
DATA SCIENCE: HERRAMIENTAS DE *MACHINE LEARNING* PARA IMPLEMENTAR MANTENIMIENTO PREDICTIVO EN UNA PLANTA PROPULSORA

♦ RESUMEN ♦

El *data science* (ciencia de datos), forma parte fundamental del análisis, control estadístico y de la aplicación de la industria 4.0. Es así, como la investigación realizada consistió en implementar un analizador de bases de datos en tiempo real que permitiera utilizar los datos capturados por diversos sensores en un sistema de propulsión. A partir de lo anterior y aplicando algoritmos de *machine learning* como una herramienta de esta ciencia, se contribuyó a la implementación del mantenimiento predictivo que permite por un lado, determinar cuándo es necesario realizar las tareas de mantenimiento y, por otro, reducir costos al detectar potenciales fallas de manera temprana.



EDUARDO CARRASCO VIDAL

Teniente 1°. Ingeniero Civil Industrial, Licenciado en Ciencias de la Ingeniería. (ecarrascov@armada.cl)

Data science, machine learning, turbina a gas, mantenimiento predictivo.

De acuerdo con la *British Standard* 3811:1974,¹ el mantenimiento involucra a todas las acciones necesarias para sostener o recuperar un sistema, equipo o componente a un estándar aceptable en el cual pueda cumplir las funciones para el cual fue concebido.

A lo largo del tiempo, desde la Revolución Industrial en el siglo XIX, la forma más simple de mantenimiento fue reparar o reemplazar el equipo cuando se tenía un problema, actividad que, en su mayoría, era realizada por los propios operarios. Pero la industrialización y el avance tecnológico crearon máquinas más complejas, debiendo contar con personal especializado para su reparación, creando los primeros departamentos de mantenimiento, centrados principalmente en solucionar fallas que se presentaran con acciones básicamente correctivas, definiéndose con ello el mantenimiento correctivo.

A partir del término de la Segunda Guerra Mundial, la complejidad y el avance tecnológico, derivado del conflicto, llevaron al desarrollo de conceptos como la confiabilidad, referido al desempeño de un equipo, sistema o arma, bajo cualquier tipo de escenario, lo cual demandaba un mayor control en el cuidado de los activos.

En línea con lo anterior, fueron la aviación y la industria naval, los primeros desarrolladores de planes de mantenimiento que permitieran asegurar esa confiabilidad tan necesaria en un posible conflicto (contexto de Guerra Fría), haciendo énfasis en lograr un diseño robusto, a prueba de fallos y con un adecuado programa de tareas por ejecutar durante su ciclo de vida, lo cual fue definido como el tiempo total en el cual el equipo aseguraba un desempeño óptimo de las tareas para las cuales fue creado, implementando así el mantenimiento preventivo.

La formación de mejores ingenieros de diseño, derivó en un mejor análisis de las fallas, tanto de las que han ocurrido

basándose en el registro de la información, como de las potenciales por ocurrir; por lo cual, ya en la década de los 70, se desarrolló el Mantenimiento Basado en la Confiabilidad (RCM²) como un estilo de gestión, basado en el estudio acabado de los equipos, análisis de los modos de falla y en la aplicación de técnicas de estadística, que permitieran perfeccionar los planes de mantenimiento.

A lo largo del tiempo, y con la implementación de nuevas herramientas computacionales que hicieron posible desarrollar, de mejor manera, la gestión de las actividades de mantenimiento, se cambió el enfoque desde una perspectiva reactiva hacia una perspectiva proactiva, pasando desde un foco centrado en acciones de reemplazo o reparación hacia actividades predictivas, utilizando la trazabilidad de la información disponible como una herramienta de asesoramiento en tiempo real para apoyar en la toma de decisiones, desarrollando con ello el mantenimiento predictivo basándose a partir de datos.

