al., 2021), las ventas (Villachica Pérez et al., 2022), la gestión de inventarios (Gonzales et



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



al., 2023), la gestión de cobranzas (Mendoza, P. y Anthony, R., 2022), la gestión de proyectos de software (Snider, N., 2020), etc.



5 Desarrollo

5.1 Reseña de la empresa

La empresa se estableció en diciembre de 1992 tras la privatización de Gas del Estado. Distribuye gas natural y gas licuado brindando servicio a 661.081 clientes y tiene una concesión de 35 años.

Su misión es ser una empresa reconocida por la calidad y fiabilidad del servicio. Contar con un personal competente y motivado, involucrado con los pensamientos estratégicos de la organización y que aporte su potencial para lograr los objetivos.

Su visión es desarrollar las fronteras del negocio, tanto en su extensión como en su uso teniendo como valores la exigencia, compromiso, audacia y cohesión.

Como estrategia la empresa está enfocada en expandirse manteniendo la calidad y fiabilidad del servicio. La transformación digital es una estrategia de innovación que se está llevando a cabo en las distintas gerencias como meta del grupo.

La organización se sitúa como una empresa del rubro energético en la Argentina y se desempeña como monopolio en su zona distribuyendo gas en media y alta presión a distintos clientes. Tiene asignada una zona geográfica de incumbencia donde presta sus servicios a grandes clientes y residenciales.

En cuanto a su estructura organizativa la empresa desarrolla la actividad con base en una provincia de Argentina. Posee una sede central y varias sucursales instaladas en las ciudades más importantes de las provincias que presta servicios. Cuenta con aproximadamente 400 empleados y recursos contratados que realizan las actividades necesarias para prestar el servicio y mantener el sistema.

La empresa se encuentra encuadrada como sociedad anónima, que es una forma de organización de tipo capitalista muy utilizada entre las grandes compañías en donde todo el capital se encuentra dividido en acciones, que representan la participación de cada socio en el capital de la compañía.

5.1.1 Marco regulatorio

La Ley 24.076 llamada marco regulatorio de la actividad del gas natural crea en su artículo 50, en el ámbito del Ministerio de Economía y Obras y Servicios Públicos, el ENTE NACIONAL REGULADOR DEL GAS (ENARGAS).



En el Artículo 52 de la mencionada ley “se fijan las facultades del ENARGAS, entre las cuales se incluye la de dictar reglamentos a los que deberán ajustarse todos los sujetos de esta Ley en materia de seguridad, normas y procedimientos técnicos” dice (ENARGAS, 2016).

En tal sentido dice el mismo autor que “el ENARGAS auspició la redacción de las NORMAS ARGENTINAS MINIMAS DE SEGURIDAD PARA EL TRANSPORTE Y DISTRIBUCION DE GAS NATURAL y OTROS GASES POR CAÑERIAS, denominada N.A.G.-100”.

La N.A.G.-100 establece los estándares de diseño, operación y mantenimiento para las instalaciones de transmisión y distribución. Agrega ENARGAS (2016) que la seguridad es la consideración básica de la norma y al no tener la N.A.G.-100 características de manual de diseño, la utilización de sus contenidos deberá ser acompañada por apropiados criterios de ingeniería”.

5.1.2 Servicio Ofrecido

La empresa comercializa gas natural en media presión a casi todas las localidades de su jurisdicción y en algunas llega con gas licuado. Además, suministra servicios de gas natural en alta presión a distintas industrias.

Describe Camuzzi (2019) que “el gas natural es un hidrocarburo derivado del petróleo, incoloro, insípido e inodoro”. El mismo autor describe su composición química:

Metano (CH₄) de 85 % a 90 %.

Etano:(C₂H₆) de 5% a 6 %.

Propano:(C₃H₈) de 1,5% a 2 %.

Butano :(C₄H₁₀) de 0,8% a 1%.

Pentano:(C₅H₁₂) de 0,10% a 0,30%.

Hexano y superiores:(C₆ y sup.) de 0,05% al 0,1%.

Anhídrido Carbónico:(CO₂) de 0,5% al 1%.

Nitrógeno:(N₂) de 1% al 2%.

Densidad relativa = 0,618 (para aire = 1) la cual facilita el venteo y la disipación hacia la atmósfera por ser más liviano que el aire.



Los límites de Inflamabilidad para mezclas de gas / aire son: 5% para el límite inferior y 15% para el superior.

A presión atmosférica y temperaturas normales su estado es gaseoso.

Por otro lado, el gas licuado "es más pesado que el aire cuando está a temperatura y presión normales, lo cual impide su fácil venteo. El propano es 1,5 veces y el butano 2 veces más pesado que el aire" (Camuzzi, 2019).

El mismo autor agrega que "son incoloros, pero al vaporizarse forman nieblas al condensarse la humedad ambiente. Debido a estas características, los riesgos consecuentes de un escape de gas licuado son substancialmente mayores que los de un escape de gas natural".

5.2 Enfoque en procesos

Los datos son un elemento fundamental en la gestión empresarial. Sus orígenes son muy diversos, incluyendo formatos estructurados, que se gestionan en bases de datos, y formatos no estructurados, que son datos que se originan sin un formato adecuado para su análisis en conjunto.

Como sucede en casi cualquier organización, la cantidad de datos que se generan diariamente es muy grande. Solo aquellos datos que se consideran importantes en principio se les dan un formato para su tratamiento con distintos fines. En la empresa en estudio, algunos de los datos importantes incluyen los económicos-financieros, los datos presentados al ente regulador, los datos de clientes, y los referentes a los procesos de operación y mantenimiento de las instalaciones inmersas en los procesos técnicos.

Los datos relacionados con los procesos de operación y mantenimiento se gestionan mediante un sistema ERP (siglas de 'Enterprise Resource Planning' o 'Planificación de Recursos Empresariales') llamado Maximo.

El enfoque del proyecto estuvo en la gerencia de operaciones.

En la empresa, existe una organización del tipo funcional desde sus comienzos, gestionando sus tareas a través de actividades enfocadas en sectores, algunas de las cuales son exigidas por un ente regulador que "ordena" determinados controles mínimos sobre las instalaciones con una frecuencia estipulada.



Estos controles y otros permiten que la estructura de distribución de energía se mantenga para ofrecer a los usuarios un producto final de calidad.

Todas las actividades están clasificadas y agrupadas en conjuntos de instrucciones de trabajo, conocidos como planes de trabajo.

Cada plan de trabajo describe los pasos para realizar una actividad específica. Estos planes tienen una revisión periódica para detectar cambios o mejoras. Están documentados en secciones de trabajo que la empresa administra y se incluyen en el sistema ERP para que cada trabajo realizado a través de una orden esté guiado por un plan de trabajo.

A modo de ejemplo, la siguiente figura muestra un plan de trabajo para el mantenimiento trimestral de vehículos de la flota de la empresa:

Figura 29

Instrucciones del plan de trabajo MAN-VEH

PLAN DE TRABAJO		
Fecha actualización:	15/1/2021	
Nombre del plan	Secuencia	Tareas
MAN-VEH	10	Revisión/cambio de aceite, filtro de aire y combustible
	20	Inspección general de mangueras y circuito eléctrico
	30	Inspección/recambio y ajustes de correas
	40	Control de alternador, batería, regulador y burro de arranque
	50	Control de amortiguadores
	60	Control de tren delantero y trasero
	70	Control de ejes y semiejes
	80	Control/recambio de frenos
	90	Control de radiador y sistema de refrigeración
	100	Revisión del embrague
	110	Verificar estado de cubiertas. Alineación y balanceo
	120	Revisión del sistema 4x4
	130	Verificación de los levanta vidrios
	140	Verificar estado de luces traseras y delanteras
	150	Verificar funcionamiento de bocina
	160	Control de AA
	170	Describir observaciones particulares



Fuente: Diseño propio.

Como se observa, el plan de trabajo incluye una serie de instrucciones que el responsable de realizar estas tareas debe cumplir. En este plan específico, todos los datos que provienen del campo se registran como observaciones y se cargan en el sistema en ese formato.

Cada orden de trabajo se genera para cada tarea que se realiza. Esta orden comprende el plan de trabajo mencionado anteriormente, la ubicación y el activo (si corresponde), la descripción del trabajo y el supervisor responsable, entre otros campos principales.

Mediante la gestión en el sistema de estas órdenes de trabajo, se ejecutan las diferentes actividades realizadas anualmente por la jefatura sobre las instalaciones. La mayoría de los trabajos se gestionan como mantenimientos preventivos. Estos trabajos se administran a través de sus planes, que cuentan con un identificador único y una frecuencia específica determinada por el ente regulador o el sector responsable.

Las distintas órdenes de trabajo son generadas y administradas en cada sector para llevar a cabo mantenimientos preventivos, correctivos o reportes de eventos. Se imprimen y se entregan a los responsables de su ejecución, que pueden ser empleados que trabajan en la misma sede, contratistas, o se envían por correo privado a otra sucursal donde el personal designado realizará la tarea. El seguimiento y control de los trabajos se realiza telefónicamente o a través de cualquier medio de mensajería instantánea; ocasionalmente, se lleva a cabo una inspección presencial.

Las órdenes de trabajo, después de ser ejecutadas, regresan al sector correspondiente por el mismo medio por el cual fueron entregadas, ya sea en persona o por correo. Cada documento tiene un formato específico dependiendo del plan de trabajo, con registros y campos que el operario debe completar con valores derivados de sus tareas. Las incidencias o fallas se registran mediante códigos preestablecidos en el sistema. Otra opción es escribir comentarios al final de la orden, los cuales reflejan situaciones no previstas o eventos importantes ocurridos durante las actividades.

En ocasiones, a partir de las fallas registradas en el sistema, se generan órdenes de trabajo para mantenimiento correctivo basadas en estas incidencias encontradas en el campo.



Los datos estructurados son aquellos que el sistema admite en formato de campos preestablecidos y con valores esperados. Los datos provenientes de comentarios escritos, aunque se pueden cargar, son difíciles de analizar en el volumen general. Algunos de estos datos cargados en el sistema se convierten en información solicitada por el ente regulador según el formato requerido.

Para proporcionar una idea general, se administran alrededor de 7,000 órdenes de trabajo por año en esta jefatura, cada una procesada en el sector correspondiente. La empresa posee un archivo físico propio y alquilado para almacenar estos registros, tanto de la jefatura en estudio como de los demás sectores de la organización. Los archivos físicos se refieren a lugares o salas destinadas al almacenamiento de papeles, carpetas o documentos que la organización considera necesarios conservar.

Recientemente se ha incorporado el análisis de datos a través de la herramienta Power BI. Para construir el informe, fue necesario acceder a los datos del sistema y crear bases de datos que puedan ser "leídas" e interpretadas por el sistema de inteligencia de negocios. Una vez logrado esto, se diseña el informe. El proceso de actualización de los datos en estas bases es manual, requiriendo la ejecución de diferentes informes en el ERP, su descarga en formato de base de datos y la actualización de estos para su uso en el tablero de inteligencia de negocios.

5.3 Gestión de procesos operativos

Una de las estrategias de la gerencia de operaciones es mantener las instalaciones seguras y confiables, lo que garantiza la calidad del producto final entregado.

La estructura de la empresa está organizada a través de gerencias, jefaturas, sectores y operarios.

La gerencia es responsable de la operación y mantenimiento del sistema que permite prestar el servicio al cliente final. Está compuesta por siete sectores, cada uno de los cuales se enfoca en mantener diferentes activos de las instalaciones. Cada sector está liderado por un supervisor que, junto con un grupo de operarios especializados, lleva a cabo las tareas de manera rutinaria. Estas tareas están documentadas en secciones de trabajo que se actualizan periódicamente en busca de mejoras. La capacitación del personal, tanto nuevo como existente, se realiza mayoritariamente por personal interno de la empresa de forma periódica.



Eventualmente, un operario de un sector participa en tareas fuera de su área específica. El personal en las sucursales, bajo la supervisión de un supervisor, recibe trabajos de todos los sectores de manera coordinada, lo que le permite adquirir un conocimiento más amplio de las tareas de la jefatura.

Las actividades, como se ha descrito, se gestionan a través de sectores especializados, manteniendo una estructura de trabajo funcional por departamento.

Como se mencionó anteriormente, la gestión de los procesos implica la aplicación del ciclo de mejora continua, que incluye las fases de planificar, actuar, verificar y mejorar.

5.4 Ciclo de mejora continua

5.4.1 Planificación global de procesos

Se realizó un análisis de los procesos involucrados en esta gerencia, destacando la cadena de valor de la empresa, que es la que origina los procesos operativos de la compañía.

Figura 30

Cadena de valor de la empresa



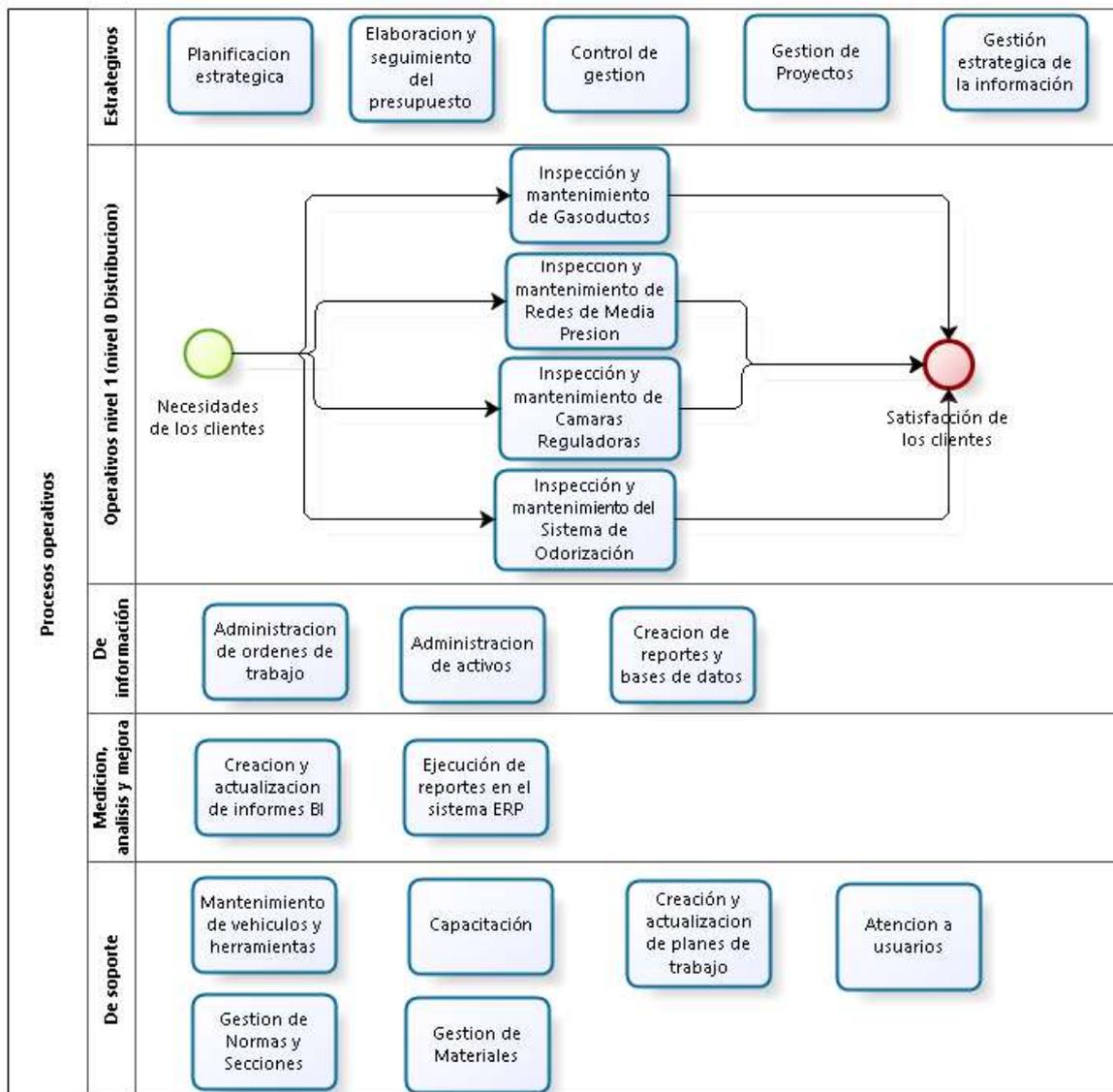
Fuente: Diseño propio.

Las actividades primarias se consideran en el mapa de procesos de la organización como los procesos operativos de nivel 0. A partir de “Distribución: Operaciones y

Mantenimiento” y desglosando los procesos que se derivan de este, se obtiene un nivel 1 en el que se construye el mapa de procesos de la gerencia de operaciones.

Figura 31

Mapa de procesos operativos



Fuente: Diseño propio.

Se construyó este mapa ampliando la clasificación de procesos clásica (estratégicos, operativos y de soporte) e incorporando los procesos “De información” y “Medición, análisis y mejora” con el fin de lograr una mejor claridad en la agrupación de estos. Aunque este



mapa es la base de la gestión de la jefatura, los procesos operativos son de “nivel 1” ya que representan el desglose de un proceso de nivel 0 de la organización.

