



UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA
FACULTAD DE CIENCIAS DE LA INGENIERÍA
CARRERA DE PETRÓLEOS

TEMA:

**“FACTIBILIDAD DEL USO DE INTELIGENCIA ARTIFICIAL EN LA
ETAPA DE PRODUCCION APLICADA A LA INDUSTRIA
HIDROCARBURIFERA EN POZOS DEL ECUADOR “**

TRABAJO DE INTEGRACIÓN CURRICULAR

AUTOR:

ARTURO ISMAEL ROSAS RAMIREZ

TUTOR:

ING. EDISON BRITO AVILA, MSC.

LA LIBERTAD, ECUADOR

2025

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA

FACULTAD DE CIENCIAS DE LA INGENIERÍA

CARRERA DE PETRÓLEOS

TEMA:

**FACTIBILIDAD DEL USO DE INTELIGENCIA ARTIFICIAL EN
LA ETAPA DE PRODUCCION APLICADA A LA INDUSTRIA
HIDROCARBURIFERA EN POZOS DEL ECUADOR**

TRABAJO DE INTEGRACIÓN CURRICULAR

AUTOR:

ARTURO ISMAEL ROSAS RAMIREZ

TUTOR:

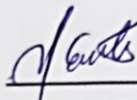
ING. EDISON BRITO AVILA, MSC.

LA LIBERTAD – ECUADOR

2025

TRIBUNAL DE SUSTENTACIÓN

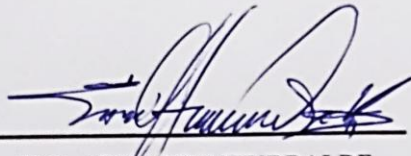
TRIBUNAL DE SUSTENTACIÓN



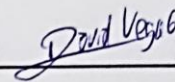
ING. MARLELIS GUTIERREZ H.
PhD.
DIRECTOR DE CARRERA



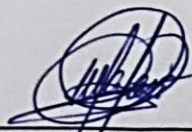
ING. EDISON BRITO AVILA. MSc.
DOCENTE TUTOR



ING. ARMANDO ITURRALDE
KURE MSc.
DOCENTE ESPECIALISTA



ING. DAVID VEGA G.
SECRETARIO DEL TRIBUNAL



ING. CARLOS MALAVE CARRERA
DOCENTE COORDINADOR UIC

DEDICATORIA

Quisiera dedicar este trabajo a todas las personas investigadoras de esta maravillosa industria que día a día realizan nuevos conocimientos para el progreso del mundo, este es mi aporte personal y humano.

Dedico esto a mis amados padres, buscando siempre abrir caminos que favorezcan mi desarrollo personal y profesional, quienes me han enseñado que no importa un título sino la persona que decides ser y formarte.

A mis hermanas quienes son mi amparo en todo momento, con sus risas, ocurrencias y aventuras son mi fortaleza en momentos difíciles son todo lo que tengo.

A mis abuelitos, esto es para ustedes donde sea que estén, ya que no pude mostrarle esto a ellos han sido mi primera familia, preocupándose sinceramente por mí.

A mis mascotas quienes abrieron mi lado humano y los quise mucho, Alex, Lily, Moka gracias por ayudarme en noches y ser parte de mi felicidad.

A todas las personas que en un momento me dieron de su amor más sincero, agradezco por las emociones y experiencias desde el momento en que nos conocimos Sofia.

Infinitas gracias, Arturo Ismael Rosas Ramírez

CERTIFICADO DE ANTIPLAGIO

En calidad de tutor del trabajo de investigación para titulación del tema **FACTIBILIDAD DEL USO DE INTELIGENCIA ARTIFICIAL EN LA ETAPA DE PRODUCCION APLICADA A LA INDUSTRIA HIDROCARBURIFERA EN POZOS DEL ECUADOR** elaborado por el estudiante **ARTURO ISMARL ROSAS RAMIREZ** egresado de la carrera de Ingeniería **PETROLEOS** de la Facultad de Ciencias de la Ingeniería, me permito declarar que una vez analizado en el sistema anti plagio **COMPILATIO**, luego de haber cumplido con los requerimientos exigidos de valoración, la presente tesis, se encuentra con un 8 % de la valoración permitida.