Este enfoque de mantenimiento, más dinámico, plantea una forma de supervisión que permite predecir potenciales degradaciones en el sistema y, con ello, planificar el momento en el cual es necesario realizar una actividad de mantenimiento, posibilitando además programar y minimizar las interrupciones. Además de disminuir el alto costo al realizar actividades de mantenimiento correctivo y prevenir la indisponibilidad del sistema que trae consigo el aplicar tal tarea.

Actualmente, para su aplicación, existen diversos *software* que hacen posible analizar grandes cantidades de datos (*big data*) y, de manera particular, en la Armada de Chile existen potenciales herramientas que permitirían generar bases de datos como son: el IPMS como un *software* de monitoreo de la condición de la maquinaria en un Patrullero de Zona Marítima (OPV) capturando a través de sensores, los diferentes estados del sistema de propulsión o también el MCAS, como

1. British Standard (1984). British Standard Glossary of Maintenance Management Terms in Terotechnology. British Standard Institution.
2. RCM: Reliability Centered Maintenance.

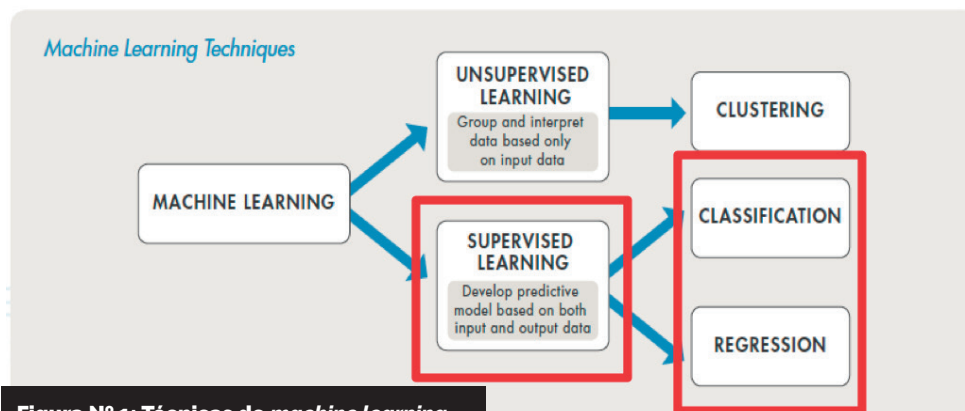


Figura N° 1: Técnicas de machine learning

monitoreo de la condición de la maquinaria en una fragata tipo 23; ambos sistemas permiten capturar información a través de sensores y, efectuando una modificación computacional, posibilitarían almacenar los datos necesarios para implementar esta tipología de mantenimiento.

Machine learning

Como se mencionó anteriormente, el gran volumen de datos (*big data*) generados por el almacenamiento de la información proveniente de sensores produce un problema generalizado en su procesamiento. En respuesta a esto, el *data science*³ cuenta con una herramienta que permite simplificar esta operación y realizar un análisis en tiempo real de la misma, generando sus propios algoritmos; este instrumento se denomina *machine learning*.

El *machine learning* es un mecanismo que permite a los computadores aprender a hacer lo que para el humano resulta natural: aprender de la experiencia. Este mecanismo utiliza algoritmos (a través de métodos computacionales)

para obtener información analizando datos sin la necesidad de contar con una determinada ecuación como modelo, es decir, los datos se procesan en una caja negra (*blackbox*) que convierte un estímulo (datos de entrada) en una respuesta (datos procesados).⁴

El *machine learning* se presenta como una ayuda esencial para el mantenimiento predictivo, permitiendo la recolección, limpieza, procesamiento y análisis a través de dos tipos de técnicas: la primera es aprendizaje supervisado, que entrena a un modelo de variables de entrada y salida conocidos que permitan predecir futuras salidas y, el segundo, es aprendizaje no supervisado que busca patrones de