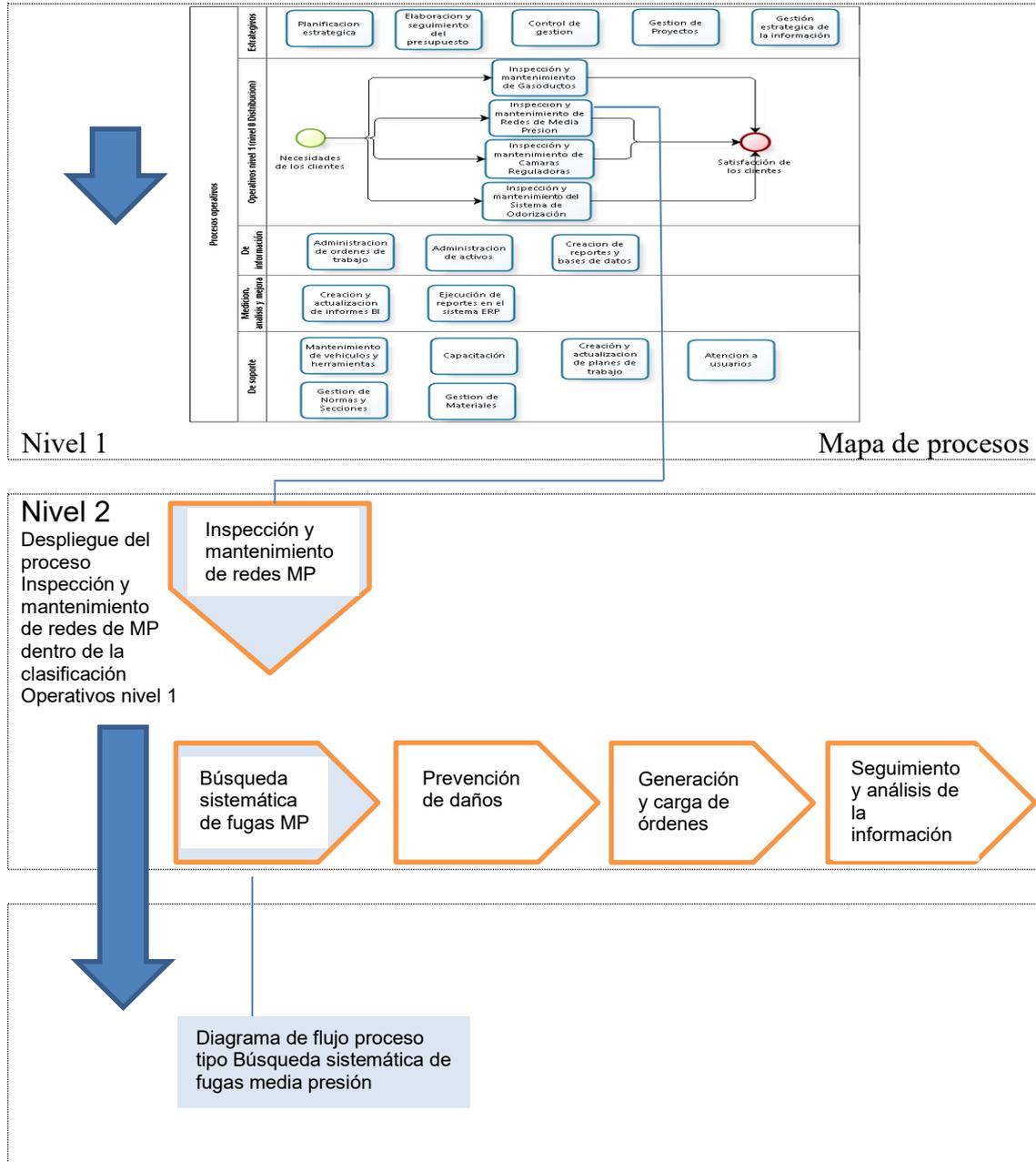
Algunas pautas utilizadas para construir este mapa de procesos son:

- Definir el alcance del mapa de procesos: No es lo mismo dibujar el mapa de procesos de un área que el de la organización.
- Inventariar los procesos: Dependiendo de las actividades que desarrolla la organización.
- Documentar los procesos.
- Clasificar los procesos.
- Interrelacionar los procesos.

Para profundizar en la gestión de estos, se tomó el proceso “Inspección y mantenimiento de redes de media presión” correspondiente a la clasificación “Operativos nivel 1” para desglosar el mismo al siguiente nivel como se muestra en la siguiente figura:

Figura 32

Desglose del mapa de procesos de la gerencia de operaciones



Fuente: Diseño propio.

Entonces, el nivel 2 para “Inspección y mantenimiento de redes de media presión” contiene los procesos:



- Búsqueda sistemática de fugas media presión: actividades que se encargan de recorrer las redes de media presión en toda la extensión de la distribuidora de gas en búsqueda de posibles pérdidas.
- Prevención de daños: actividades que se encargan de evitar o minimizar los daños producidos por terceros sobre las cañerías de gas.
- Generación y carga de órdenes: son actividades que permiten que cada trabajo que se realice este encuadrado en algún procedimiento que la empresa tiene descripto. A su vez se recogen en la misma la mano de obra vinculada y las fallas que se encuentran.
- Seguimiento y análisis de la información: el seguimiento contempla las actividades de gestión que cada supervisor debe realizar en cuanto al control de los datos y su correcta carga en el sistema. El análisis se realiza a través de informes del sistema que permiten tener una primera evaluación del estado de los activos.

5.4.2 Planificación individual de procesos

Según lo expuesto en la figura anterior, el proceso de nivel 2 llamado “Búsqueda sistemática de fugas media presión”, se considera un proceso tipo dado que sus actividades son suficientemente explicativas del objetivo del proceso en sí y no requieren agruparse como otro proceso adicional o de otro nivel.

5.4.2.1 Diagrama de Flujo.

En el diagrama de flujo del proceso “Búsqueda sistemática de fugas media presión” se pudo identificar 7 calles o responsables de realizar determinadas actividades y que son:

- Coordinadores.
- Contratistas A y B.
- Sistema Maximo – ERP.
- Sistema comercial.
- Personal propio.
- Contratistas C y D.
- Operador de guardia.

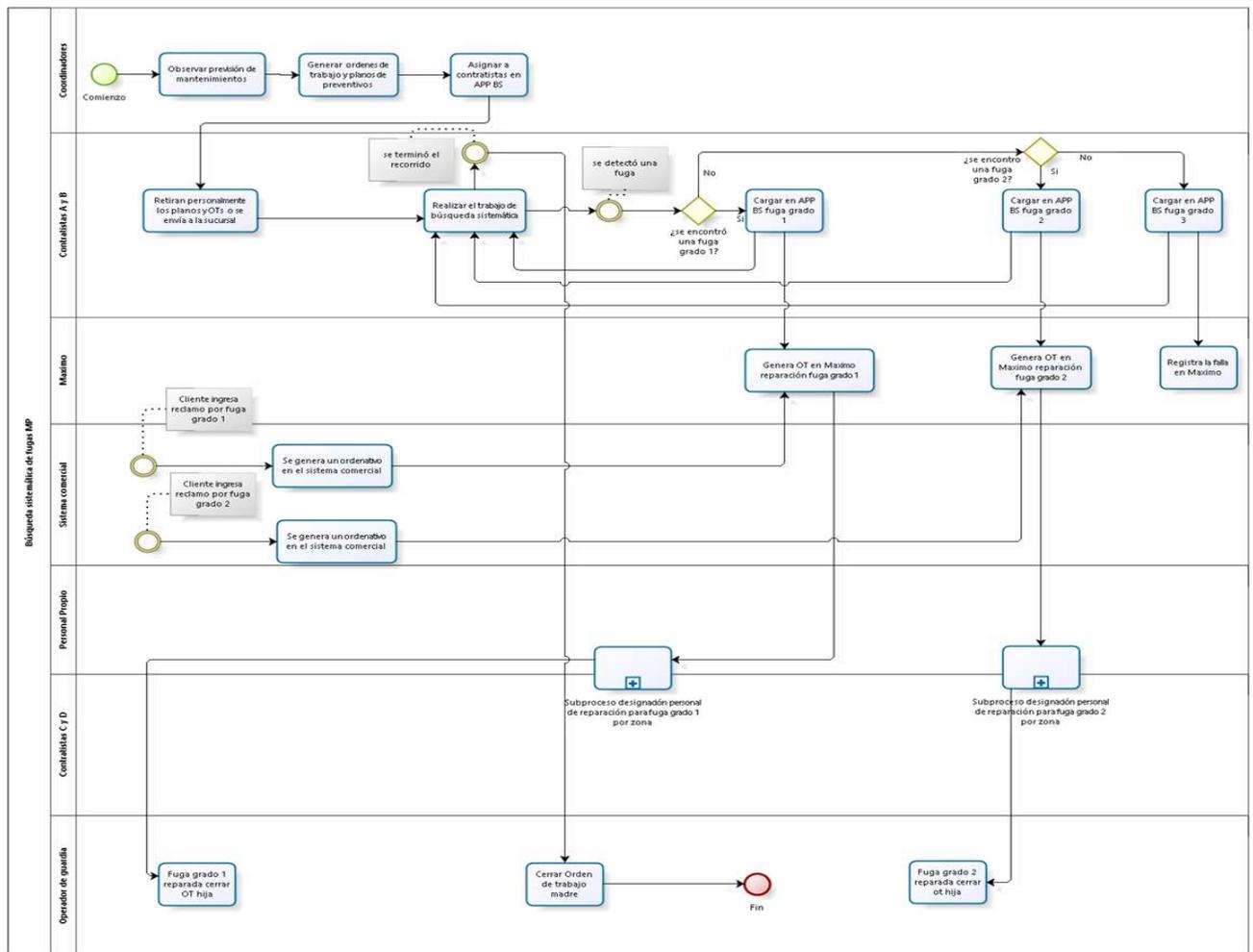
Además, se diagramaron dos subprocesos con el fin acotar el proceso principal y a fin de comprender mejor la dinámica de este.

- Subproceso designación personal de reparación para fuga grado 1 por zona.
- Subproceso designación personal de reparación para fuga grado 2 por zona.

En la figura siguiente se muestra el diagrama de proceso con todas sus actividades y los responsables de su ejecución.

Figura 33

Proceso búsqueda sistemática de fugas media presión



Fuente: Diseño propio.



Los coordinadores son personas con conocimiento del proceso a quienes el supervisor les delega actividades de gestión relacionadas con los mantenimientos preventivos y el contacto con las empresas contratistas.

Las contratistas A y B son las encargadas de recorrer las cañerías de media presión en busca de fugas, utilizando una orden de trabajo y un plano de las instalaciones. Además, cuentan con un celular con una aplicación satelital del sistema de gestión de activos Maximo, que les permite registrar las fallas encontradas. Si algunas de las fallas son fugas de grado 1 o 2, se transmiten al sistema principal, generando una orden de trabajo para su reparación según el tipo de falla.

El sistema Maximo es un ERP de la empresa IBM adquirido por la organización para gestionar activos. El sistema comercial es otro software relacionado con Maximo, que registra diversas actividades, incluyendo los reclamos ingresados por teléfono. Específicamente, si los reclamos se refieren a fugas, se registran en el sistema comercial y también se genera una orden de trabajo que debe ser atendida.

El personal propio de la organización realiza algunas reparaciones de fugas, dependiendo de la disponibilidad y la ubicación de estas. Esta gestión está integrada en los subprocesos diagramados, aunque no se profundiza en ellos para mantener el enfoque en el proceso general.

Las contratistas C y D se dedican a reparar la mayor parte de las fugas de grado 1 y 2 detectadas. Su operación está contratada por zona geográfica, dividiendo la zona de acción de la organización en dos partes.

El operador de guardia, entre otras actividades, también puede registrar reclamos de fugas. Sin embargo, en nuestro proceso se destaca que su responsabilidad principal es cerrar los trabajos reportados como terminados, cambiando el estado de las órdenes de trabajo a "cerrado".

5.4.2.2 Ficha del proceso.

La ficha del proceso "Búsqueda sistemática de fugas en media presión" es muy útil para conocer rápidamente quién es responsable del proceso, qué necesidades atiende, qué variables intervienen y cuáles son sus objetivos, entre otras características.



Tener documentados los procesos a través de su diagrama y ficha resulta muy valioso cuando se busca detectar mejoras en el mismo.

Figura 34

Ficha del proceso Búsqueda Sistemática de Fugas Media Presión

-LOGO-		FICHA DE DESCRIPCIÓN DEL PROCESO	
NOMBRE DEL PROCESO		RESPONSABLE DEL PROCESO	
Búsqueda sistemática de fugas en MP		Supervisor de BS	
FINALIDAD			
Recorrer los km de redes de media presión en búsqueda de fugas y atender los reclamos de clientes. Reparar las de grado 1 de forma inmediata. Reparar las de grado 2. Registrar las fugas grado 3.			
LÍMITES DEL PROCESO			
Primera actividad		Última actividad	
<ul style="list-style-type: none"> • Atender reclamos de clientes por fugas • Recorrer las redes en búsqueda de fugas 		<ul style="list-style-type: none"> • Todas las redes recorridas • Reparar las fugas grado 1 en el mismo día • Reparar fugas grado 2 	
Entradas		Salidas	
<ul style="list-style-type: none"> • Reclamos de clientes • Mantenimientos preventivos 		<ul style="list-style-type: none"> • Ordenativo del sistema comercial atendido • Ordenes de trabajo de BS cerradas • Ordenes de trabajo de fugas a reparar cerradas. 	
Proveedores del proceso		Clientes del proceso	
<ul style="list-style-type: none"> • Cliente • Departamento de búsqueda sistemática 		<ul style="list-style-type: none"> • Usuarios de gas • ENARGAS 	
AGENTES DEL PROCESO			
<ul style="list-style-type: none"> • Atención al cliente • Cliente • Sector de BS • ENARGAS • Sector MR 			
INDICADORES DEL PROCESO			
Indicador		Gestionado por	
1. Fugas grado 1 por km recorridos 2. Cantidad de fugas grado 1 reparadas 3. Cantidad de fugas grado 2 reparadas 4. Costo medio de reparación de fuga grado 1 4. Cumplimiento de frecuencia (ENARGAS) 6. Satisfacción en el servicio al cliente		1. Supervisor de Búsqueda Sistemática	
DOCUMENTACIÓN RELACIONADA			
<ul style="list-style-type: none"> • Ordenes de trabajo, partes diarios, ordenativos del sistema comercial, planos recorridos. 			

Fuente: Diseño propio.



En este trabajo, centraremos especial atención en los indicadores de fugas de grado 1 y en aquellos relacionados con el ente regulador del gas.

5.4.2.3 Ficha de los indicadores.

La ficha de cada indicador expone rápidamente muchas características como por ejemplo quien es el responsable, como se calcula, de donde se extrae la información para su cálculo y cuál es su objetivo entre otras variables de interés.

A continuación, se expone la ficha de los siguientes indicadores:

- Fugas grado 1 por km de cañería en MP.
- Cantidad de fugas grado 1 reparadas en el año en curso.
- Costo medio de reparación de fugas grado 1 por terceros.
- Cumplimiento de frecuencia (ENARGAS).

Figura 35

Ficha del indicador Fugas grado 1 por km de cañería en MP

-LOGO-				FICHA TÉCNICA DEL INDICADOR					
INDICADOR		Fugas grado 1 por km de cañería en MP		CÓDIGO		IPBS001			
INFORMACIÓN GENERAL DEL INDICADOR									
DESCRIPCIÓN GENERAL						AREA			
Cálculo de las fugas grado 1 encontradas y reparadas por kilómetros recorridos en redes de media presión						Mantenimiento de redes			
OBJETIVOS									
Facilitar la interpretación de las manifestaciones de las redes de media presión en cuanto a los problemas detectados en toda su longitud									
RESPONSABLES									
Supervisor de Búsqueda Sistemática									
FORMA DE REGISTRO EN EL SISTEMA DE INFORMACIÓN									
INTERFACE MANUAL						INTERFACE AUTOMÁTICA		X	
MÉTODO DE CÁLCULO DEL INDICADOR									
FÓRMULA				META					
FpK= CF1 / KMP				Que el valor no aumente					
FpK= fuga por kilómetros CF1= cantidad de fugas grado 1 detectadas KMP= longitud de la red de media presión				OBJETIVO		META		DESAFIO	
				FpK= año anterior		FpK < año anterior		FpK= ¾ año anterior	
PREMISAS DEL INDICADOR									



Estar disponible para análisis de proceso	
VARIABLE REQUERIDA	FUENTE DE INFORMACIÓN
Longitud de la red de media presión en Km	Sistema comercial
Cantidad de fugas grado 1 detectadas	Aplicación mobile de búsqueda sistemática

Fuente: Diseño propio.

Figura 36

Cantidad de fugas grado 1 reparadas

-LOGO-	FICHA TÉCNICA DEL INDICADOR		
INDICADOR	Cantidad de fugas grado 1 reparadas	CÓDIGO	IPBS002
INFORMACIÓN GENERAL DEL INDICADOR			
DESCRIPCIÓN GENERAL			AREA
Recuento de fugas grado 1 reparadas en el año en curso			Mantenimiento de redes
OBJETIVOS			
Conocer la evolución de las fugas grado 1 en el sistema			
RESPONSABLES			
Supervisor de Búsqueda Sistemática			
FORMA DE REGISTRO EN EL SISTEMA DE INFORMACIÓN			
INTERFACE MANUAL		INTERFACE AUTOMÁTICA	X
MÉTODO DE CÁLCULO DEL INDICADOR			
FÓRMULA		META	
CFG1=Recuento de fugas grado 1		Que el valor no aumente	
	OBJETIVO	META	DESAFIO
	CFG1= año anterior	CFG1 < año anterior	CFG1= ¾ año anterior
PREMISAS DEL INDICADOR			
Estar disponible para análisis de proceso			
VARIABLE REQUERIDA		FUENTE DE INFORMACIÓN	
Las fugas cargadas en el sistema		Aplicación mobile de búsqueda sistemática	

Fuente: Diseño propio.