**CERTIFICADO DE ANÁLISIS**
magister

TESIS ROSAS ARTURO -FACTIBILIDAD DE USO DE INTELIGENCIA ARTIFICIAL EN POZOS PRODUCTORES

8%
Textos sospechosos

3% Similitudes
0% similitudes entre comillas
0% entre las fuentes mencionadas

6% Idiomas no reconocidos

27% Textos potencialmente generados por la IA (ignorado)

Nombre del documento: TESIS ROSAS ARTURO -FACTIBILIDAD DE USO DE INTELIGENCIA ARTIFICIAL EN POZOS PRODUCTORES.docx	Depositante: EDISON ANDRÉS BRITO AVILA	Número de palabras: 21.553
ID del documento: ad2702acba8d7961b362338fee15e4e50e083b9	Fecha de depósito: 13/6/2025	Número de caracteres: 152.335
Tamaño del documento original: 7,03 MB	Tipo de carga: Interface	
	fecha de fin de análisis: 13/6/2025	

FIRMA DEL TUTOR



Firmado electrónicamente por:
EDISON ANDRES BRITO AVILA

Validar únicamente con FirmaEC

ING. EDISON BRITO AVILA MSC.

C.I.: 0927177428

DECLARACIÓN DE AUTORÍA

Yo, **ARTURO ISMAEL ROSAS RAMIREZ**, declaro bajo juramento que el presente trabajo de titulación denominado **“FACTIBILIDAD DEL USO DE INTELIGENCIA ARTIFICIAL EN LA ETAPA DE PRODUCCION APLICADA A LA INDUSTRIA HIDROCARBURIFERA EN POZOS DEL ECUADOR”**, no tiene antecedentes de haber sido elaborado en la Facultad de Ciencias de la Ingeniería, Carrera de Petróleos, lo cual es un trabajo exclusivamente inédito y perteneciente de mi autoría.

Por medio de la presente declaración cedo los derechos de autoría y propiedad intelectual, correspondientes a este trabajo, a la Universidad Estatal Península de Santa Elena, según lo establecido por la ley de propiedad intelectual, por su reglamento y por la normativa institucional vigente.

Atentamente,



ARTURO ISMARL ROSAS RAMIREZ

AUTOR DE TESIS

C.I. 0923670558

CERTIFICACIÓN DEL TUTOR

ING. EDISON BRITO AVILA MSC.

TUTOR DE PROYECTO DE INVESTIGACIÓN

Universidad Estatal Península de Santa Elena

En mi calidad de Tutor del presente trabajo **“FACTIBILIDAD DEL USO DE INTELIGENCIA ARTIFICIAL EN LA ETAPA DE PRODUCCION APLICADA A LA INDUSTRIA HIDROCARBURIFERA EN POZOS DEL ECUADOR”**, previo a la obtención del Título de Ingeniero PETROLEOS elaborado por el Sr. **ARTURO ISMARL ROSAS RAMIREZ**, egresado de la carrera de PETROLEOS, Facultad Ciencias de la Ingeniería de la Universidad Estatal Península de Santa Elena, me permito declarar que luego de haber orientado, estudiado y revisado, la apruebo en todas sus partes.

FIRMA DEL TUTOR



ING. EDISON BRITO AVILA MSC.

TUTOR DE TESIS

CERTIFICADO DE GRAMATOLOGIA

CERTIFICADO DE GRAMATOLOGÍA

LICENCIADA EN CIENCIAS DE LA EDUCACIÓN MÁSTER EN EDUCACIÓN
Y DESARROLLO SOCIAL
MÁSTER EN EDUCACIÓN Y DESARROLLO SOCIAL
DOCENTE DE LENGUA Y LITERATURA – LEXICOLOGÍA

A petición del interesado tengo a bien certificar que he realizado la revisión y el análisis del contenido del presente trabajo de integración curricular.