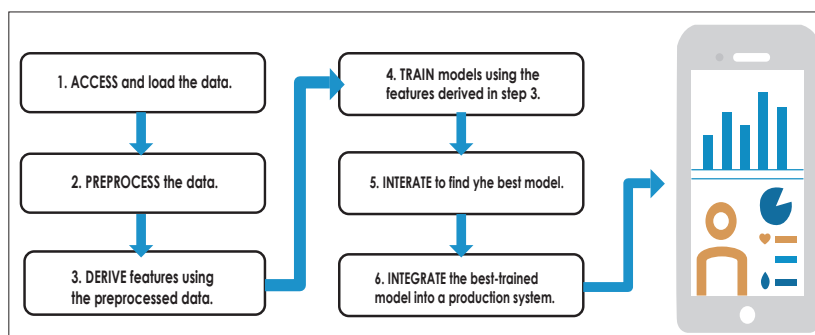


Figura N° 2: Flujo de Trabajo

reconocimiento de estructuras sólo basada en los datos de entrada.

3. Ciencia de datos: Unifica conocimientos de estadística, análisis de datos, métodos científicos, algoritmos y más, para extraer información relevante de distintas formas de datos basada en el procesamiento de grandes cantidades (Big Data) y permitir su implementación en forma diversificada.

4. Una de las características principales es que los algoritmos se actualizan a medida que se ingresa mayor información, por lo cual, el algoritmo se hace más exacto con el tiempo, lo cual es ideal para ser aplicado en sistemas de monitoreo en tiempo real.

A partir de lo anterior, para aplicar el *machine learning* en la resolución de un problema de datos, se plantea un flujo de trabajo basado en seis pasos:

Adquisición y carga de datos: Es el proceso de captura a través de sensores de los datos de operación de un sistema y su almacenamiento en distintos *data-set* como información estructurada SQL⁵ y procesable.

Preprocesamiento de los datos: Obteniendo características e identificando valores atípicos (*Outliers*) que se encuentren por fuera del resto de los datos, y con ello decidir cuáles son relevantes para la resolución del problema. Parte del preprocesamiento de los datos es dividir los datos en dos partes. Una base de datos se utilizará para entrenamiento (*Train set*) y la otra base de datos se empleará para verificar el adecuado ajuste del algoritmo (*Test Set*),⁶ esto permitirá utilizar datos que no fueron incluidos en el modelo procesado y con ello reconocer como reacciona el algoritmo ante datos desconocidos.

Obtención de características: Obtener características de los datos procesados es la actividad más importante del flujo de trabajo, lo cual también es conocido como ingeniería de datos, lo que permite transformar data cruda en información que puede ser utilizada en un algoritmo de *machine learning*.

Construcción y entrenamiento del modelo: Cuando se construye un modelo es recomendable comenzar por un algoritmo simple, debido a que será más rápida su obtención e interpretación. Un ejemplo de esto es, si se utiliza en un modelo con etiquetas, o de clasificación, como: *standing*, *sitting*, *walking*, *running* y *dancing*, para conocer cómo se ajusta a las variables de respuesta, se puede realizar una matriz de confusión que permite comparar las clasificaciones realizadas por el modelo

TRUE CLASS	Sitting	>99%		<1%		
	Standing	<1%	99%	<1%		
	Walking		<1%	>99%	<1%	
	Running			1%	93%	5%
	Dancing		<1%	<1%	40%	59%
		Sitting	Standing	Walking	Running	Dancing
		PREDICTED CLASS				

Figura N° 3: Matriz de confusión

con las creadas de forma particular, de manera de obtener su ajuste, como se muestra en la figura 3.

Cada punto de la matriz simboliza que el algoritmo de clasificación al ingresar variables desde sensores que tengan características de *sitting*, tenga un 99% de probabilidad de clasificarse en dicha etiqueta, en tanto que, al ingresar un estímulo que tenga las características de *dancing*, sólo exista un 59% de probabilidad de ser clasificado en esa categoría; lo anterior permite comparar el nivel de eficiencia de los distintos modelos a través de su matriz.