Figura 37

Costo medio por fugas grado 1 reparadas por terceros

-LOGO-				FICHA TÉCNICA DEL INDICADOR			
INDICADOR		Costo medio por fugas grado 1 reparadas por terceros		CÓDIGO		IPBS004	
INFORMACIÓN GENERAL DEL INDICADOR							
DESCRIPCIÓN GENERAL						AREA	
Costo medio que la empresa paga a un contratista para realizar las reparaciones en las localidades asignadas por contrato						Mantenimiento de redes	
OBJETIVOS							
Llevar un control del costo por fuga grado 1 reparada							
RESPONSABLES							
Supervisor de Búsqueda Sistemática							
FORMA DE REGISTRO EN EL SISTEMA DE INFORMACIÓN							
INTERFACE MANUAL		X		INTERFACE AUTOMATICA			
MÉTODO DE CÁLCULO DEL INDICADOR							
FÓRMULA				META			
CFG1R= \sum costo por fugas grado 1 reparadas / Recuento de fugas grado 1				Monitorear los costos			
OBJETIVO		META		DESAFIO			
CFG1R= año anterior		CFG1R < año anterior		CFG1R= ¾ año anterior			
PREMISAS DEL INDICADOR							
Estar disponible para el monitoreo de los costos y la negociación del próximo contrato							
VARIABLE REQUERIDA				FUENTE DE INFORMACIÓN			
Recuento de fugas grado 1				Aplicación mobile de búsqueda sistemática			
Costos por fugas				Partes mensuales de la contratista			

Fuente: Diseño propio.



Figura 38

Cumplimiento de frecuencia de mantenimientos preventivos previstos por el ENARGAS

-LOGO-	FICHA TÉCNICA DEL INDICADOR		
INDICADOR	Cumplimiento de frecuencia de mantenimientos preventivos previstos por el ENARGAS	CÓDIGO	IPBS005
INFORMACIÓN GENERAL DEL INDICADOR			
DESCRIPCIÓN GENERAL		AREA	
Refiere a mantenimientos preventivos asociados con la búsqueda sistemática anual por sistemas que el ente regulador solicita por año		Mantenimiento de redes	
OBJETIVOS			
Llevar un control de la ejecución de los mantenimientos preventivos			
RESPONSABLES			
Supervisor de Búsqueda Sistemática			
FORMA DE REGISTRO EN EL SISTEMA DE INFORMACIÓN			
INTERFACE MANUAL		INTERFACE AUTOMÁTICA	X
MÉTODO DE CÁLCULO DEL INDICADOR			
FÓRMULA		META	
AMP = (Cantidad de MP ejecutados a la fecha / Cantidad de MP previstos a la fecha) *100		Cumplir con los requerimientos	
	OBJETIVO	META	DESAFIO
	AMP = 100%	AMP = 100%	AMP = 100%
PREMISAS DEL INDICADOR			
Cumplir con los requerimientos del ENARGAS			
VARIABLE REQUERIDA		FUENTE DE INFORMACIÓN	
Cantidad de MP ejecutados a la fecha		ERP Maximo	
Cantidad de MP previstos a la fecha		ERP Maximo	

Fuente: Diseño propio.

5.4.3 Ejecución del proceso

Dentro de la gestión de procesos, en el ciclo de mejora continua, el “hacer” representa la práctica de lo que se ha planificado, comenta Pardo Álvarez (2017).



En el proceso de “Búsqueda sistemática de fugas en MP”, dado que los responsables incluyen tanto personal de la empresa como contratistas, los procedimientos y actividades no son exactamente iguales. El personal de la empresa energética sigue procedimientos a través de órdenes de trabajo. Cada una de estas órdenes, mediante un plan de trabajo, indica a los operarios las actividades a realizar en una secuencia determinada. La empresa contratista, además de cumplir con estos procedimientos, debe atender las cláusulas del contrato establecido.

Los planes de trabajo se actualizan periódicamente. Según Pardo Álvarez (2017), gran parte de los problemas en la ejecución de los procesos son consecuencia de una mala planificación.

Esta etapa también puede presentar diversas situaciones que impacten en la eficiencia del proceso, como la ausencia o la mala carga de registros de datos relacionados con las actividades de este.

Generalmente, la empresa utiliza personal técnico en sus sectores para realizar la carga de datos de las órdenes de trabajo. En este proceso, la carga es manual desde el lugar de trabajo. Sin embargo, existe resistencia por parte del personal a realizar estas tareas administrativas. Pardo Álvarez (2017) señala que es necesario explicar la importancia de esta labor y que se pueden utilizar diversos argumentos, tales como:

- Los registros posibilitan la trazabilidad del servicio.
- Se puede encontrar información cuando se ha producido algún tipo de incidencia.
- Permiten demostrar que se actúa según lo planificado.
- Se puede demostrar transparencia ante terceros en las actuaciones realizadas.
- Facilitan el análisis para detectar posibles mejoras.
- Cada empleado puede justificar su trabajo.
- Proporcionan evidencia solicitada por clientes y otras partes interesadas, en este caso por ejemplo el ENARGAS.

La ejecución de todas las actividades de este proceso no solo requiere planificación y capacitación, sino también herramientas como recursos para realizar el procedimiento. Cuando hablamos de herramientas, podemos citar vehículos, dispositivos para registrar datos (como teléfonos digitales), explosímetros, herramientas generales de mano, entre otros.



Pardo Álvarez (2017) menciona que no es conveniente escatimar en la adquisición de equipos, ya que los beneficios son muchos, tales como:

- Su funcionamiento será mas consistente y esto se relaciona con la calidad del servicio.
- Los trabajos no se demoraran por un menor número de averías.
- Los trabajadores estaran mas motivados si cuentan con mas y mejores equipos.
- La eficiencia mejora.

5.4.4 Actividad búsqueda sistemática: ubicación de la fuga

Las pérdidas de gas se detectan recorriendo la traza de las cañerías con un equipo especial llamado “exposímetro”. Según Camuzzi (2019), este dispositivo es “capaz de detectar y medir concentraciones de gas en la atmósfera, tanto en tareas de detección como en la localización y clasificación de fugas”.

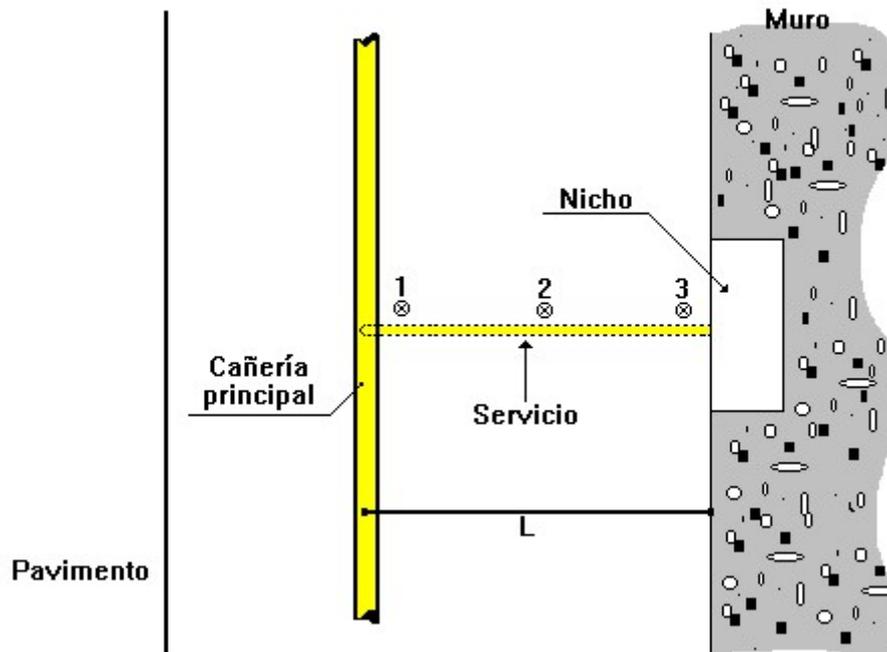
Menciona Camuzzi (2019) que conocido el lugar exacto en el cual el detector indica la máxima lectura se actuará según el caso:

a) Servicio domiciliario

Se perfora primero sobre la máxima lectura del detector, luego sobre la tapa de servicio y por último en la línea municipal. Con estas tres lecturas podemos definir claramente donde está el escape.

Figura 39

Sondeo para fuga en servicio



Fuente: Recuperado de Camuzzi (2019).

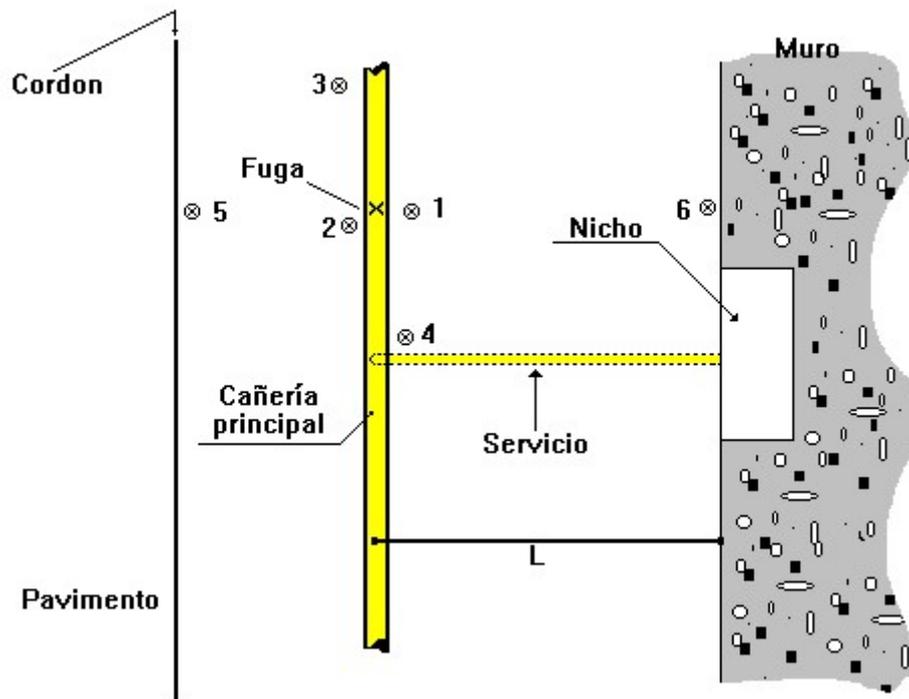
b) Sobre el sistema de distribución

Se harán tantas perforaciones y evaluaciones como sea necesario, tratando de tener suficientes valores para poder dibujar la planta del perfil de pérdidas en la superficie, esto significa conocer el máximo valor y los puntos cero de la periferia.

Generalmente a medida que nos alejamos del punto de fuga, el valor va bajando, tendiendo al valor cero.

Figura 40

Sondeo para fuga en cañería principal



Fuente: Recuperado de Camuzzi (2019).

5.4.4.1 Clasificación de la pérdida.

Basada en una evaluación de la ubicación y de la magnitud de una pérdida, dice ENARGAS (2016) que “uno de los grados de pérdidas siguientes debe ser asignado, estableciendo así la prioridad de reparación de esta”. El mismo autor detalla:

Grado 1: una pérdida que representa un riesgo probable o existente para las personas o la propiedad, y requiere reparación inmediata o acción continua hasta que las condiciones dejen de ser riesgosas. En la práctica se reparan el mismo día que se detectaron.

Grado 2: una pérdida que se define no riesgosa al momento de la detección, pero requiere una reparación programada basada en un posible riesgo futuro. En la práctica se repara en uno o dos meses.



Grado 3: una pérdida que no es riesgosa al momento de la detección y puede esperarse razonablemente que se mantenga en ese estado. Estas últimas se inspeccionan periódicamente.

Uno de los criterios que generan estas clasificaciones es la cantidad de gas que la avería deja en el lugar del hecho. De ahí es que se establece el tiempo de reparación mencionado.

5.4.5 Verificación del proceso

Pardo Álvarez (2017) menciona que la verificación es la tercera etapa de la gestión por procesos. Esta etapa se centra en recopilar datos sobre el funcionamiento de los procesos mediante un conjunto limitado de indicadores, garantizando que la selección de estos sea representativa de los objetivos generales y específicos.

Dice Pardo Álvarez (2017) que esta fase quizás sea la más crítica para poder mantener el proceso bajo control y a partir de aquí proponer mejoras que optimicen los resultados.

Los indicadores que se han elegido para representar a este proceso son:

- IND1: Fugas grado 1 por km recorridos**
- IND2: Cantidad de fugas grado 1 reparadas**
- IND3: Cantidad de fugas grado 2 reparadas
- IND4: Costo medio de reparación de fuga grado 1**
- IND5: Cumplimiento de frecuencia (ENARGAS)**
- IND6: Satisfacción en el servicio al cliente

Se elige el cálculo del rendimiento del proceso como mecanismo de control, utilizando todos los indicadores mencionados. Sin embargo, se realizó un análisis detallado posterior de los indicadores relacionados con las fugas de grado 1, con el objetivo de incorporar el “machine learning” como una ventaja.

A continuación, se realiza la asignación de importancia de cada indicador empleando la matriz pareada.



Figura 41

Importancia de indicadores del proceso

	IND1	IND2	IND3	IND4	IND5	IND6	Total	Imp.
IND1		1(IND2)	2(IND1)	1(IND4)	1(IND5)	0	2(+1)=3	0,16
IND2			1(IND2)	0	1(IND5)	1(IND6)	2(+1)=3	0,16
IND3				1(IND4)	1(IND5)	1(IND6)	0(+1)=1	0,05
IND4					1(IND5)	1(IND6)	2(+1)=3	0,16
IND5						0	4(+1)=5	0,26
IND6							3(+1)=4	0,21
							19	1

Fuente: Diseño propio.

Los valores de importancia obtenidos en este análisis nos permitirán calcular el rendimiento del proceso.

En el siguiente paso calculamos el rendimiento de los indicadores:

a) IND1: Fugas grado 1 por km recorridos

$$Rto = (\text{Valor límite del cumplimiento} / \text{Valor del indicador}) * 100$$

$$Rto = (\text{Valor año anterior} / \text{Valor año actual}) * 100$$

$$Rto = (0,99 / 0,84) * 100 = \mathbf{117,8 \%}$$

b) IND2: Cantidad de fugas grado 1 reparadas

$$Rto = (\text{Valor límite del cumplimiento} / \text{Valor del indicador}) * 100$$

$$Rto = (\text{Valor año anterior} / \text{Valor año actual}) * 100$$

$$Rto = (11778 / 10971) * 100 = 107,3\%$$

c) IND3: Cantidad de fugas grado 2 reparadas

$$Rto = (\text{Valor límite del cumplimiento} / \text{Valor del indicador}) * 100$$

$$Rto = (\text{Valor año anterior} / \text{Valor año actual}) * 100$$

$$Rto = (1069 / 1196) * 100 = 89,4\%$$

d) IND4: Costo medio de reparación de fuga grado 1

$$CFG1R = \sum \text{costo por fugas grado 1 reparadas} / \text{Recuento de fugas grado 1}$$



$$\text{Rto} = (\text{Valor límite del cumplimiento} / \text{Valor del indicador}) * 100$$

$$\text{Rto} = ((\text{Valor año anterior} * \text{inflación anual}) / \text{Valor año actual}) * 100$$

El cálculo del costo por fugas grado 1 reparadas tiene varios conceptos especificados y aceptados por las partes en el contrato firmado con terceros:

Costo por fugas grado 1 reparadas= Costo básico + Costo reparación vereda + Costo exceso por km de traslado

Estos costos se aprueban mensualmente mediante partes de trabajos.