“FACTIBILIDAD DEL USO DE INTELIGENCIA ARTIFICIAL EN LA ETAPA DE PRODUCCION APLICADA A LA INDUSTRIA HIDROCARBURIFERA EN POZOS DEL ECUADOR”, Del señor: *ARTURO ISMAEL ROSAS RAMIREZ C.I.: 0923670558*, egresado de la Universidad Estatal Península de Santa Elena, Facultad de Ciencias de la Ingeniería, Carrera de Ingeniería en Petróleos.

Que el mencionado trabajo, en el contexto general, cumple con los requisitos lingüísticos y normativos dados por la Real Academia Española para el uso del idioma español.

Certificación, que otorgo en la ciudad de Santa Elena a los doce días del mes de junio del dos mil veinticinco.

CERTIFICA



Lic. Narcisca Del Jesús García Cajape. MSc.

C.I. 1307095818

Telf: 0984898687

LICENCIADA EN CIENCIAS DE LA EDUCACION ESPECIALIDAD
EDUCACION BASICA
MAGISTER EN EDUCACION Y DESARROLLO SOCIAL
N.º DE REGISTRO DE SENESCYT: 1032-13-86036224

AGRADECIMIENTOS

Gracias a Dios, a toda mi familia por estar siempre conmigo, no fue un camino fácil, pese a la distancia, me hicieron saber que la familia siempre en los mejores y peores momentos.

De igual manera mis agradecimientos a la Universidad Estatal Península de Santa Elena, a todos mis docentes que a lo largo de la carrera aportaron en mi formación universitaria y personal.

A su vez agradecer al equipo del Bloque 58, donde realice mis prácticas por sus enseñanzas y conocimientos brindados, son un grupo de personas dedicadas y proactivas.

Finalmente quiero expresar mi sincero agradecimiento al Ing. Bernardo Burgos colaborar invaluable en este proceso, a mi tutor de tesis Ing. Édison apoyo durante todo este camino.

CONTENIDO

Pág.

CONTENIDO

DEDICATORIA	iii
CERTIFICADO DE ANTIPLAGIO	iv
DECLARACIÓN DE AUTORÍA.....	v
CERTIFICACIÓN DEL TUTOR.....	vi
AGRADECIMIENTOS.....	vii
CONTENIDO.....	ix
LISTA DE FIGURAS.....	xii
LISTA DE TABLAS.....	xv
LISTA DE ANEXOS	xvii
NOMENCLATURA	xviii
RESUMEN	xx
CAPITULO I: INTRODUCCIÓN.....	1
1. PROBLEMA DE INVESTIGACIÓN.....	1
1.1 Antecedentes.....	3
1.2 Hipótesis	7
1.3 Objetivos De La Investigacion.....	8
1.3.1 Objetivo General.....	8
1.3.2 Objetivos específicos.....	8

1.4 Alcance	9
1.5 Variables	11
1.5.1 Variables Dependientes	11
1.5.2 Variables Independientes.....	12
CAPITULO II: MARCO TEÓRICO	13
2. CONCEPTOS FUNDAMENTALES DE INTELIGENCIA ARTIFICIAL	13
2.1 Evolución De La Ia	13
2.2 Ia: General Y Específica	15
2.3 Aprendizaje Automático (Machine Learning).....	21
2.3.1 Redes Neuronales Artificiales (Rna)	25
2.3.2 Sistemas Expertos.....	28
2.3.3 Métricas De Regresión	36
2.3.4 Métricas De Clasificación.....	38
2.3.5 Métricas De Agrupamiento (Clustering)	40
2.4 Modelos Predictivos.....	41
A) Random Forest	41
B) Xgboost	41
C) Catboost.....	41
D) Máquina De Vectores De Soporte (Svm)	41
E) Red Neuronal Artificial (Rna).....	42
F) Anaconda Y Jupyter Notebook.....	42
CAPITULO III: METODOLOGÍA.....	55
3. METODOLOGÍA DE LA INVESTIGACIÓN.....	55