Realización de iteraciones para encontrar el mejor modelo: Del análisis obtenido de la matriz de confusión, se puede obtener la real efectividad del modelo utilizado, esto se logra comparando los resultados con las matrices obtenidas por otros modelos.

Integración del mejor modelo en un sistema de producción: Se debe generar una herramienta que permita al usuario obtener conclusiones y ayudar en la toma de decisiones. Este tipo de plataforma puede ser representada en un *dashboard* o una interfaz adecuada para cada nivel técnico del

5. SQL: Structured Query Language, sistema diseñado para administrar y recuperar información de sistemas de gestión de bases de datos relacionales.

6. Generalmente, se utiliza un 80% del total de la base de datos para generar el algoritmo (*blackbox*), para luego verificar su efectividad simulando el 20% restante.

operador, ya sea en una aplicación a través de *smartphone* o en un panel de operación del sistema de monitoreo en tiempo real.

Por último, cabe destacar que, dentro de las técnicas de aprendizaje supervisado, existen modelos de clasificación (respuestas discretas) y modelos de regresión (respuestas continuas), de lo cual se puede especificar lo siguiente: si la data posee etiquetas, categorizaciones o puede ser separada en grupos específicos, se deben utilizar algoritmos de clasificación; por otra parte, si se trabaja con un rango de datos, si la naturaleza de la respuesta es un número real como la temperatura de operación o el tiempo hasta que se produzca una falla, se deben utilizar algoritmos de regresión.

Aplicación de machine learning en una turbina GE LM2500

Se ha seleccionado para su análisis una planta propulsora caracterizada como una combinación de dos motor diésel eléctricos *Siemens* (4,5 MW cada uno) con una turbina a gas *General Electric*

LM2500 (20MW), lo que se traduce en una propulsión CODLAG tal como se utiliza en las fragatas F-125 de la *Deutsche Marine*. El componente principal que se analizará en esta investigación es la turbina que, mecánicamente, maneja las dos hélices de paso controlable (CPP), por intermedio de una caja de engranajes de conexión cruzada.

Esta máquina térmica rotativa de combustión interna utiliza como medio de trabajo el gas y tiene un ciclo termodinámico completo que incluye un fluido de entrada como aire y un fluido de

Nº	Variable	Símbolo	Unidad
1	Lever Position	lp	(.)
2	Ship Speed	v	knot
3	Gas Turbine Shaft Torque	GT_T	kN m
4	Gas Turbine Rate of Revolutions	GT_n	r/min
5	Gas Generator Rate of Revolutions	GG_n	r/min
6	Starboard Propeller Torque	T_s	kN
7	Port Propeller Torque	T_p	kN
8	HP Turbine Exit Temperature	T_{48}	°C
9	GT Compressor Inlet Air Temperature	T_1	°C
10	GT Compressor Outlet Air Temperature	T_2	°C
11	HP Turbine Exit Pressure	P_{48}	bar
12	GT Compressor Inlet Air Pressure	P_1	bar
13	GT Compressor Outlet Air Pressure	P_2	bar
14	GT Exhaust Gas Pressure	P_{exh}	bar
15	Turbine Injection Control	TIC	%
16	Fuel Flow	n_f	kg/s