Al no disponer de un valor real de rendimiento adoptamos:

$$\text{Rto} = 75\%$$

e) IND5: Cumplimiento de frecuencia (ENARGAS)

$$\text{Rto} = (\text{Valor límite del cumplimiento} / \text{Valor del indicador}) * 100$$

$$\text{Rto} = (\text{Cantidad de preventivos realizados} / \text{Cantidad de preventivos}) * 100$$

Se conoce que se realizaron todos los mantenimientos preventivos en tiempo y forma, entonces:

$$\text{Rto} = 100\%$$

f) IND6: Satisfacción en el servicio al cliente

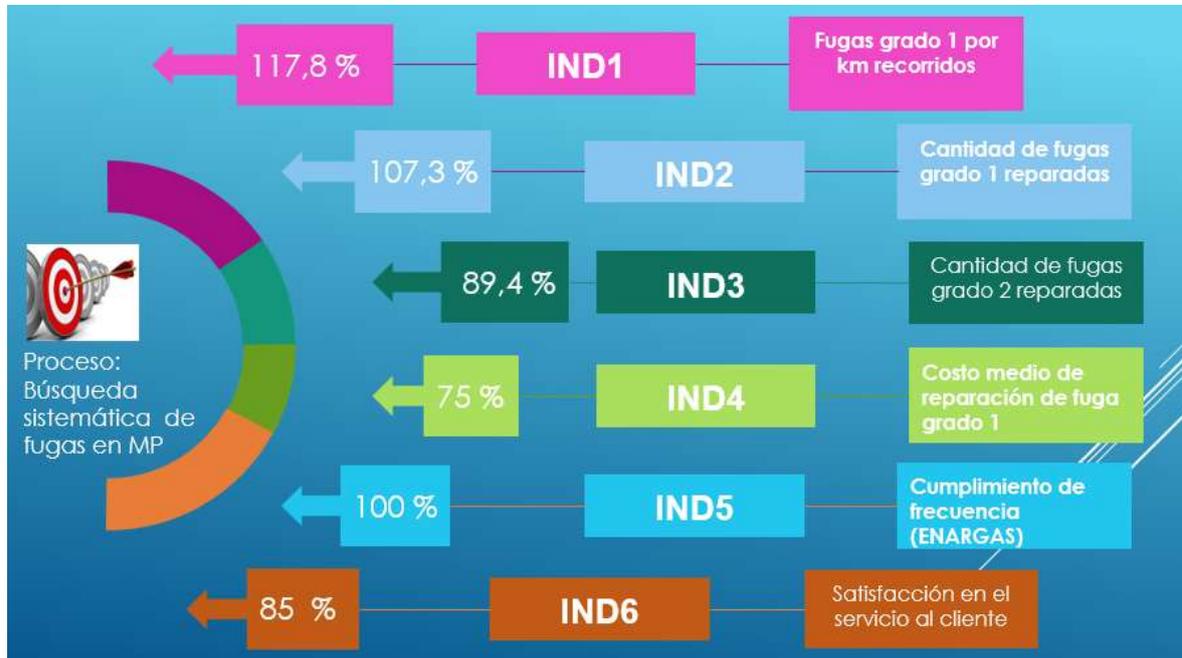
Se desconoce cómo la empresa mide este indicador y cuál es su valor actual. Para poder avanzar en el análisis de este proceso, se asignó un valor de:

$$\text{Rto} = 85 \%$$

Ahora, para calcular el rendimiento del proceso de "Búsqueda sistemática de fugas en MP", se consideró tanto el rendimiento de los indicadores como su importancia en el proceso. Por lo tanto, el rendimiento de los indicadores en su conjunto es:

Figura 42

Resumen rendimiento de indicadores



Fuente: Diseño propio.

Figura 43

Rendimiento del proceso “Búsqueda sistemática de fugas en MP”

Proceso: Búsqueda sistemática de fugas en MP				
Indicador	Nomenclatura	Importancia	Rendimiento	Imp. * Rto
IND1: Fugas grado 1 por km recorridos	IPBS001	0,16	117,8 %	18,8 %
IND2: Cantidad de fugas grado 1 reparadas	IPBS002	0,16	107,3 %	17,2 %
IND3: Cantidad de fugas grado 2 reparadas	IPBS003	0,05	89,4 %	4,5 %



IND4: Costo medio de reparación de fuga grado 1	IPBS004	0,16	75 %	12 %
IND5: Cumplimiento de frecuencia (ENARGAS)	IPBS005	0,26	100 %	26 %
IND6: Satisfacción en el servicio al cliente	IPBS006	0,21	85 %	17,8
Rendimiento del proceso		1		96,3 %

Fuente: Diseño propio.

5.4.6 Actuación para la mejora del proceso

Al analizar el rendimiento del proceso, se observó que es más que óptimo, alcanzando un 96,3%. No obstante, siempre hay margen para mejorar, y en ese sentido, se trabajó sobre las siguientes premisas:

- Predecir la cantidad de fugas grado 1 para el año siguiente.
- Implementar mecanismos para mejorar el indicador “Costo medio de reparación de fuga grado 1”.
- Tener instalaciones más seguras y confiables.
- Mejorar el rendimiento del indicador 6 “Satisfacción en el servicio al cliente”.

Con base en estos puntos se desarrolló un modelo predictivo el cual permitió conocer las cantidades de fugas grado 1 que se pueden presentar el año siguiente en los períodos que se ejecutan los mantenimientos preventivos.

5.5 Uso de “machine learning” para optimizar el proceso búsqueda sistemática de fugas media presión

Para utilizar esta tecnología en la mejora de nuestro proceso lo primero que hicimos fue definir las variables que vamos a utilizar para el desarrollo del modelo:



- Variable dependiente: cantidad de fugas grado 1 reparadas en los últimos 5 años contabilizadas por cada mes.
- Variable independiente: el tiempo expresado en meses y años.

A la hora de optar el tipo de aprendizaje automático se eligió el clásico dado que tenemos datos simples y categorías bien definidas. Además, como los datos están clasificados en números utilizamos un algoritmo supervisado.

Como el modelo predice un número, en este caso la cantidad de fugas grado 1 para el próximo año, se utilizó un algoritmo de regresión que analice la serie de tiempo que tenemos como base de datos.

Para ello nos basamos en la metodología de Box – Jenkins descrita en el marco teórico.

Como lenguaje de programación optamos por Python que es muy utilizado en ciencia de datos. El ide a través del cual escribiremos el código es Jupyter Notebook.

5.5.1 Escenario inicial

Para innovar en mejoras del proceso seleccionado, se trabajó sobre los indicadores:

- IND2: Cantidad de fugas grado 1 reparadas
- IND4: Costo medio de reparación de fuga grado 1
- IND5: Cumplimiento de frecuencia (ENARGAS)
- IND6: Satisfacción en el servicio al cliente

La base de datos obtenida nos permitió tener datos del indicador número 2 no solo del año anterior sino de 5 años atrás. Conocer con anticipación el valor futuro nos permitirá lograr eficiencia y eficacia en la planificación de recursos.

El indicador número 4, el costo medio de reparación de fuga grado 1, se utilizó para reformular su ecuación agregando una variable relacionada con el modelo de predicción:

IND4: Costo medio de reparación de fuga grado 1

$$CFG1R = \sum \text{costo por fugas grado 1 reparadas} / \text{Recuento de fugas grado 1}$$



$$Rto = (\text{Valor límite del cumplimiento} / \text{Valor del indicador}) * 100$$

$$Rto = ((\text{Valor año anterior} * \text{inflación anual}) / \text{Valor año actual}) * 100$$

$$\text{Costo por fugas grado 1 reparadas} = (\text{Costo básico} + \text{Costo reparación vereda} + \text{Costo exceso por km de traslado}) * CP$$

Donde CP es Coeficiente de predicción. Esta variable tomará dos valores:

CP = 0,8 cuando las predicciones estén dentro del margen de error propuesto

CP = 1 cuando las cantidades de fugas grado 1 pronosticadas se encuentren fuera del rango de predicción

De esta manera, se planteó reducir el costo si las predicciones resultan acertadas. Esto debe negociarse al momento de elaborar el contrato, dado que la visión es de "ganar-ganar". La empresa energética se beneficiaría de la reducción del costo por reparación, mientras que el contratista podría gestionar las reparaciones de manera más eficaz y eficiente a lo largo del año. Cabe recordar que estas empresas suelen prestar servicios a varias compañías de diferentes sectores.

El indicador número 5 también se verá beneficiado con todas estas medidas dado que se busca tener más medidas de control que respalden las operaciones en seguras y confiables.

El indicador número 6, satisfacción del cliente, también está dentro del margen de optimización dado que esta operatoria de predicciones y eficientización sin duda mejora el servicio y el sistema de prestación de este.

5.5.2 De lo general a lo particular con “machine learning”

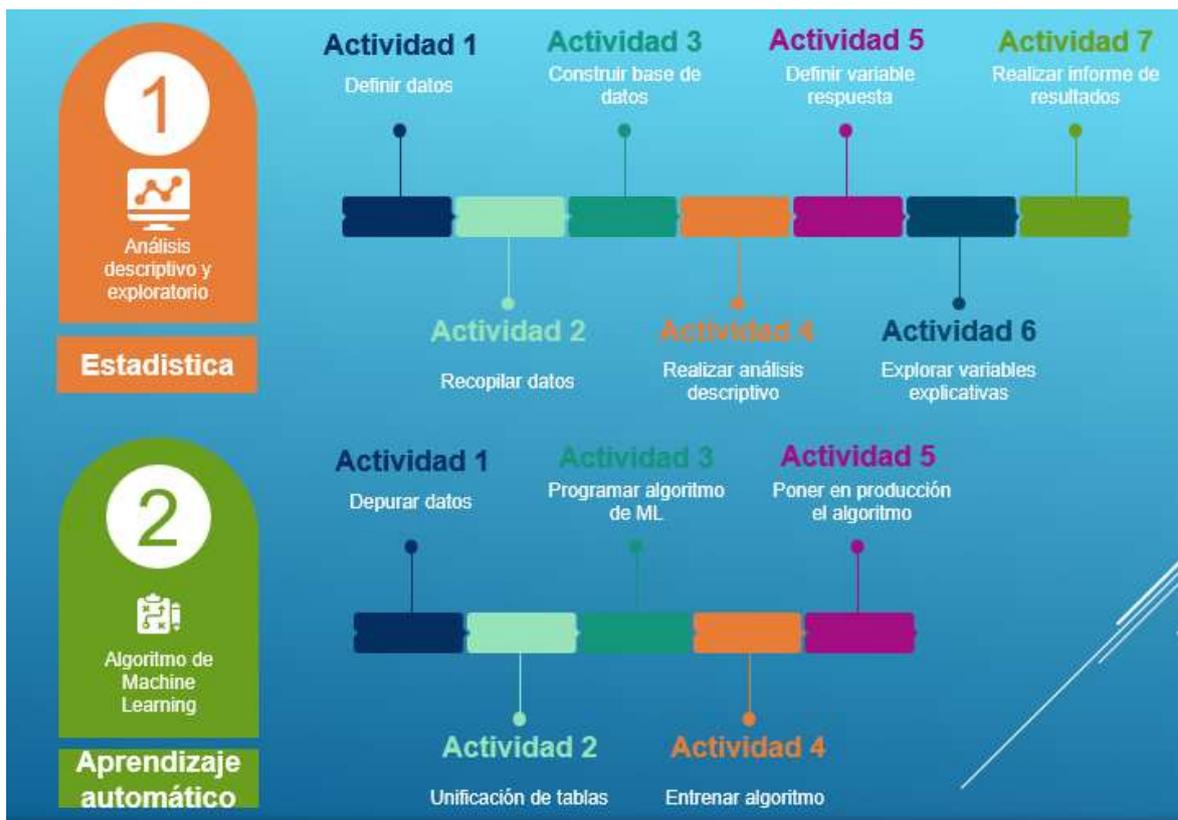
Llegamos a este punto desarrollando la gestión por procesos y destacando las ventajas de aplicar el ciclo de mejora continua al proceso seleccionado. En esta fase, aplicamos los conocimientos adquiridos en estadística de negocios, especialmente el esquema de series de tiempo basado en la metodología Box-Jenkins. Utilizamos este procedimiento para desarrollar ARIMA y SARIMA con el lenguaje Python. El código está publicado en el anexo B.

El trabajo de mantenimiento predictivo se construyó a partir del siguiente esquema de actividades, que nos ayudó a ordenar y entender cada tarea realizada. Este modelo consta de dos etapas bien definidas:

- Estadística: contiene una serie de pasos que van desde la obtención y verificación de la base de datos hasta realizar un despliegue descriptivo y exploratorio de los mismos.
- Aprendizaje automático: Mientras se armó el algoritmo se fueron depurando los datos para cada instancia. Los mismos se dividieron en entrenamiento y test para programar ARIMA – SARIMA.

Figura 44

Modelo de proyecto para mantenimiento predictivo



Fuente: Diseño propio.



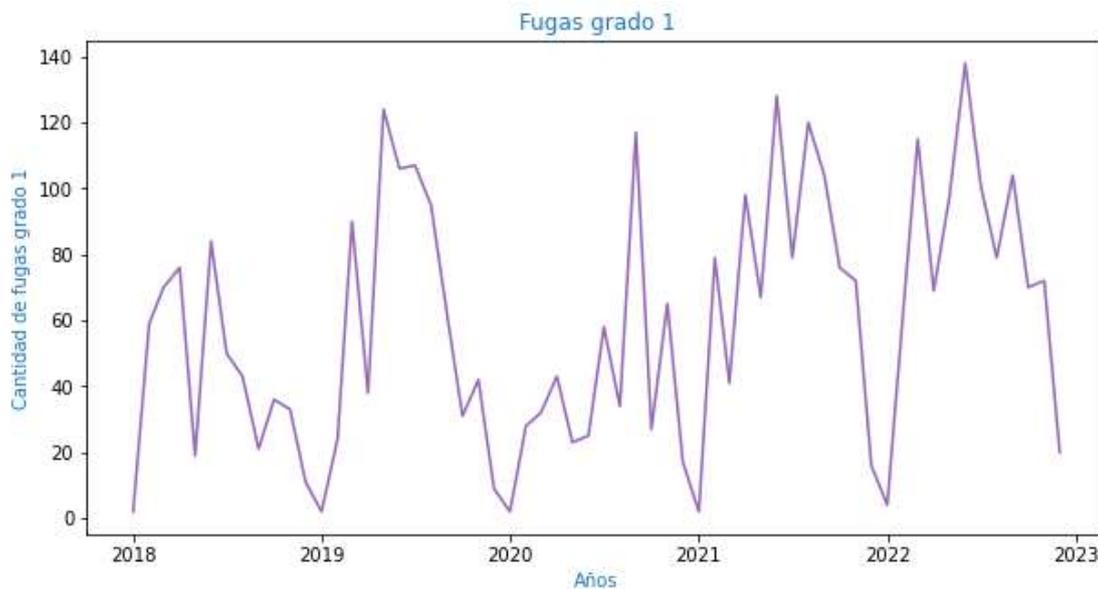
5.5.2.1 Estadística.

En concordancia con lo expuesto en la figura anterior, dentro de esta etapa, se definieron siete actividades que nos ayudaron a construir el modelo de “machine learning” para predecir la cantidad de fugas de grado 1 en el próximo año.

- 1) Definir datos: los datos con los que trabajamos corresponden a una serie de tiempo que registra las cantidades de fugas de grado 1 detectadas y reparadas por día durante un período de 5 años.
- 2) Recopilar datos: la base de datos utilizada para desarrollar este trabajo final fue solicitada al ENARGAS mediante un correo electrónico en el que se explicaron los detalles del caso. Esta base de datos no hace referencia a ninguna empresa en particular y está exenta de los procesos y mecanismos de gestión desarrollados. Se utilizó como ejemplo para concluir esta propuesta.
- 3) Construir base de datos: la evaluación de los datos suministrados por ENARGAS se realizó verificando el formato de las fechas y los valores asignados a cada registro. Se efectuó un primer análisis preliminar cronológico para asegurar que los valores no se desvíen de un rango aceptable.
- 4) Realizar análisis descriptivo: se utilizó el software Jupyter Notebook como interfaz de programación en lenguaje Python. Al comenzar el análisis descriptivo, el primer paso fue graficar los datos. Se obtuvo la siguiente gráfica:

Figura 45

Serie de tiempo de fugas grado 1



Fuente: Diseño propio.

- 5) Definir variable respuesta: la variable respuesta en nuestro trabajo final es la cantidad de fugas grado 1 detectadas y reparadas.
- 6) Explorar variables explicativas si bien en este trabajo final la variable tiempo es explicativa, también es posible analizar otras variables que no se consideraron debido a la dificultad de acceso a ellas. A continuación, se enumeran otras variables que podrían ser explicativas:
 - El tipo de cañería: acero o polietileno.
 - La antigüedad de la cañería.
 - La ubicación o localidad de la fuga.
 - Si es acero, los niveles de protección catódica y estado del revestimiento de la cañería.
 - La presencia o no de alguna interferencia.
 - La presencia o no de alguna obra de terceros.

La importancia de contar con estas variables citadas, o con algunas de ellas, radica en que ayudaría a comprender mejor el fenómeno de la ocurrencia de fugas en

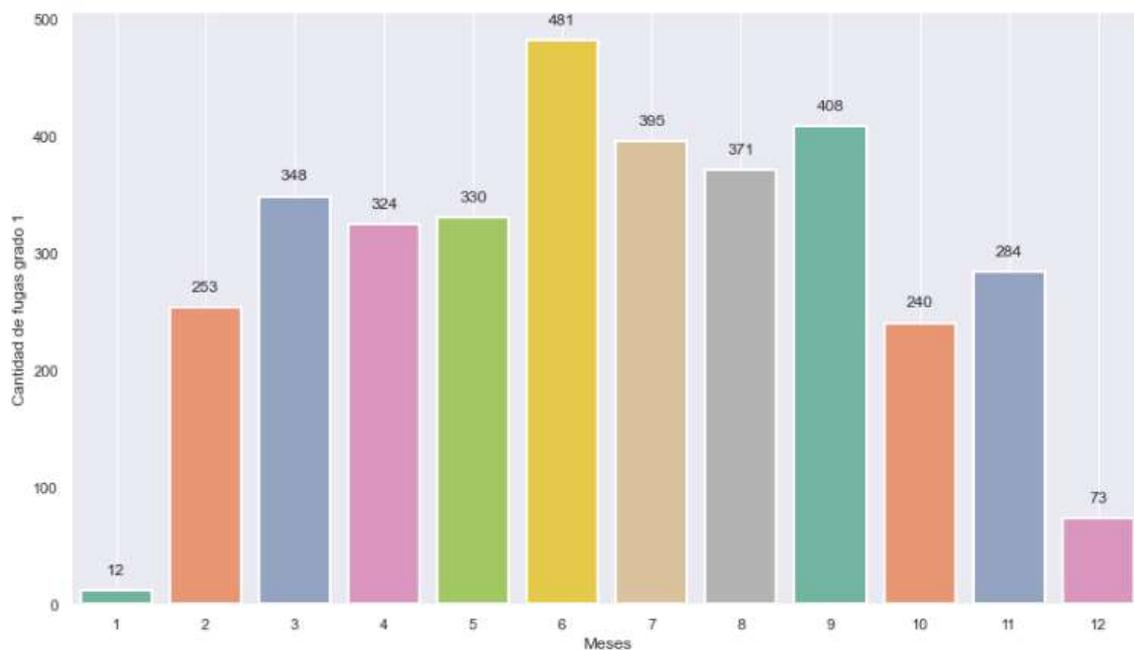


cañerías activas. En el análisis de todos los datos, se incluiría la correlación entre la fuga detectada y cualquiera de las demás variables mencionadas, proporcionando una comprensión más profunda de los factores presentes en el momento de la detección de la pérdida de gas. Trabajar con más variables que tengan un grado de correlación con la variable de respuesta permite a los algoritmos entender mejor el mecanismo de relación entre ellas y mejorar las predicciones.