3.1 Enfoque De Investigación.....	55
3.2 Recopilación De Datos E Información	55
3.3 Población Y Muestra.....	57
3.3 Aplicación Tecnica Con Machine Learning	57
CAPITULO IV: ANÁLISIS Y DISCUSIÓN DE RESULTADOS.....	80
4. EVALUACIÓN Y COMPARACIÓN DE RESULTADOS	80
4.1 Resultados en Clasificación	81
4.2 Resultados en Regresión:.....	82
4.3. Resultados y Validación	83
4.5. Resultados Variables Relevantes	85
4.6. Casos de Éxito y Casos Fallidos	89
4.6.1 Casos de Éxito: Detecciones Precisas de Fallas	89
4.6.2 Casos Fallidos: Falsos Positivos y Falsos Negativos	90
4.6.3 Evaluación sobre Series Temporales Largas.....	92
CAPITULO V: CONCLUSIONES Y RECOMENDACIONES.....	94
5.1 CONCLUSIONES.....	94
5.2 RECOMENDACIONES	96

LISTA DE FIGURAS

	Pág.
Figura 1 Alan Turing padre de la IA Fuente: Alan Turing: El Padre de la #IA (Kopuro.com)	13
Figura 2 Machine learning subcampo de la IA (Alonso, 2020)	18
Figura 3 El comienzo de la Industria4.0 autor: prevencion integral.com	27
Figura 4 Matriz de confusión autor: wikipedia.com	39
Figura 5 Icono de Anaconda Autor: Ubunlog.com	42
Figura 6 Icono Jupyter Notebook fuente:jumpingrivers	43
Figura 7 Icono Jupyter Notebook fuente:jumpingrivers	44
Figura 8 Sistema en lo que Jupyter Notebook puede ser ejecutado autor: aprende machine	44
Figura 9 Componentes basicos BES Autor: monografías.com	45
Figura 10 Recolección y carga de datos autor: propia en JN.....	58
Figura 11 Datos types autor: Jupyter Notebook.....	59
Figura 12 Carga de librerías especializadas	60
Figura 13 Preprocesamiento y limpieza del dataset Autoria propia	60

Figura 14 Primeras filas del dataset autoria propia.....	61
Figura 15 Resumen estadístico autoría: propia	61
Figura 16 Creacion columna falla y umbrales.....	61
Figura 17 Fallas critica y funcionamiento normal autor:Jupyter Notebook	62
Figura 18 Distribución de umbrales TI,PDES,TM,PIP.....	63
Figura 19 Variables features y target.....	64
Figura 20 División del conjunto de datos 80-20	65
Figura 21 Codigos para curvas y matrices en Jupyter Notebook	66
Figura 22 Curva ROC-FALLA REAL autoria:Jupyter Notebook	66
Figura 23 Entrenamiento del modelo autor: JN	67
Figura 24 Código para Evaluación del modelo autor: JN.....	67
Figura 25 Precision, recall, accuracy y F1-score, autoria: Jupyter Notebook..	68
Figura 26 Variables en orden de importancia numerica.....	69
Figura 27 Regresion para estimar falla en el tiempo Autor: Jupyter Notebook	69
Figura 28 Grafica regresion Sin límite de tiempo	70
Figura 29 Grafica regresion Ventana de 90 días	70
Figura 30 Validación cruzada para robustes del modelo	71
Figura 31 Falsos positivos y negativos - XGBoost.....	72

Figura 32 Falsos positivos y negativos - CATBoost.....	73
Figura 33 Falsos positivos y negativos -REDES NEURONALES	74
Figura 34 Falsos positivos y negativos- Vector Machine	75
Figura 35 Modelo con el mejor F1-score autor: JUPYTER NOTEBOOK	76
Figura 36 Secuencias resumidas del modelado autor: Autoría propia	77
Figura 37 Comparación de los 5 modelos autor: autoría propia.....	78
Figura 38 Línea de tiempo en predicción de falla a 7 días autor: JN.....	79
Figura 39 Diagrama de flujo de creación del modelado con inteligencia artificial autor: Arturo Rosas 2025	80
Figura 40 F1-Score y su relación con cada modelo autor: autoría propia	82
Figura 41 Validación cruzada en interfaz Jupyter Notebook (F1-SCORE) Autor: Autoría propia	84

LISTA DE TABLAS

Pag.