Figura N° 5: Resumen de parámetros de funcionamiento

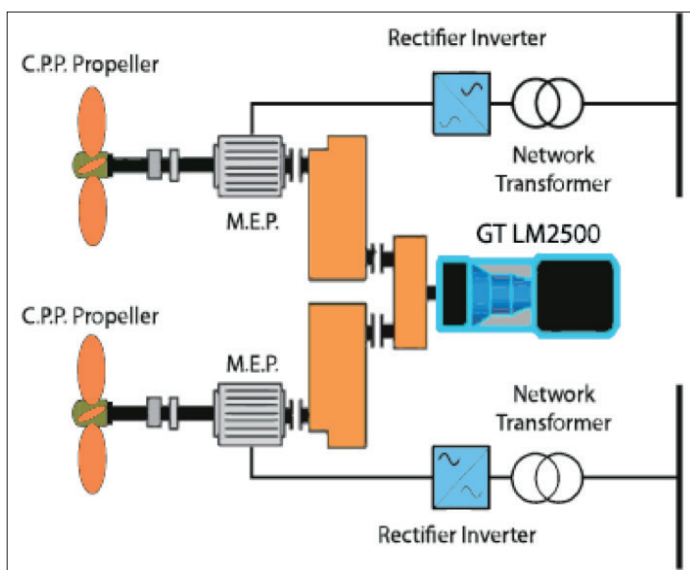


Figura N° 4: Diagrama componentes principales propulsión CODLAG

egreso como gas derivado de la combustión, convirtiendo la energía calórica, contenida en el combustible en trabajo mecánico rotacional a través de su eje.

Lo anterior, es controlado en modo local a través del telégrafo (*Telegraph* o *lever*) que, de acuerdo a una posición determinada, genera la demanda de velocidad requerida para gobernar el buque.

En la figura N° 5 se muestra un resumen de los parámetros involucrados en el proceso de utilización de una turbina a gas, consistentes en 16 datos.

El sistema de monitoreo de la condición de la maquinaria de la fragata almacena de manera automática, estos parámetros de operación en una base de datos relacional, sin embargo, el acceso a los datos de operación de la turbina en la *Deutsche Marine* se encuentran regulados como información clasificada, por lo cual se decidió utilizar una base de datos provista por la *University of California Irvine* (UCI) en su repositorio de *machine learning*, la cual contiene una simulación de los 16 parámetros de una turbina de similares características, durante un año de operación (11.934 mediciones).

Además, esta base de datos incluye dos variables adicionales: *Compressor Degradation Coefficient* (kM_C) y *Gas Turbine Degradation Coefficient* (kM_t), ambos representan la degradación sobre las horas de servicio remanentes de ambos componentes. En resumen, se tiene una base de datos de 18 variables, consistente en 16 simuladas (datos de operación), más 2 calculadas (coeficientes de degradación).

De acuerdo a lo descrito anteriormente, utilizando la base de datos del repositorio (UCI) y aplicando el flujo de trabajo descrito en la presente investigación, se aplicaron herramientas de *data science* con técnica de aprendizaje supervisado (respuesta conocida) en un modelo de clasificación, para crear un sistema de análisis de la turbina a gas que permitiera monitorear en tiempo real sus horas de servicio remanentes de acuerdo a un determinado nivel de confiabilidad.

Modelo de clasificación

Preliminarmente, se deben establecer criterios para crear etiquetas al conjunto de datos (necesarios para

un modelo de clasificación). Estos se establecieron basándose en el rango permisible de los respectivos coeficientes de degradación provistos en la base de datos, utilizando las siguientes etiquetas: Operación normal, tomar precaución o reparación urgente.⁷ Estas etiquetas han sido designadas de manera aleatoria para que el operador del sistema de monitoreo automatizado tenga una alerta del comportamiento de la turbina, con lo cual se obtiene el detalle que se muestra en la figura N° 6.

	Etiqueta	Cantidad
1	Operación normal	5049
2	Tomar precaución	4590
3	Reparación urgente	2295
	Total	11934

Figura N° 6: Resumen de cantidad de etiquetas clasificadas

El objetivo es observar de mejor manera la real incidencia de la posición del telégrafo en función de las variables anteriormente descritas, se designan colores (Rojo:

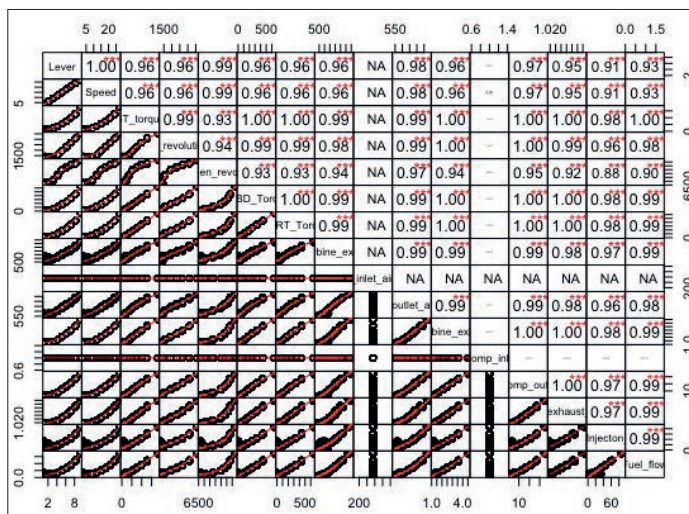


Figura N° 7: Comparación de variables en función del telégrafo

7. Cabe destacar que los rangos de aceptación de cada criterio se encuentran especificados en la Tesis realizada por el autor de este artículo.

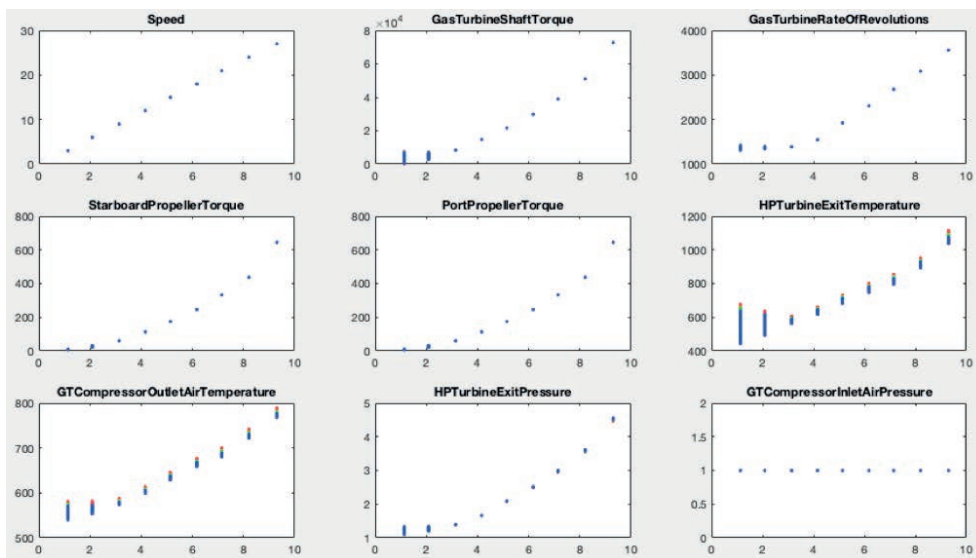


Figura N° 8: Matriz de correlación de las 16 variables

reparación urgente, verde: tomar precaución, azul: operación normal), para representar cada una de las etiquetas en el gráfico de la figura N° 7.

De acuerdo con lo anterior, es posible determinar que los parámetros de temperatura de salida de la turbina (HP Turbine exit temperature) y temperatura de salida del compresor (GT Compressor outlet air temperature) tienen mayor incidencia en la pérdida de eficiencia.

De igual manera, se elaboró una matriz de correlación para determinar el nivel de relación de cada variable y de cuáles variables no tenían mayor incidencia en la resolución del problema. (Figura N° 8).

Como se observa en la matriz de correlación, los parámetros de presión y temperatura de entrada (inlet air, inlet pressure) son siempre 1 bar y 15 °C (simulación), por lo cual no tienen un nivel de significancia para la resolución del problema y pueden ser descartados de la simulación.

Realizado el análisis preliminar, se efectúa una partición de la totalidad de los datos en dos grandes grupos. El primero de ellos (*Training set*) correspondiente al 80% para efectuar el entrenamiento de la base de datos y, el 20% restante (*Test Set*), se utilizará como variable de comprobación del algoritmo seleccionado como modelo de predicción, quedando los resultados como se muestra en la figura N° 9.