- 7) Realizar informes de resultados: anteriormente, observamos la serie de tiempo en un gráfico XY. Ahora, en primer lugar, presentamos la cantidad total de fugas resumidas para cada mes a lo largo de los 12 meses.

Figura 46

Serie de tiempo resumida en cada mes

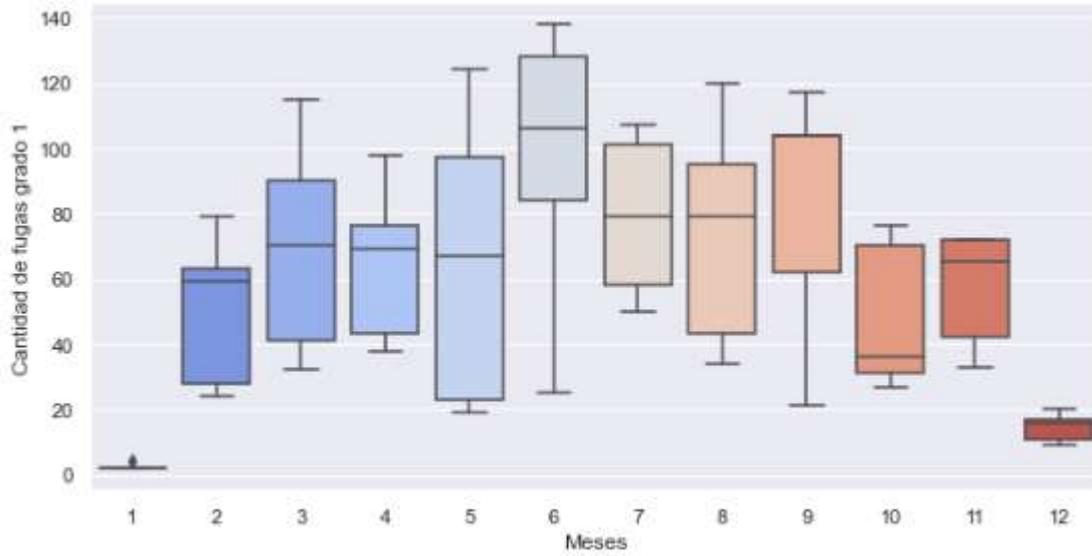


Fuente: Diseño propio.

Se apreció una mayor concentración de valores sobre los meses seis al nueve de dicho periodo. Otra forma de describir lo visto en la anterior figura es a través de un gráfico de cajas como el que se expone a continuación:

Figura 47

Serie de tiempo resumida en cada mes a través de un gráfico de cajas

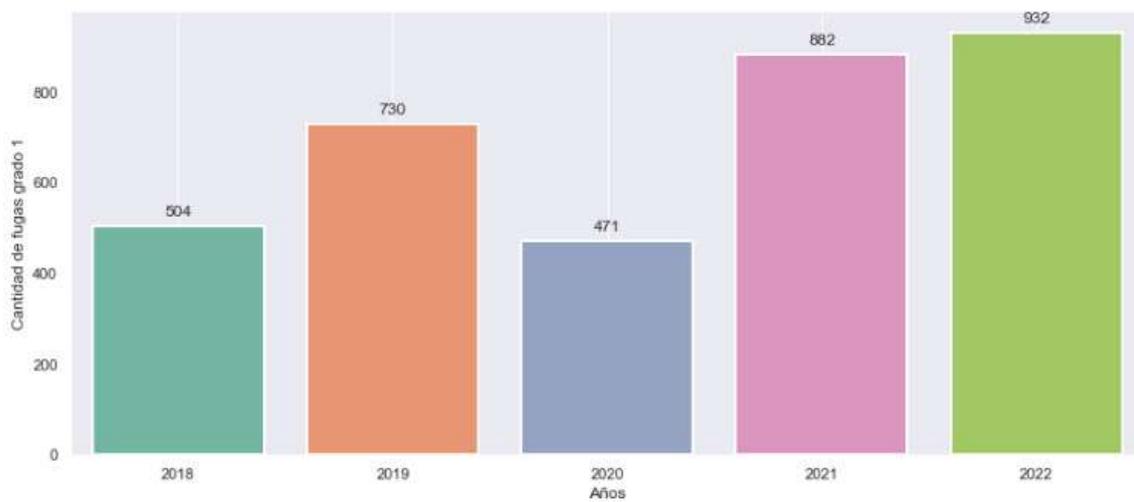


Fuente: Diseño propio.

En segundo lugar, se observó la suma total de las fugas de grado 1 por año.

Figura 48

Serie de tiempo resumida en cada mes



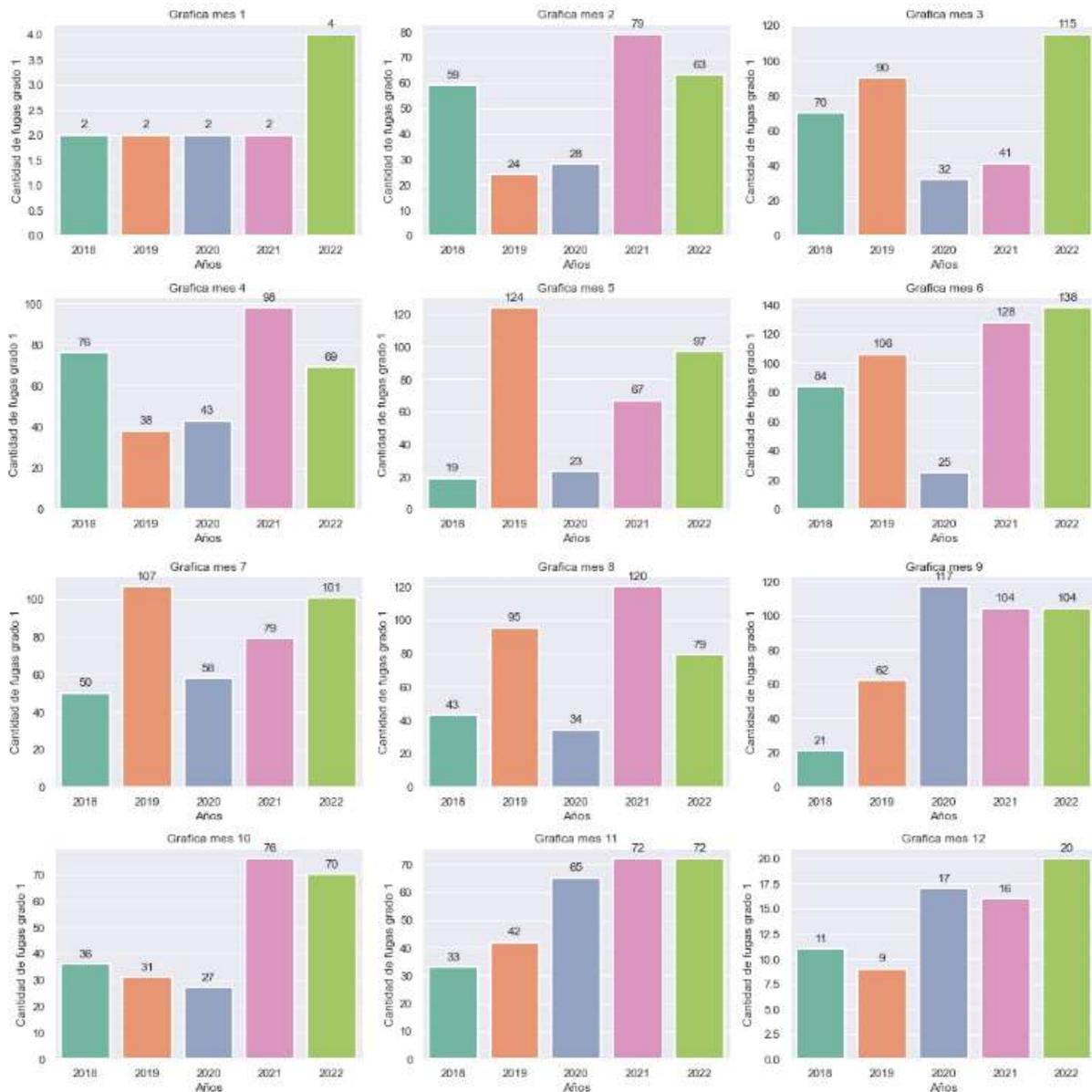
Fuente: Diseño propio.

Se observó que la cantidad de fugas detectadas y reparadas ha aumentado en los últimos años.

En tercer lugar, se muestra la evolución de las fugas de grado 1 por mes, observada a lo largo de los años.

Figura 49

Fugas grado 1 por mes evaluadas a través de los años



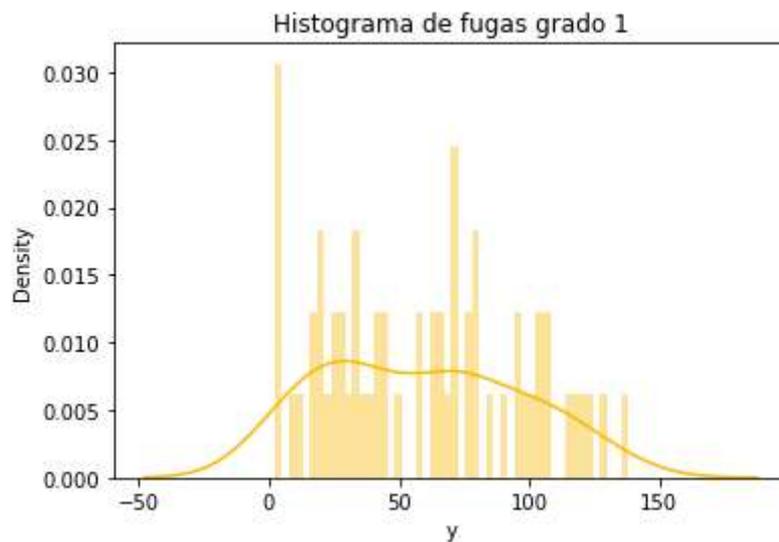
Fuente: Diseño propio.



En cuarto lugar, observamos la distribución de los valores de las fugas, analizando su densidad.

Figura 50

Histograma de densidad de fugas grado 1



Fuente: Diseño propio.

En quinto lugar, para evaluar si la serie puede ser modelada y no es simplemente ruido, utilizamos el estadístico Q de Ljung-Box. Para series mensuales, se consideran 36 datos o coeficientes de autocorrelación. Con un nivel de confianza del 95%, un valor de Q superior a 50 indicaría un rechazo del ruido blanco, es decir, la posibilidad de modelizar la serie (Chasco, 2017). Entonces:



Figura 51

Estadístico Q de Ljung-Box

	lb_stat	lb_pvalue
1	10.440449	1.232853e-03
2	20.830801	2.996739e-05
.....		
36	95.443366	2.792640e-07

Fuente: Diseño propio.

5.5.2.2 Aprendizaje automático.

En la segunda etapa del “modelo de proyecto para mantenimiento predictivo” propuesto, las primeras dos actividades son únicas para cualquier algoritmo utilizado:

- 1) **Depuración de datos:** en este punto, con los datos que tenemos, no es necesario realizar una depuración adicional dado que nuestra base de datos solo contiene dos variables.
- 2) **Unificación de tablas:** cuando se trabaja con más variables, puede haber varias que sean explicativas. En este punto es necesario unificar distintas bases de datos para analizar el grado de influencia de cada variable sobre la variable de respuesta. Sin embargo, la base de datos que utilizamos solo contiene una variable explicativa.

Las actividades 3 y 4 pueden unificarse para lograr una mayor fluidez en el desarrollo, expuesto a continuación:

- 3) **Programar y entrenar algoritmo de machine learning:** para realizar esta actividad, nos basamos en la metodología de Box-Jenkins desarrollada en Python. Esta tarea incluyó la creación del código en Python para analizar la serie de tiempo y realizar una predicción. Tras depurar la base de datos, la dividimos en dos partes para implementar el algoritmo de predicción:



- Entrenamiento del algoritmo: son los datos que el programa utiliza para construir internamente una ecuación que represente la serie de tiempo, de acuerdo con los valores que le suministramos. Para esto, se usa el 80 % de los datos disponibles en la primera parte.
- Test: son los datos que el programa utiliza para evaluar qué tan bien predice a partir de la ecuación que construyó en la etapa anterior. En esta fase, se utiliza el 20 % restante de los datos para comparar la precisión de las predicciones del programa con los valores ya conocidos. A partir de este punto, se calcula el margen de error.

Se probaron distintos modelos. El primer modelo consistió en un enfoque sugerido mediante programación, en el cual el mismo código evaluó diversas configuraciones para determinar cuál de los modelos funciona mejor entre las muchas posibilidades. El segundo y el tercer modelo se desarrollaron a partir de las observaciones realizadas en los gráficos de autocorrelación y autocorrelación parcial. A continuación, se detalla el análisis realizado.

En primer lugar, se realizó un análisis de estacionalidad de nuestra serie mediante la prueba de Dickey-Fuller. Creamos una función y calculamos los valores de la prueba para nuestros datos:

Figura 52

Test de Dickey - Fuller para la base de datos

```
Resultados de la prueba de Dickey-Fuller para columna: y
Test Statistic                -3.247383
p-value                       0.017401
No Lags Used                  3.000000
Número de observaciones utilizadas 56.000000
Critical Value (1%)          -3.552928
Critical Value (5%)          -2.914731
Critical Value (10%)         -2.595137
dtype: float64
Conclusion:====>
Rechazar la hipótesis nula
Los datos son estacionarios
```

Fuente: Diseño propio.

La conclusión se basa en evaluar el valor p con el valor crítico del 5%. Dado que $0,05 > 0,017$, esto indica que la serie es estacionaria. Al comparar el valor de la prueba con el



valor crítico correspondiente al 5%, obtenemos que $-3,2473 < -2,9147$. Ambas condiciones, como vimos a través de Zarzosa Espina (2013), indican que la serie es estacionaria. Sin embargo, esto asume el criterio del valor crítico del 5%. Al observar más detenidamente, notamos que el valor crítico del 1% arroja una conclusión diferente. Por lo tanto, avanzaremos diferenciando la serie una vez más para evaluar los nuevos parámetros.

Figura 53

Test de Dickey - Fuller para la base de datos diferenciada una vez

```
Resultados de la prueba de Dickey-Fuller para columna: y
Test Statistic          -1.265541e+01
p-value                 1.340118e-23
No Lags Used           0.000000e+00
Número de observaciones utilizadas  5.800000e+01
Critical Value (1%)    -3.548494e+00
Critical Value (5%)    -2.912837e+00
Critical Value (10%)   -2.594129e+00
dtype: float64
Conclusion:====>
Rechazar la hipótesis nula
Los datos son estacionarios
```

Fuente: Diseño propio.

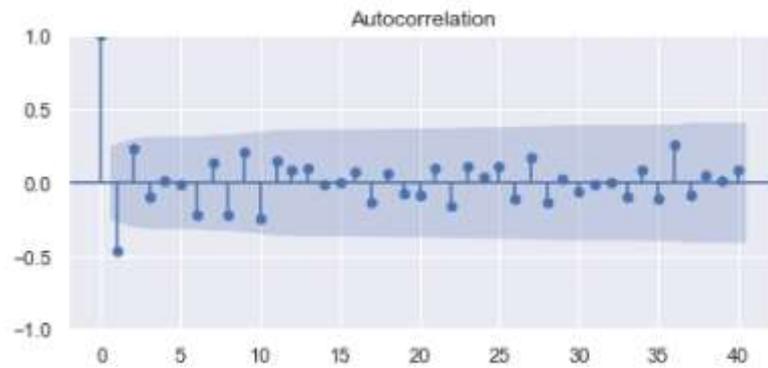
Se puede observar nuevamente que la serie es estacionaria al comparar el valor p con el valor crítico del 5%. En este caso, $0,05 > 1,340118e-23$. Al comparar el valor de la prueba con el valor crítico del 5%, obtenemos que $-12,65541 < -2,912837$. Ambas condiciones confirman que la serie es estacionaria. Además, al considerar el valor crítico del 1%, también llegamos a la misma conclusión: la serie es estacionaria, ya que $-12,65541 < -3,548494$.

En segundo lugar, identificamos nuestro modelo. Para ello, recurrimos a los gráficos de autocorrelación simple y parcial de nuestros datos diferenciados una vez y obtenemos los siguientes valores:



Figura 54

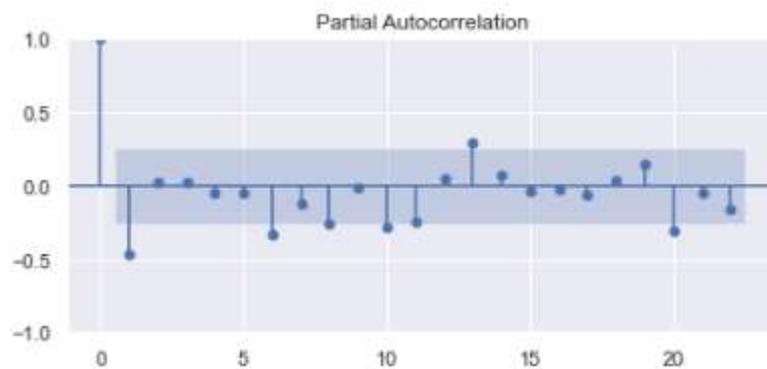
Gráfico de autocorrelación simple de la serie diferenciada una vez



Fuente: Diseño propio.