Tabla 1 Comparación de IAG VS IAE Autor: Autoria propia.....	20
Tabla 2 Umbrales técnicos para clasificación de fallas	62
Tabla 3 Variables utilizadas en el modelo autoria propia	64
Tabla 4 Métricas de evaluación de modelos autor: autoria propia.....	79
Tabla 5 Resultados de diferentes modelos Autor: autoría propia	81
Tabla 6 Predicción de métricas en base al tiempo de falla autor: Autoría propia	82
Tabla 7 Parte del Dataset de historial de pozo "X" autor: Petroecuador	88
Ilustración 1 Programacion Clasica Y Machine Learning Autor: Alberto Maisueche Cuadrado	22
Ilustración 2 Relación con otros campos del machine learning autor: Villen 2019	24
Ilustración 3 Arquitectura de una red neuronal autor: Recuperado de: https://naukas.com/2015/12/09/acertando-quinielas-redes-neuronales/ :.....	35

LISTA DE ECUACIONES

Pág.

Ecuación 1 FORMULA MAE ERROR ABSOLUTO MEDIO.....	36
Ecuación 2 FORMULA MSE ERROR CUADRATICO MEDIO	36
Ecuación 3 FORMULA RMSE	37
Ecuación 4 FORMULA RMSLE.....	37
Ecuación 5 COEFICIENTE DE DETERMINACION R2	37
Ecuación 6 FORMULA R2 AJUSTADO	38
Ecuación 7 Metrica Precision- Accuracy	38
Ecuación 8 Metrica Precisión (Precision)	38
Ecuación 9 Metrica Exhaustividad (Recall)	38
Ecuación 10 Metrica Puntaje F1 (F1 Score).....	39
Ecuación 11 Metrica Puntaje $F\beta$ ($F\beta$ Score).....	39
Ecuación 12 Metrica ROC (AUC-ROC).....	39
Ecuación 13 Índice de Silueta (Silhouette Score).....	40
Ecuación 14 Matriz Índice de Davies-Bouldin	40
Ecuación 15 Matriz Adjusted Rand Index - ARI.....	40

LISTA DE ANEXOS

Pag.

ANEXO 1 PRINCIPALES COMPONENTES Y CONFIGURACIÓN TÍPICA DE BOMBA BES.....	101
ANEXO 2 PARTE DEL DATASET CON SUS PARAMETROS ELECTRICOS	103
ANEXO 3 EVALUACION CON METRICAS EN JUPYTER NOTEBOOK.....	104
ANEXO 4 INSTALACION DE PROGRAMAS PARA MODELADOS.....	104
ANEXO 5 INSTALACION DE MODELO GXBOOST	105
ANEXO 6 VALIDACIONES EN SERIE RED NEURONAL.....	105
ANEXO 7 PREDICCION EN FALLA SOBRE PRESION	106
ANEXO 8 MODELO CALIBRADO CATBOOST	106
ANEXO 9 F1-SCORE CON 5 MODELOS	106

NOMENCLATURA

SLA	Sistema Levantamiento Artificial
AI	Artificial Intelligence (Inteligencia Artificial)
ML	Machine Learning (Aprendizaje Automático)
DL	Deep Learning (Aprendizaje Profundo)
EDA	Exploratory Data Analysis (Análisis Exploratorio de Datos)
MAE	Mean Absolute Error (Error Absoluto Medio)
MSE	Mean Squared Error (Error Cuadrático Medio)
RMSE	Root Mean Squared Error (Raíz del Error Cuadrático Medio)
RMSLE	Root Mean Squared Logarithmic Error
SW	Swab
CPB	Clay Pebble Bed
PB	Passage Beg
BSW	Contenido de agua y sedimentos
R ²	Coefficient of Determination (Coeficiente de Determinación)
AUC	Area Under the Curve (Área Bajo la Curva)
ROC	Receiver Operating Characteristic (Curva Característica Operativa del Receptor)
TP	True Positive (Verdadero Positivo)
TN	True Negative (Verdadero Negativo)
FP	False Positive (Falso Positivo)
FN	False Negative (Falso Negativo)
ACC	Accuracy (Precisión global)