	Training set (80%)	Cantidad	Total
1	Operación normal	3.997	9.547
2	Tomar precaución	3.682	
3	Reparación urgente	1.868	
	Test set (20%)		Cantidad
1	Operación normal	1.052	2.387
2	Tomar precaución	908	
3	Reparación urgente	427	
	Total	11.934	11.934

Figura N° 9: Resumen de cantidad de etiquetas clasificadas

Definido lo anterior, y establecidos los grupos aleatorios, se realizó la simulación del grupo de entrenamiento (*Training set*), utilizando las 16 variables continuas, considerando como respuesta la variable discreta equivalente a la etiqueta de clasificación (Operación normal, tomar

precaución, reparación urgente), de acuerdo al siguiente resultado:⁸

K-Nearest Neighbor (KNN): Se utilizó un modelo de KNN Weighted obteniendo un porcentaje de 94,8% de eficiencia, con la matriz de confusión de la figura N° 10.

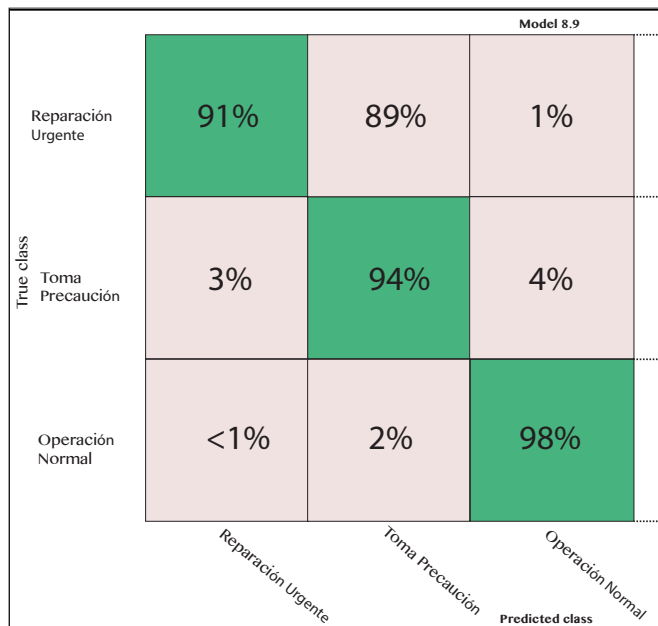


Figura N° 10: Resumen de cantidad de etiquetas clasificadas

A partir de los datos expuestos, y considerando los grados de respuesta y confiabilidad de los modelos, se decidió elegir como algoritmo de predicción el KNN *Weighted* para crear la *blackbox*.

Efectuado lo anterior, se realizó lo siguiente: en primer lugar se tiene el detalle del 20% que no se utilizó para crear el algoritmo (*Test set*), equivalente a las etiquetas de la figura 11. En segundo lugar, utilizando la *blackbox*, generada por el algoritmo a través de Matlab, se aplicó una función de predicción a esta base de datos (*Test set*)

para crear una nueva base de datos llamada *trained set* (detallada en la figura 11). Cabe destacar que esto se realizó para verificar la efectividad del algoritmo utilizando la base de datos desconocida para el modelo que representaría la entrada de nuevos datos derivados de la operación de la turbina a gas, para luego ser comparada con una nueva base de datos (*Trained Set*), obtenida tras aplicar esta función como datos entrenados, obteniendo los resultados que se muestran en la figura N° 11.