Figura 55

Gráfico de autocorrelación parcial de la serie diferenciada una vez



Fuente: Diseño propio.

En tercer lugar, realizaremos la estimación del posible modelo. Para analizar los modelos potenciales, tomamos los datos originales y los dividimos en dos partes:

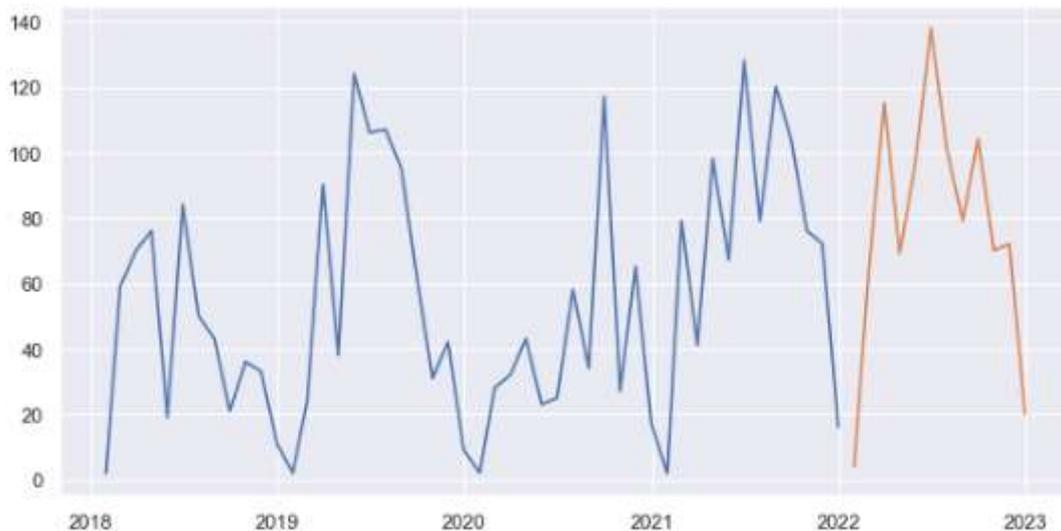
- El 80 % se utilizó para entrenar el modelo.
- El 20 % restante se empleó para pruebas de ajuste.

Los datos, ya divididos, se muestran de la siguiente manera:



Figura 56

Serie de datos dividida en entrenamiento y test



Fuente: Diseño propio.

El **primer modelo** que se analizó es el propuesto por el algoritmo de Python denominado autoarima. Esta función requiere ciertos parámetros de configuración y devuelve el mejor modelo según un algoritmo interno. En la siguiente imagen, se muestra parte de la salida.

Figura 57

Salida de autoarima

```
Best model: ARIMA(1,1,0)(1,1,0)[12]  
Total fit time: 7.992 seconds  
ARIMA(1,1,0)(1,1,0)[12]
```

Fuente: Diseño propio.

Para evaluar la eficiencia del modelo, procedemos a revisar sus métricas:



Figura 58

Métricas del modelo autoarima

```
Evaluation metric results:-  
MSE is : 910.8614509442351  
MAE is : 21.864401283643318  
RMSE is : 30.180481290798447  
MAPE is : 52.14027055265969 %  
R2 is : 0.30735895750089837
```

Fuente: Diseño propio.

El **segundo modelo** se desarrolló a partir de la evaluación de los gráficos de autocorrelación simple y parcial de la serie de datos diferenciada una vez. El modelo propuesto es el siguiente:

Figura 59

Modelo Arima - Sarima observando los gráficos de autocorrelación simple y parcial diferenciada una vez

```
modelo2 = SARIMAX(train_h, order=(1,1,2), seasonal_order=(0,1,0,12))  
modelo2_fit = modelo2.fit()  
print(modelo2_fit.summary())
```

Fuente: Diseño propio.

Las métricas obtenidas del segundo modelo son:

Figura 60

Métricas del segundo modelo

```
Evaluation metric results:-  
MSE is : 760.0965991870453  
MAE is : 19.09461184503592  
RMSE is : 27.569849458911545  
MAPE is : 35.176640776349885 %  
R2 is : 0.42200419140015977
```

Fuente: Diseño propio.



El **tercer modelo** es una modificación del segundo, basada en una nueva interpretación de los gráficos de autocorrelación mencionados. El modelo observado es el siguiente:

Figura 61

Segunda interpretación de los gráficos de autocorrelación

```
modelo3 = SARIMAX(train_h, order=(0,1,0), seasonal_order=(1,1,0,12))
modelo3_fit = modelo3.fit()
print(modelo3_fit.summary())
```

Fuente: Diseño propio.

Las métricas del tercer modelo son:

Figura 62

Métricas del tercer modelo

```
Evaluation metric results:-
MSE is : 961.8838329894994
MAE is : 20.12632190350796
RMSE is : 31.01425209463384
MAPE is : 20.397407298846673 %
R2 is : 0.26856030612094994
```

Fuente: Diseño propio.

Podemos resumir las métricas de los distintos modelos en la siguiente tabla:

Figura 63

Resumen de métricas de los modelos

	Metricas	Autoarima	Modelo2	Modelo3
0	MSE	910.86	760.09	910.88
1	MAE	21.86	19.09	20,12
2	RMSE	30.18	27.56	31,01
3	MAPE	52.14	35.17	20.39
4	R2	0.30	0.42	0.26



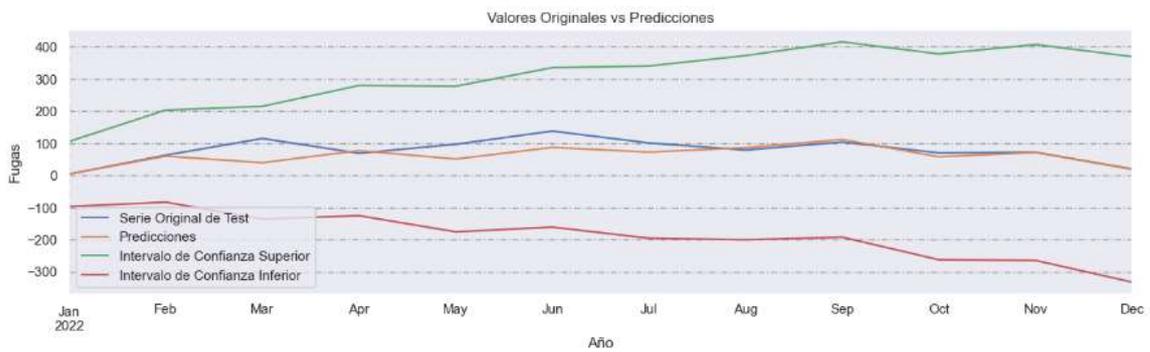
Fuente: Diseño propio.

En este trabajo final, se optó por observar el MAPE de cada modelo y se eligió el que tenía el valor más bajo. En este caso, el modelo 3 presenta un MAPE del 20%.

En el gráfico siguiente, se muestra cómo ajusta el modelo elegido:

Figura 64

Ajuste del modelo elegido



Fuente: Diseño propio.

En cuarto lugar, se realizó la evaluación del modelo propuesto. En este sentido, tenemos:

- La prueba de Durbin-Watson.
- Gráfico autocorrelación de los residuos.

La prueba de Durbin-Watson se realizó con Python y arrojó el siguiente valor:

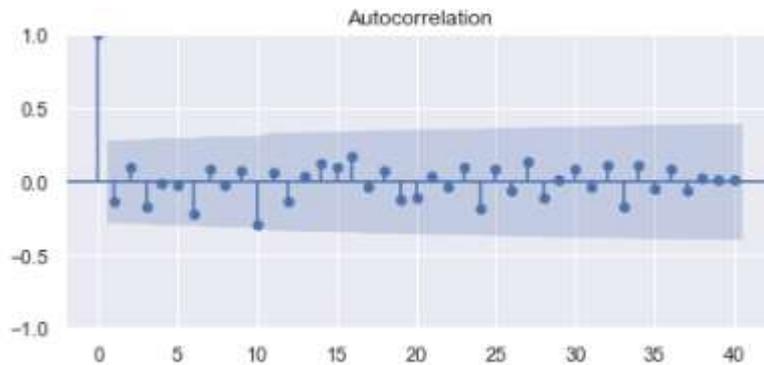
- 3.4286676311486906

Este valor indica que los residuos no están correlacionados, aunque puede encontrarse en la zona de indecisión, como se observó.

Por otro lado, el gráfico de autocorrelación de los residuos se muestra así:

Figura 65

Gráfico de autocorrelación de los residuos



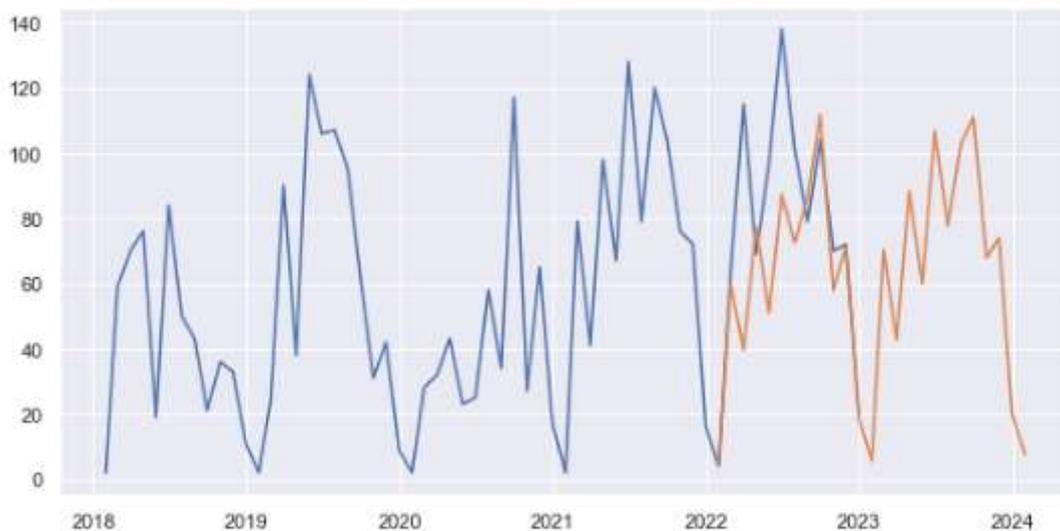
Fuente: Diseño propio.

En el gráfico, salvo el rezago de orden 10, que está en el límite de la banda de confianza, los demás rezagos no presentan un margen de autocorrelación significativo.

En quinto lugar, se observó cómo ajusta el modelo y su pronóstico. En la siguiente gráfica, se utilizó el año 2022 como etapa de prueba. La línea azul muestra los valores reales de fugas para ese período, mientras que la línea roja representa los valores de predicción en la etapa de prueba y el pronóstico para el siguiente año, en este caso, 2023.

Figura 66

Ajuste en test y pronóstico



Fuente: Diseño propio.



Los valores de pronóstico para el año 2023, que se muestran en la figura anterior, también se presentan a continuación en formato de tabla:

Figura 67

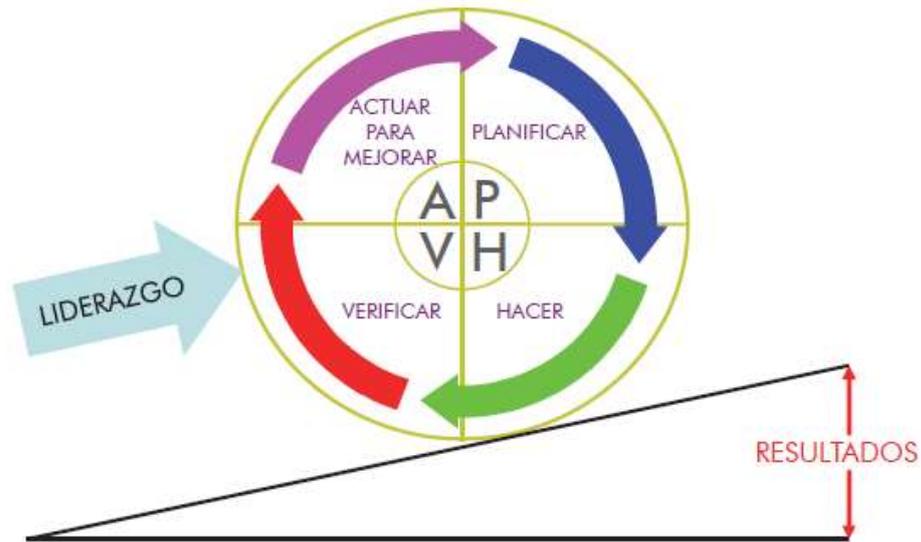
Valores de la predicción del año siguiente

Fecha	Valor de la predicción
2023-01-31	5,72
2023-02-28	70,3
2023-03-31	42,53
2023-04-30	88,33
2023-05-31	60,01
2023-06-30	106,65
2023-07-31	77,61
2023-08-31	102,78
2023-09-30	110,88
2023-10-31	67,79
2023-11-30	74,01
2023-12-31	19,96

Fuente: Diseño propio.

- 4) **Poner en producción el algoritmo:** consiste en integrar esta herramienta desarrollada en el proceso. Esto significa utilizarla en la gestión del proceso de ‘Búsqueda sistemática de fugas’.

Cuando hablamos de gestión, nos referimos al ciclo de mejora continua. Recordemos:



Fuente: Recuperado de Pardo Álvarez (2017).

Es importante llevar a cabo el ciclo de mejora continua. En este trabajo final, hemos abordado cómo se implementan las etapas de planificar, hacer, verificar y actuar para lograr mejoras. También discutimos cómo evitar limitarse a las etapas de hacer y actuar, que pueden llevar a una 'diagonal perversa'.

Al implementar el desarrollo de predicción, nos centramos en las etapas de verificar y actuar. Verificamos nuestros indicadores y realizamos controles sobre los resultados parciales o finales obtenidos. Al actuar, además de observar los resultados, cuestionamos y analizamos los métodos de trabajo actuales, incluyendo nuestro proyecto, para mejorar el desempeño del proceso.

5.6 Como trabajar el modelo a futuro

El modelo es una herramienta adicional que contribuyó a la mejora del proceso de 'búsqueda de fugas en redes de media presión'.

En primer lugar, se debe continuar con la gestión del proceso mediante el seguimiento de sus indicadores y su frecuencia de revisión, monitoreando así los objetivos del proceso.

En segundo lugar, es necesario aplicar el ciclo de mejora continua para asegurar el cumplimiento de esos objetivos, con un enfoque en la mejora continua.



En tercer lugar, después de transcurridos los doce meses, disponemos de nuevos datos sobre el desempeño y cumplimiento en la reparación de las fugas de grado 1. Esto nos lleva a observar dos puntos:

- Se dispone de nuevos datos correspondientes a un año de trabajo, que deben compararse con la predicción del modelo para analizar su desempeño.
- Estos nuevos datos se añadirán a la base de datos existente. No se recomienda extender el período de datos a más de 5 a 7 años.

En cuarto lugar, con la nueva base de datos se debe generar una nueva predicción utilizando la estructura del proyecto como base. Esta nueva predicción no implica necesariamente el uso del mismo modelo; en su lugar, se deberá realizar un análisis similar para desarrollar un modelo que podría variar debido a la incorporación del nuevo año de datos.

5.7 Ventajas para la organización

5.7.1 Mejora la planificación de los recursos

Gestionar un proceso analizando los datos que arroja su ejecución permite contar con más y mejor información para tomar mejores decisiones.

El uso del "machine learning" como herramienta de mejora continua aplicada al proceso "búsqueda sistemática de fugas" sirve para realizar un análisis de datos mediante un algoritmo y predecir situaciones futuras.

Además, incorporar la predicción de fugas de grado 1 para el año siguiente permite planificar los recursos con mayor eficiencia. Por un lado, conocer la cantidad de personal necesario para cubrir las actividades y el momento preciso en que se requiere. Por otro lado, las herramientas y repuestos necesarios para realizar las reparaciones. La compra de estos repuestos y la gestión del stock también se benefician al saber cuántos materiales se precisan. Algunos de ellos requieren importación, y este proceso puede ser complicado en la actualidad.



5.7.2 Mejora en la calidad del servicio

Las intervenciones de la empresa no solo se centran en la detección de fugas, sino también en su reparación. Incorporar tecnología al proceso promueve un sistema más seguro y confiable, impactando positivamente en la calidad del servicio. Este trabajo describe el uso de "machine learning" para mejorar un proceso específico como proyecto, y la organización puede aplicar esta experiencia en otros procesos internos.

Desde la perspectiva de la gerencia, como cliente interno, se busca ser eficaz y eficiente. Los datos recopilados de los diferentes procesos son la materia prima necesaria para alcanzar estas metas. La información generada por el uso de "machine learning" como herramienta de mejora continua contribuye a optimizar la gestión del proceso mencionado y abre nuevas oportunidades para el análisis de datos dentro de la organización. Sin duda, este trabajo traza un nuevo camino para lograr mejoras en los procesos y aumentar la satisfacción del cliente interno.