PREC Precision (Precisión positiva)

REC Recall / Sensitivity (Exhaustividad o Sensibilidad)

F1 F1 Score (Puntaje F1)

DBI Davies–Bouldin Index (Índice de Davies-Bouldin)

ARI Adjusted Rand Index (Índice de Rand Ajustado)

SIL Silhouette Score (Índice de Silueta)

CV Cross Validation (Validación Cruzada)

KNN K-Nearest Neighbors (K Vecinos más Cercanos)

SVM Support Vector Machine (Máquina de Vectores de Soporte)

RF Random Forest (Bosque Aleatorio)

XGB XGBoost (Extreme Gradient Boosting)

CB CatBoost

NN Neural Network (Red Neuronal)

BES Bomba Electrosomergible (aplicable a tu caso)

JNB Jupyter Notebook

ANA Anaconda

GPU Graphics Processing Unit (Unidad de Procesamiento Gráfico)

CPU Central Processing Unit (Unidad Central de Procesamiento)

IoU Intersection over Union (Intersección sobre Unión; usado en segmentación)

MLOps Machine Learning Operations (Operaciones de ML)

K-Fold Cross Validation (Validación Cruzada K-Fold)

GridSearchCV Grid Search Cross Validation (Búsqueda de Hiperparámetros con Validación Cruzada)

“FACTIBILIDAD DEL USO DE INTELIGENCIA ARTIFICIAL EN LA ETAPA DE PRODUCCION APLICADA A LA INDUSTRIA HIDROCARBURIFERA EN POZOS DEL ECUADOR”

Autor: ROSAS RAMIREZ ARTURO ISMAEL

Tutor: EDISON BRITO

RESUMEN

La inteligencia artificial (IA) está revolucionando el siglo en el que vivimos tanto en educación, ciberseguridad, automatización e incluso en todo tipo de industria, incluyendo la producción de hidrocarburos.

En la actualidad, se han desarrollado avances tecnológicos que permiten optimizar el proceso de identificación de patrones de flujo, a partir de la implementación de técnicas de inteligencia artificial, debido a que existen limitaciones para monitorear varios patrones, interior de las tuberías, toma de decisiones, malinterpretación de datos entre otros. Unos avances que ha hecho la IA en las industrias están:

Control de Calidad Preciso, la IA permite supervisar el comportamiento y la operación de cada componente del equipo de fabricación. Al recopilar datos del proceso, se detectan errores y defectos que afectan el producto final.

Automatización y Monitoreo en Tiempo Real, la IA recopila datos en tiempo real sobre el funcionamiento de las plantas de producción. Esto automatiza procesos rutinarios y permite monitorear y recomendar intervenciones en la producción. Áreas como el mantenimiento predictivo, la calidad y la detección de defectos se benefician de estas soluciones.

Agilidad y Sostenibilidad: La IA también ofrece adaptabilidad rápida a cambios en la demanda o especificaciones del producto. Además, op La inteligencia artificial (IA) está transformando profundamente múltiples sectores en este siglo, desde la educación y la ciberseguridad hasta la automatización industrial y la producción de hidrocarburos.

En la industria energética, por ejemplo, los avances tecnológicos basados en IA han mejorado significativamente la capacidad de identificar patrones de flujo en tuberías, algo que antes resultaba difícil de monitorear de forma eficiente.

Gracias a estos sistemas, hoy es posible reducir errores en la interpretación de datos y tomar decisiones con mayor precisión.