A partir de los resultados expuestos en la figura 13, se puede señalar que si se implementa esta herramienta de predicción a la turbina, al ingresar nueva data al sistema a través de sus sensores de operación, existe un 99,668% que el sistema de monitoreo encienda una alerta (Tomar precaución), cuando existan condiciones que disminuyan el nivel de confiabilidad o que demanden tomar precaución del equipo, permitiendo al

	Etiqueta	Cantidad Test Set	Cantidad Trained Set	Eficacia
1	Operación normal	1052	1053	99.905%
2	Tomar precaución	908	905	99.668%
3	Reparación urgente	427	429	99.533%
	Total	2387	2387	99.702%

Figura 11: Resumen de resultados tras aplicar el modelo

8. Objeto presentar en forma resumida los resultados, sólo se describe el mejor algoritmo utilizado para la simulación de un total de 10 algoritmos.

usuario tomar medidas paleativas para aumentar la eficiencia o prevenir una falla mayor.

Cabe destacar que el algoritmo aprende de la experiencia, por lo cual, el nivel de eficiencia en su predicción irá en aumento a medida que se ingrese más información al sistema (operando el equipo).

Conclusiones y comentarios finales

En virtud de lo demostrado en esta investigación, se puede señalar que la aplicación de este tipo de tecnología en la industria del mantenimiento beneficia no sólo a la reducción de costos, si no que, al efectuar una predicción avanzada al sistema, permite una adecuada planificación y con ello aumentar el nivel de disponibilidad de los equipos (reducción del tiempo de parada), factor preponderante en las operaciones navales.

Como fue expuesto, las herramientas de *machine learning* demostraron un alto

grado de eficacia en la determinación de los estados (etiquetas), así como su capacidad para trabajar con grandes cantidades de datos, constituyéndose como una aplicación fundamental en la implementación de una estrategia de mantenimiento predictivo.

Dentro de las ventajas de la utilización de este tipo de herramientas, destacan que no es necesario conocer al detalle los algoritmos matemáticos involucrados, siendo una herramienta de fácil comprensión y de grandes prestaciones para cualquier tipo de usuario u organización.

Finalmente, cabe destacar que la transformación digital es entendida como la integración de la tecnología en todas las áreas de una organización, esta tendencia no solo implica el replanteamiento de la gestión actual, sino que representa una oportunidad para simplificar procesos, automatizando el trabajo de las personas a través de herramientas computacionales.



BIBLIOGRAFÍA

1. Atamuradov et al. (2017) *Prognostics and Health Management for Maintenance Practitioners, Review Implementation and Tools*. International Journal of Prognostics and Health Management.
2. Bastos et al. (2014) *Application of Data Mining in a Maintenance System for Failure Prediction*. Taylor and Francis Group of Safety, Reliability and Risk Analysis: London, UK.
3. Carrasco, E. [Alid P.G.] (2019). *Herramientas de Machine Learning para el análisis del mantenimiento predictivo en una planta de propulsión naval*. Tesis para optar al Título de Ingeniero Civil Industrial. Universidad Andrés Bello, Concepción.
4. Coble and Hines (2011) *Applying the General Path Model to Estimation of Remaining Useful Life*. International Journal of Prognostics and Health Management, University of Tennessee: Knoxville, USA.
5. Coraddu, A. et al (2014) *Machine Learning Approaches for improving condition-based maintenance of naval propulsion plants*. IMechE, Journal of Engineering for the Maritime Environment. Vol 230 (I).
6. Dragomir et al (2009) *Review of prognostic problem in condition-based maintenance*. IFAC and in collaboration with the IEEE Control Systems Society.
7. Espino, Carlos (2017) *Análisis Predictivo: técnicas y modelos utilizados y aplicaciones del mismo*. Tesis para optar al grado académico de Ingeniero Informático. Universitat Oberta de Catalunya, España.
8. Fernández J. (2009) *Ciclo Brayton -Turbinas a Gas*. Universidad Tecnológica Nacional. Editorial: UTN; Mendoza, Argentina.
9. Hederra F. (2018) *Información mediante Network Centric Warfare en la Armada. ¿Cómo y Cuando?*. Revista de Marina N° 967, pp. 24-30. ISSN 0034-8511.
10. MATLAB (2018) *Statistics and Machine Learning Toolbox - White Paper*. The MathWorks Inc.