Además, contar con un proceso más eficiente contribuye a mejorar la satisfacción del cliente externo, ENARGAS, al permitir un mejor control y cumplimiento de los requisitos establecidos por el ente.

5.7.3 Reducción de costo

Durante el desarrollo de este trabajo, se planteó reducir el costo del contrato con terceros si las predicciones resultan adecuadas en tiempo y forma. Esto permitiría, al elaborar el contrato, introducir ajustes en el costo de las reparaciones con el objetivo de lograr una visión de "ganar-ganar". La empresa energética se beneficiaría al reducir el costo de las reparaciones, mientras que el contratista podría gestionar las reparaciones de manera más eficaz y eficiente a lo largo del año, de manera similar a lo mencionado en el apartado anterior. Estas empresas suelen prestar servicios a varias compañías de diferentes sectores, y esta planificación "anticipada" representa una gran ventaja.



6 Conclusiones

En la actualidad, tiene gran relevancia que las organizaciones implementen una gestión efectiva de sus procesos. Sin embargo, los esfuerzos para lograrlo dependerán del grado de madurez de cada organización.

En el presente trabajo se aborda esta temática en una empresa distribuidora de gas natural, identificando su cadena de valor. A partir de esto, se aplicó un modelo predictivo de “machine learning” para mejorar un proceso específico.

Para desarrollar lo expuesto, se revisó el marco teórico relacionado y se elaboró un código en Python para implementar el modelo mencionado, lo que permitió generar predicciones sobre la cantidad esperada de fugas de grado 1 en el sistema.

Quedó expuesto, por lo tanto, cómo los resultados obtenidos proporcionan mecanismos para gestionar el proceso a través del ciclo de mejora continua y del análisis periódico de las predicciones.

La aplicación del modelo predictivo en este proceso ha evidenciado las siguientes tres ventajas para la organización en estudio:

- a) Mejora en la planificación de recursos.
- b) Mejora en la calidad del servicio.
- c) Reducción de costos.

El desafío será expandir el número de variables explicativas en el modelo, incorporando datos relacionados con el proceso. Además, es fundamental que la organización cuente con un gobierno de datos eficaz para garantizar que la información esté disponible en tiempo y forma para realizar estos análisis.

En síntesis, este trabajo final de maestría incorpora “machine learning” para generar predicciones que aporten mejoras al proceso de “búsqueda sistemática de fugas” y mejoren su rendimiento. Se considera que, si se traslada este mecanismo de trabajo a toda la organización y se aplica la gestión por procesos, se pueden lograr mejoras significativas en el rendimiento de la cadena de valor de la empresa.



7 Bibliografía

Abadía, J. (2010). La Inteligencia Artificial. *Researchgate*.

Álvarez, J. (2013). *Configuración y usos de un mapa de procesos*. España: AENOR Internacional.

Arranz, J. y Zamora, M. (2002). Unizar. *Obtenido de*
<https://www.ciberconta.unizar.es/leccion/autocorrelacion/analisis%20de%20autocorrelacion.PDF>

Ayala, G. (2022). *Estadística básica*. Valencia: Universidad de Valencia.

Bagnato, J. (2020). *Aprende Machine Learning*. España: Leanpub.

Camuzzi. (2019). Camuzzigas. *Obtenido de* <https://www.camuzzigas.com/wp-content/uploads/2019/03/n-sm-107-03-relevamiento-de-fugas.doc>

Chang, C. (2022). Gravatar. *Obtenido de* <https://gravitar.biz/bi/machine-learning/>

Chasco, C. (2017). Tema 10 Predicción Modelos ARIMA F. *Obtenido de*
<https://www.youtube.com/watch?v=2fenJJukcQk>

Club BPM (2011). *El libro del BPM: Tecnologías, Conceptos, Enfoques Metodológicos y Estándares*. Madrid, España. Editorial: Centro de Encuentro BPM, s.l.

Cristobal, J. (2016). markdown.es. *Obtenido de* <https://markdown.es/>

De la Fuente Fernández, S. (2022). *Obtenido de*
<https://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf>

De la Fuente Fernández, S. (2023). *Obtenido de*
<https://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-clasico.pdf>

EBC Academia. (2017). *Obtenido de* <https://www.youtube.com/watch?v=P0F-1cvBR1s>



ENARGAS. (2016). *ENARGAS*. Obtenido de

<https://www.enargas.gob.ar/secciones/normativa/pdf/normas-tecnicas/NAG-100.pdf>

García Herrera, J. (2020). Inteligencia artificial en las organizaciones. *Universidad militar nueva granada*. Obtenido de

<https://repository.unimilitar.edu.co/server/api/core/bitstreams/920840bf-a391-450c-b930-78e5ae889be4/content>.

Gonzales, V., Alessandro, R., Perez, M. y Noe, R. (2023). Sistema basado en Machine Learning para la mejora de gestión de inventario en A&M Confecciones.

Universidad César Vallejo. Recuperado de
<https://hdl.handle.net/20.500.12692/134667>.

Hitpass, B. (2012). BPM Busines Process Management-Fundamentos y Conceptos de Implementación. *Santiago de Chile: BPM CENTER*.

ISA. (2021). Obtenido de <https://isa-spain.org/ia-ml-en-la-industria/>

ISOTools. (2015). ISO 9001:2008 Requisitos del Sistema de Gestión de la Calidad. Obtenido de <https://www.isotools.com.co/iso-9001-2008-requisitos-sistema-gestion-calidad/>

Kasmier, L. y Díaz Mata, A. (1996). *Estadística aplicada a la administración y la economía*. México: McGRAW-HILL.

Lind, D., Marchal, W. y Wathen, S. (2012). *Estadística aplicada a los negocios y la economía*. Mexico: McGRAW-HILL/INTERAMERICANA EDITORES.

Martínez Zabaleta, M. y Luna, R. (2023). Inteligencia empresarial y su rol en la generación de valor en los procesos de negocios. *Tendencias*, 24(1), 226-251.



- Medina, A., Rivera, D., Hernández-Nariño, A. y Rodríguez, R. (2019). Procedimiento para la gestión por procesos: métodos y herramientas de apoyo. *Ingeniare, revista chilena de ingeniería*, 328 - 342.
- Mendoza, P. y Anthony, R. (2022). *Machine learning en la mejora de la gestión de cobranza en la Empresa Externa S.A.C.* Universidad César Vallejo. Recuperado de <https://hdl.handle.net/20.500.12692/106382>
- Miranda Chinlli. (2021). *Modelización de Series Temporales modelos clásicos y SARIMA.* España: Universidad de Granada.
- Moreno, J. (2019). Predicción de energía eólica con modelos autorregresivos. *Madrid: Universidad Autónoma de Madrid.*
- Nieves D., Cervantes L., Osuna D. y Olguín C. (2024). La inteligencia artificial en las empresas y las nuevas tendencias de su desarrollo. *Gestión de las organizaciones y de las tendencias administrativas. DOI: 10.55204/pmea.73.c154*
- Pardo Álvarez, J. (2017). *Gestión por procesos y riesgo operacional.* España: AENOR Internacional.
- Parra, F. (2015). *Economía Aplicada 2.* Creative Commons.
- Pedrero, V., Reynaldos-Grandón, K., Ureta-Achurra, J. y Cortez-Pinto, E. (2021). Generalidades del Machine Learning y su aplicación en la gestión sanitaria en servicios de urgencia. *Revista médica de Chile*, 248-254.
- Peña, D. (2010). *Análisis de series temporales.* Madrid: Alianza Editorial S.A.
- Perez Grau, S. y Zabaleta de Armas, M. (2015). *Contabilidad de gestión.* Universidad Simón Bolívar.
- Quintero, S. (2023). Aprende Python. *España:* <https://aprendepython.es/>.



- Quinto, N., Villodas, A., Montero, C., Cueva, D. y Vera, S. (2021). La inteligencia artificial y la toma de decisiones gerenciales. *Revista de Investigación Valor Agregado*, 8(1), 52-69.
- Quisbert, N. (s.f.). *Modelos ARIMA*. Universidad Católica Boliviana.
- Riabiz, J. (2022). Curso de Python inicial. *Rosario*.
- Robledo, P. (2017). Diferencias entre Procesos, Procedimientos e Instrucciones de trabajo. *Obtenido de* <https://albatian.com/es/blog/diferencias-entre-procesos-procedimientos-e-instrucciones-de-trabajo/>
- Rubín, C. (2024). La IA en la administración de negocios actual. *Cuadernos Del CIMBAGE*, 1(26), 61-76.
[https://doi.org/10.56503/CIMBAGE/Vol.1/Nro.26\(2024\)/3021](https://doi.org/10.56503/CIMBAGE/Vol.1/Nro.26(2024)/3021)
- Ruiz, M. (Julio de 2015). Página Web de Mari Carmen Ruiz Abellón. *Obtenido de* https://www.dmae.upct.es/~mcr Ruiz/Telem06/Teoria/apuntes_procesos.pdf
- Sanz, J., Calvo, M., Pérez, R., Zapata, M. y Panchón, F. (2009). *Guía para una gestión basada en procesos*. España: Instituto Andaluz de Tecnología.
- Snider, N. (2020). *Machine Learning en la Dirección de Proyectos de Desarrollo de Aplicaciones en Pymes de la República Argentina*. Recuperado de https://www.researchgate.net/publication/347279728_Machine_Learning_en_la_Direccion_de_Proyectos_de Desarrallo_de Aplicaciones_en Pymes_de la Republica Argentina/citation/download
- UNAM. (2022a). Repositorio Digital de la Facultad de Ingeniería: Conceptos fundamentales de series de tiempo. *Obtenido de* <http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/363/>



A5.pdf?sequence=5&isAllowed=y#:~:text=Una%20serie%20de%20tiempo%20es,
de%20un%20fen%C3%B3meno%20o%20proceso.

UNAM. (2022b). Repositorio Digital de la Facultad de Ingeniería: Modelación ARIMA.

Obtenido de

<http://www.ptolomeo.unam.mx:8080/jspui/bitstream/132.248.52.100/363/7/A7.pdf>

UNAM. (2022c). Repositorio Digital de la Facultad de Ingeniería: Modelación clásica.

Obtenido de

<http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/363/A6.pdf?sequence=6&isAllowed=y>

Villachica Pérez, Y., Ordoñez Cuthbert, D. y Mckensy Sambola, D. (2022). *Modelo predictivo basado en Machine Learning dirigido a PYMES de venta, caso de estudio Bluefields*. Ciencia E Interculturalidad, 30(01), 139–146.

<https://doi.org/10.5377/rci.v30i01.14267>

Yajure Ramírez, C. (2022). *Uso de algoritmos de Machine Learning para analizar los datos de energía eléctrica facturada en la Ciudad de Buenos Aires durante el período 2010–2021*. Ciencia, Ingenierías y Aplicaciones, 5(2), 7–37.

Zaratiegui, J. (1999). La gestión por procesos: su papel e importancia en la empresa.

Economía industrial, 81-88.

Zarzosa Espina, P. (2013). Universidad de Valladolid on line. *Obtenido de*

<https://www.youtube.com/watch?v=14Tks1IstOE>

8 Anexos

8.1 Anexo A

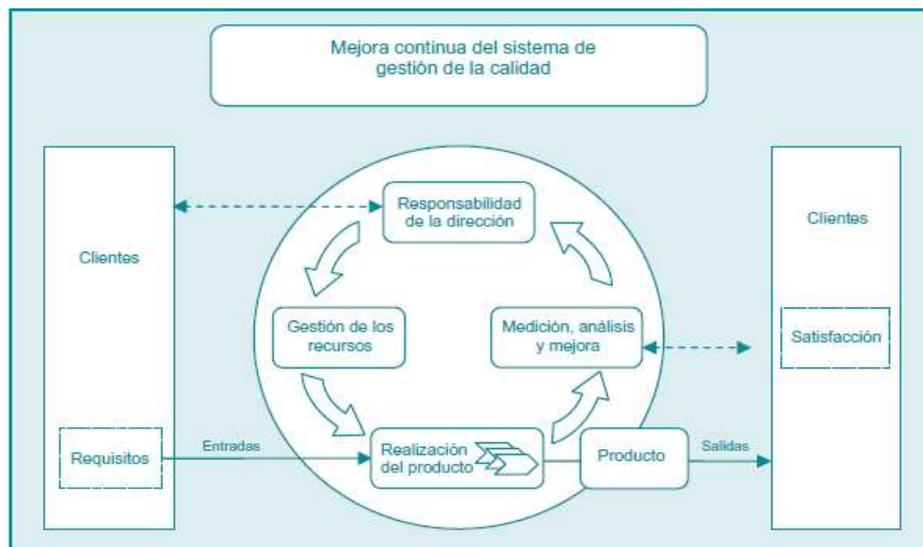
8.1.1 La Gestión basada en procesos: el enfoque ISO 9001

La norma ISO 9001:2008 “se encuentra basada en un modelo enfocado en los procesos en la que la organización debe establecer los procesos y gestionarlos de forma sistemática. Se presenta de manera gráfica el modelo en el que se enfocan los procesos utilizados en la norma ISO 9001” (ISOTools, 2015).

La figura muestra un modelo de un sistema de gestión de la calidad basado en procesos:

Figura 68

Modelo de un sistema de gestión de la calidad basado en procesos (según ISO 9001:2000)



Fuente: Recuperado de Sanz, Calvo, Pérez, Zapata, & Panchón (2009)

Según Sanz, Calvo, Pérez, Zapata, & Panchón (2009) los pasos para el establecimiento, implementación y mantenimiento de un sistema de gestión de la calidad son:

- Identificar los procesos necesarios para el sistema de gestión de la calidad y su aplicación a través de la organización.
- Determinar la secuencia e interacción de estos procesos.



- c. Determinar los criterios y métodos necesarios para asegurarse de que tanto la operación como el control de estos procesos sean eficaces.
- d. Asegurarse de la disponibilidad de recursos e información necesarios para apoyar la operación y el seguimiento de estos procesos.
- e. Realizar el seguimiento, la medición y el análisis de estos procesos.
- f. Implementar las acciones necesarias para alcanzar los resultados planificados y la mejora continua de estos procesos.

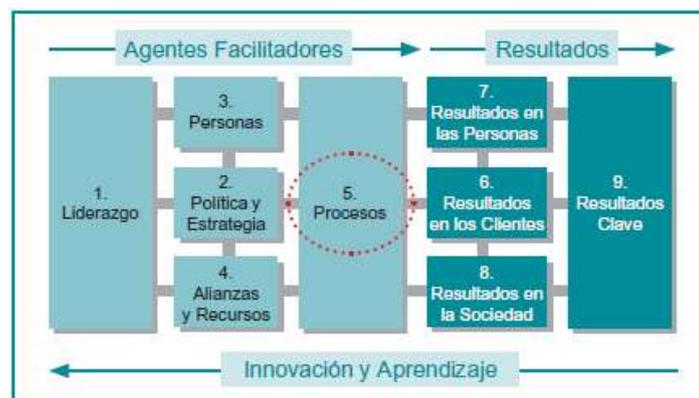
8.1.2 La Gestión basada en procesos: el modelo EFQM

La Fundación Europea para la gestión de la calidad con sede en Bruselas propone el modelo de excelencia EFQM común marco práctico y no prescriptivo para la gestión organizacional diseñado para ayudar a las organizaciones a ser más competitivas.

En la figura “se aprecian los nueve criterios que forman el modelo, agrupados en agentes facilitadores (que tratan sobre lo que la organización hace) y en resultados (que tratan de lo que la organización logra), de forma que estos últimos son consecuencia de los primeros” (Sanz, Calvo, Pérez, Zapata, & Panchón, 2009).

Figura 69

Estructura del modelo de excelencia de la EFQM



Fuente: Recuperado de Sanz, Calvo, Pérez, Zapata, & Panchón (2009).



8.2 Anexo B

8.2.1 Algoritmo desarrollado en Python

Carga de Herramientas

```
# Tratamiento de datos
```

```
import pandas as pd  
import numpy as np  
np.set_printoptions(suppress=True)  
import math as math  
from datetime import date
```

```
# Gráficos
```

```
import seaborn as sns  
import matplotlib.pyplot as plt  
import hvplot.pandas  
import statsmodels.api as sm  
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf  
import statsmodels.graphics.tsaplots as sgt  
from scipy.stats.distributions import chi2  
import statsmodels.tsa.stattools as tsa  
from statsmodels.tsa.seasonal import seasonal_decompose #libreria para descomponer la  
serie
```

```
# Modelado y Forecasting
```

```
from statsmodels.tsa.arima_process import ArmaProcess  
from statsmodels.tsa.arima_model import ARIMA  
from statsmodels.tsa.arima_model import ARMA  
from statsmodels.tsa.stattools import adfuller  
import scipy.stats  
import pylab  
from skforecast.ForecasterAutoreg import ForecasterAutoreg  
from skforecast.model_selection import grid_search_forecaster  
from skforecast.model_selection import backtesting_forecaster  
from skforecast.model_selection_statsmodels import backtesting_sarimax  
from skforecast.model_selection_statsmodels import grid_search_sarimax  
from sklearn.linear_model import Ridge  
from sklearn.preprocessing import StandardScaler  
from sklearn.pipeline import make_pipeline  
from sklearn.metrics import mean_absolute_error
```

```
# Métrica de Evaluación
```



```
from sklearn.metrics import mean_squared_error
from statsmodels.tools.eval_measures import rmse
from sklearn import metrics
from pmdarima.arima import auto_arima, ndiffs, nsdiffs
from statsmodels.tsa.statespace.sarimax import SARIMAX
from sklearn.metrics import mean_absolute_error, mean_squared_error
from statsmodels.stats.diagnostic import acorr_ljungbox, acorr_breusch_godfrey
from scipy.stats import jarque_bera, shapiro, boxcox, probplot
from datetime import timedelta
```

```
# Configuración warnings
```

```
import warnings
warnings.filterwarnings('ignore')
```

Proceso ETL

```
#Leo la base de datos
base= r'C:\Users\VPCEA35FL\Desktop\Andrés\Maestria\Proyecto trabajo
final\Resultados\Nueva base\BASE_ENARGAS.xlsx'
```

```
#Defino la hoja del excel que utilizamos
datos = pd.read_excel(base, sheet_name="fugasB")
datos.head(4)
datos.info()
```

Análisis descriptivo / Exploratorio

```
datos.y.describe()
```

```
#Grafico mi serie
fig, axx = plt.subplots(figsize=(10,5))
axx.set_title('Fugas grado 1', color='C0')
axx.set_xlabel('Años', color='C0')
axx.set_ylabel('Cantidad de fugas grado 1', color='C0')
axx.plot(datos['ds'],datos['y'],color='tab:purple')
```

```
datos['mes'] = datos['ds'].dt.month
datos['año'] = datos['ds'].dt.year
print(datos)
```

```
datos_mes=datos.y.groupby(datos['mes']).sum().to_frame()
datos_mes
```



```
datos_mes=datos_mes.reset_index()
print (datos_mes)

plt.figure(figsize=(14,8))
ax=sns.barplot(x=datos_mes.mes, y=datos_mes.y, order=None,palette='Set2' ,
linewidth=2)

# Agregar etiquetas a las barras
for p in ax.patches:
    ax.annotate(format(p.get_height(), '.2f'), (p.get_x() + p.get_width() / 2., p.get_height()),
    ha = 'center', va = 'center', xytext = (0, 10), textcoords = 'offset points')
plt.xlabel('Meses')
plt.ylabel('Cantidad de fugas grado 1')
plt.grid()

datos_bis= datos.copy()
data1=(datos_bis.groupby(by=['mes','año']).sum())
print (data1)

data1=data1.reset_index()
data1.head(10)

sns.set(rc={'figure.figsize':(10, 5)})
sns.boxplot(data=data1, x='mes', y='y',palette="coolwarm")
plt.xlabel('Meses')
plt.ylabel('Cantidad de fugas grado 1')
datos_año=datos.y.groupby(datos['año']).sum().to_frame()
datos_año
datos_año=datos_año.reset_index()
print (datos_año)

plt.figure(figsize=(14,6))
ax=sns.barplot(x=datos_año.año, y=datos_año.y, order=None, palette='Set2' , linewidth=2)

# Agregar etiquetas a las barras
for p in ax.patches:
    ax.annotate(format(p.get_height(), '.2f'), (p.get_x() + p.get_width() / 2., p.get_height()),
    ha = 'center', va = 'center', xytext = (0, 10), textcoords = 'offset points')
plt.xlabel('Años')
plt.ylabel('Cantidad de fugas grado 1')
plt.grid()

datos.index = datos['ds']
datos.head (4)

datos.info()
```



```
data=(datos.groupby(by=['mes','año']).sum())
print (data)

data=data.reset_index()
data.head(6)

enero = data ['mes']== 1
data_enero = data[enero]
data_enero.head()

# Generemos gráficos de barras para cada mes en todos los años con valore acumulados

fig = plt.figure(figsize=(15,15))

for i in range (1,13):
    mes = data['mes']==(i)

    data_mes = data[mes]

    x= data_mes.año
    y= data_mes['y']
    ax=plt.subplot(4,3,i)
    ax=sns.barplot(x=x, y=y, order=None, palette='Set2' , linewidth=2)
    for p in ax.patches:
        ax.annotate(format(p.get_height(), '.2f'), (p.get_x() + p.get_width() / 2.,
        p.get_height()),
        ha = 'center', va = 'center', xytext = (0, 10), textcoords = 'offset points')
    ax.set_xlabel('Años')
    ax.set_ylabel('Cantidad de fugas grado 1')
    ax.set_title('Grafica mes '+ str(i))
    fig.tight_layout()

datos=datos.resample('M').sum()
datos.y.head(12)

sns.distplot(datos['y'],bins=50,color='#fcba03')
plt.title('Histograma de fugas grado 1')
#la distribución es

datos=datos.reset_index()
datos.head (4)

datos.drop(columns=['mes','año'],inplace=True)
datos.columns

#Grafico mi serie
fig, axx = plt.subplots(figsize=(20, 4))
```



```
axx.set_title('Fugas grado 1', color='C0')  
axx.plot(datos['ds'], datos['y'], color='tab:purple')
```

Estadístico Q de Ljung-Box

```
acorr_ljungbox(datos.y, lags = 36)
```

1 Paso Box - Jenkins: Evaluación de estacionalidad / Test Dickey-Fuller

```
datos_bis.head()  
datos.set_index("ds", inplace=True)  
datos.head(4)
```

```
# Definimos la función para el test de Dickey – Fuller
```

```
def Prueba_Dickey_Fuller(series , column_name):  
    print(f'Resultados de la prueba de Dickey-Fuller para columna: {column_name}')  
    dftest = adfuller(series, autolag='AIC')  
    dfoutput = pd.Series(dftest[0:4], index=['Test Statistic', 'p-value', 'No Lags Used', 'Número  
de observaciones utilizadas'])  
    for key, value in dftest[4].items():  
        dfoutput['Critical Value (%s)%key'] = value  
    print(dfoutput)  
    if dftest[1] <= 0.05:  
        print("Conclusion:====>")  
        print("Rechazar la hipótesis nula")  
        print("Los datos son estacionarios")  
    else:  
        print("Conclusion:====>")  
        print("No se puede rechazar la hipótesis nula")  
        print("Los datos no son estacionarios")
```

```
Prueba_Dickey_Fuller(datos["y"], "y")
```

```
# Definimos la función para graficar el test de Dickey - Fuller
```

```
def test_stationarity(timeseries):  
  
    #Rolling statistics  
    rolmean_r = timeseries.rolling(7).mean()  
    rolstd_r = timeseries.rolling(7).std()  
  
    #Plot rolling statistics:  
    fig = plt.figure(figsize=(12, 6))  
    orig = plt.plot(timeseries, color='blue', label='Original')  
    mean = plt.plot(rolmean_r, color='red', label='Rolling Mean')  
    std = plt.plot(rolstd_r, color='black', label='Rolling Std')  
    plt.legend(loc='best')
```



```
plt.title('Rolling Mean & Standard Deviation')
plt.show()

test_stationarity(datos.y)

#descompongo la serie original en tendencia, estacionalidad y residuos
descomposicionR = seasonal_decompose(datos.y, model = "additive")
descomposicionR.plot()
plt.show()

# Diferencio la serie una vez
Ddatos= datos.copy()
Ddatos['y']= datos['y'].diff()
Ddatos.head(4)

# Remuevo el primer registro
Ddatos.dropna(inplace=True)
Ddatos.head(4)
Ddatos=Ddatos.reset_index()
Ddatos.head (4)

# Grafico la serie con diferenciación simple
fig, axx = plt.subplots(figsize=(20, 4))
axx.set_title('Fugas grado 1', color='C0')
axx.plot (Ddatos ['ds'],Ddatos ['y'] ,color = 'tab:purple')

# Bandas De Confianza
BDC = 2 * math.sqrt(1/60)
print ('Bandas De Confianza: +/-', BDC)
Ddatos.set_index("ds", inplace=True)

# Gráfico autocorrelación con logaritmo y diferenciación simple

fig, ax = plt.subplots(figsize=(7, 3))
plot_acf(Ddatos.y, ax=ax, lags=40)
plt.show()

# Gráfico autocorrelación parcial con logaritmo y diferenciación simple

fig, ax = plt.subplots(figsize=(7, 3))
plot_pacf(Ddatos.y, ax=ax, lags=22)
plt.show()
Prueba_Dickey_Fuller(Ddatos["y"],"y")

test_stationarity(Ddatos.y)
```

2 Paso Box - Jenkins: Identificación



Gráfico autocorrelación con diferenciación simple

```
fig, ax = plt.subplots(figsize=(7, 3))  
plot_acf(Ddatos.y, ax=ax, lags=40)  
plt.show()
```

Gráfico autocorrelación parcial con diferenciación simple

```
fig, ax = plt.subplots(figsize=(7, 3))  
plot_pacf(Ddatos.y, ax=ax, lags=22)  
plt.show()
```

3 Paso Box - Jenkins: Estimación de posible modelo

```
# Armo mi ARIMA - SARIMA  
# El set de datos empieza el 2018-01-31 y termina el 2022-12-31.  
# Se dividen los datos en 2 conjuntos, uno de entrenamiento, y otro de test."
```

```
datos.head(12)  
datos.info()
```

```
# Separación Train/Test  
# Seleccionamos el primer 80% para train
```

```
train_h = datos.iloc[:round(len(datos)*.8)]
```

```
# Seleccionamos el 20% restante para test
```

```
test_h = datos.iloc[round(len(datos)*.8):]
```

```
# Corroboramos que los límites se hayan establecido según lo planificado
```

```
print("Datos para entrenamiento")  
print(train_h)  
print("-----")  
print("Datos para test")  
print(test_h)  
plt.plot(train_h.y)  
plt.plot(test_h.y)
```

```
# Función para graficar las predicciones. Toma como parámetros:  
# serie_test: la porción de la serie con los datos de test  
# serie_pred: las predicciones hechas por el modelo  
# conf_int: el intervalo de confianza de las predicciones
```



```
def graficar_pred(serie_test, serie_pred, conf_int, titulo="Valores Originales vs Predicciones", xlabel="Año", ylabel="Fugas"):
    fig_x = plt.figure(figsize=(12,3), dpi=100)
    axes_x = fig_x.add_axes([0.0, 0.0, 1, 1])

    # Graficamos los datos tests y las predicciones
    serie_test.plot(label="Serie Original de Test")
    serie_pred.plot(label="Predicciones")

# Pasamos dataframe con los intervalos de confianza a series de #Pandas y graficamos
conf_int.iloc[:,1].plot(label="Intervalo de Confianza Superior")
conf_int.iloc[:,0].plot(label="Intervalo de Confianza Inferior")

axes_x.set_xlabel(xlabel)
axes_x.set_ylabel(ylabel)
axes_x.set_title(titulo)
axes_x.legend(loc=0)
axes_x.grid(True, color='0.6', dashes=(5,2,1,2))

# Función para calcular las métricas de las predicciones

def evaluar_metricas(y_true, y_pred):

    def mean_absolute_percentage_error(y_true, y_pred):
        y_true, y_pred = np.array(y_true), np.array(y_pred)
        return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
    print('Evaluation metric results:-')
    print(f'MSE is : {metrics.mean_squared_error(y_true, y_pred)}')
    print(f'MAE is : {metrics.mean_absolute_error(y_true, y_pred)}')
    print(f'RMSE is : {np.sqrt(metrics.mean_squared_error(y_true, y_pred))}')
    print(f'MAPE is : {mean_absolute_percentage_error(y_true, y_pred)} %')
    print(f'R2 is : {metrics.r2_score(y_true, y_pred)}',end='\n\n')

# Vamos a observar el primer modelo propuesto en forma automática y
# veremos cómo predice

modelo_auto=auto_arima(train_h,start_p=0,d=1,start_q=0,
                        max_p=4,max_d=2,max_q=4, start_P=0,
                        D=1, start_Q=0, max_P=2,max_D=1,
                        max_Q=2, m=12, seasonal=True,
                        error_action='warn',trace=True,
                        supress_warnings=True,stepwise=True,
                        random_state=20,n_fits=50)
print(modelo_auto)

# MODELO 1: surge de auto_arima
```



```
modelo_optimo = SARIMAX(train_h, order=(1,1,0), seasonal_order=(1,1,0,12))
modelo_optimo_fit = modelo_optimo.fit()
print(modelo_optimo_fit.summary())
fig_res_1h = modelo_optimo_fit.plot_diagnostics(figsize=(12,10))

# Utilizando el modelo de auto_arima realizamos la predicción
# correspondiente al periodo de test

predicciones1h = modelo_optimo_fit.get_forecast(len(test_h))
predicciones1h.predicted_mean.info()
# predicciones1h.predicted_mean es una serie de pandas; One-dimensional
# (Pandas DataFrame; Two-dimensional)
# convierto predicciones1h.predicted_mean a dataframe

predicciones1h.predicted_mean = predicciones1h.predicted_mean.to_frame()
predicciones1h.predicted_mean.info()

# Graficamos la porción de test de la serie original junto con las
# predicciones y los intervalos de confianza

graficar_pred(test_h.y, predicciones1h.predicted_mean.predicted_mean,
predicciones1h.conf_int())

test_h.info()

predicciones1h.predicted_mean.predicted_mean

test_h.y

evaluar_metricas(test_h.y, predicciones1h.predicted_mean.predicted_mean)

Test_data={'Metricas':['MSE','MAE','RMSE','MAPE','R2']}
Test_data= pd.DataFrame(data=Test_data)
Test_data

Test_data=Test_data.assign(Autoarima=['910.86','21.86','30.18','52.14','0.30'])
Test_data

plt.plot(datos.y)
plt.plot(predicciones1h.predicted_mean.predicted_mean)

# MODELO 2: Surge de ver los gráficos de autocorrelación de la serie
# diferenciada una vez

modelo2 = SARIMAX(train_h, order=(1,1,2), seasonal_order=(0,1,0,12))
modelo2_fit = modelo2.fit()
print(modelo2_fit.summary())
```



```
fig_res_1h = modelo2_fit.plot_diagnostics(figsize=(12,10))

# Utilizando el modelo realizamos la prediccion correspondiente al
# periodo de test

predicciones2h = modelo2_fit.get_forecast(len(test_h))
predicciones2h.predicted_mean.info()

# predicciones2h.predicted_mean es una serie de pandas; One-dimensional
#(Pandas DataFrame; Two-dimensional)
# convierto predicciones1h.predicted_mean a dataframe

predicciones2h.predicted_mean = predicciones2h.predicted_mean.to_frame()
predicciones2h.predicted_mean.info()
# Graficamos la porción de test de la serie original junto con las
# predicciones y los intervalos de confianza

graficar_pred(test_h.y,predicciones2h.predicted_mean.predicted_mean,
predicciones2h.conf_int())

test_h.info()

predicciones2h.predicted_mean.predicted_mean

test_h.y

evaluar_metricas(test_h.y, predicciones2h.predicted_mean.predicted_mean)

Test_data=Test_data.assign(Modelo2=['760.09','19.09','27.56','35.17','0.42'])
Test_data

plt.plot(datos.y)

plt.plot(predicciones2h.predicted_mean.predicted_mean)

# MODELO 3: similar al modelo 2, con ajustes de parámetros

modelo3 = SARIMAX(train_h, order=(0,1,0), seasonal_order=(1,1,0,12))
modelo3_fit = modelo3.fit()

print(modelo3_fit.summary())

fig_res_1h = modelo3_fit.plot_diagnostics(figsize=(12,10))

# Utilizando el modelo realizamos la predicción correspondiente al
# periodo de test
```



```
predicciones2h = modelo3_fit.get_forecast(len(test_h))
predicciones2h.predicted_mean.info()

# predicciones2h.predicted_mean es una serie de pandas; One-dimensional #(Pandas
DataFrame; Two-dimensional)
# convierto predicciones1h.predicted_mean a dataframe

predicciones2h.predicted_mean = predicciones2h.predicted_mean.to_frame()
predicciones2h.predicted_mean.info()

# Graficamos la porción de test de la serie original junto con las predicciones y los
intervalos de confianza

graficar_pred(test_h.y,predicciones2h.predicted_mean.predicted_mean,
predicciones2h.conf_int())

test_h.info()

predicciones2h.predicted_mean.predicted_mean
test_h.y

evaluar_metricas(test_h.y, predicciones2h.predicted_mean.predicted_mean)

Test_data= Test_data.assign(Modelo3=['910.88','20,12','31,01','20.39','0.26'])
Test_data

plt.plot(datos.y)

plt.plot(predicciones2h.predicted_mean.predicted_mean)

4 Paso Box - Jenkins: Evaluación

from statsmodels.stats.stattools import durbin_watson

# Realizar la prueba de Durbin-Watson

durbin_watson(modelo3_fit.resid)

# Gráfico autocorrelación de los residuos

fig, ax = plt.subplots(figsize=(7, 3))
plot_acf(modelo_optimo_fit.resid, ax=ax, lags=40)
plt.show()

# Gráfico de línea de errores residuales
```



```
residuals = pd.DataFrame(modelo3_fit.resid)  
residuals.plot(figsize = (16,5));  
plt.show();
```

5 Paso Box - Jenkins: Pronóstico

```
predicciones3h = modelo3_fit.get_forecast(steps=25, signal_only=False)  
predicciones3h
```

```
predicciones3h.predicted_mean.info()
```

```
# predicciones3h.predicted_mean es una serie de pandas; One-dimensional  
# (Pandas DataFrame; Two-dimensional)  
# convierto predicciones1h.predicted_mean a dataframe
```

```
predicciones3h.predicted_mean = predicciones3h.predicted_mean.to_frame()
```

```
predicciones3h.predicted_mean.info()
```

```
predicciones3h.predicted_mean
```

```
plt.plot(datos.y)
```

```
plt.plot(predicciones3h.predicted_mean)
```

```
predicciones3h.predicted_mean
```