Algunas de las aplicaciones más destacadas de la IA en entornos industriales incluyen:

- **Control de calidad automatizado:** mediante sensores y análisis en tiempo real, la IA puede detectar fallos en los componentes de fabricación antes de que afecten al producto final.
- **Monitoreo continuo:** sistemas inteligentes recopilan datos, minutos a minuto sobre el estado de las plantas, lo que permite ajustar procesos al instante y anticipar fallos antes de que ocurran.
- **Mayor flexibilidad y sostenibilidad:** la IA ayuda a las empresas a adaptarse rápidamente a cambios en la demanda o en las condiciones del mercado, al mismo tiempo que mejora la eficiencia en el uso de recursos.

Aunque la implementación de estas tecnologías puede enfrentar ciertos desafíos, como la necesidad de inversión o la capacitación del personal, su impacto positivo en la velocidad, precisión y eficiencia de los procesos industriales ya es una realidad en muchas organizaciones.

Optimiza la gestión de recursos, contribuyendo a la sostenibilidad y eficiencia en la producción.

La IA mejora la velocidad, precisión y eficiencia de las operaciones de fabricación. Aunque existen desafíos en la implementación, su potencial es innegable

PALABRAS CLAVE: inteligencia artificial, producción de petróleo, levantamiento artificial, machine learning

“FEASIBILITY OF THE USE OF ARTIFICIAL INTELLIGENCE IN THE PRODUCTION STAGE APPLIED TO THE HYDROCARBON INDUSTRY IN WELLS IN ECUADOR”

Author: ROSAS RAMIREZ ARTURO ISMAEL

Tutor: EDISON BRITO

ABSTRACT

Artificial intelligence (AI) is revolutionizing the century we live in, in education, cybersecurity, automation, and even all types of instruction, including hydrocarbon production.

Currently, technological advances have been developed that allow for the optimization of the flow pattern identification process through the implementation of artificial intelligence techniques. This is due to the limitations of monitoring various patterns, the interior of pipes, decision-making, and data misinterpretation, among others. Some advances that AI has made in industry include:

Precise Quality Control: AI makes it possible to monitor the behavior and operation of each component of manufacturing equipment. By collecting process data, errors and defects that affect the final product are detected.

Automation and Real-Time Monitoring: AI collects real-time data on the operation of production plants. This automates routine processes and allows for monitoring and recommending interventions in production. Areas such as predictive maintenance, quality, and defect detection benefit from these solutions. **Agility and Sustainability:** AI also offers rapid adaptability to changes in demand or product specifications. Furthermore, artificial intelligence (AI) is profoundly transforming multiple sectors in this century, from education and cybersecurity to industrial automation and hydrocarbon production.

In the energy industry, for example, AI-based technological advances have significantly improved the ability to identify flow patterns in pipelines, something that was previously difficult to monitor efficiently. Thanks to these systems, it is now possible to reduce errors in data interpretation and make decisions with greater precision.

Some of the most notable applications of AI in industrial settings include:

- Automated quality control: Using sensors and real-time analytics, AI can detect faults in manufacturing components before they affect the final product.
- Continuous monitoring: Intelligent systems collect minute-by-minute data on plant status, allowing for instant process adjustments and anticipating failures before they occur.
- Greater flexibility and sustainability: AI helps companies quickly adapt to changes in demand or market conditions, while improving resource efficiency. Although the implementation of these technologies may face certain challenges, such as the need for investment or staff training, their positive impact on the speed, accuracy, and efficiency of industrial processes is already a reality in many organizations.

It optimizes resource management, contributing to sustainability and production efficiency.

AI improves the speed, accuracy, and efficiency of manufacturing operations. Although there are implementation challenges, its potential is undeniable.

KEY WORDS: artificial intelligence, oil production, artificial lift, automation, machine learning.

CAPITULO I: INTRODUCCIÓN

1. PROBLEMA DE INVESTIGACIÓN

El sector hidrocarburífero ecuatoriano atraviesa una etapa desafiante, marcada por la necesidad de mejorar sus procesos productivos, garantizar la seguridad operativa y avanzar hacia modelos más sostenibles. En este escenario, la inteligencia artificial (IA) se perfila como una herramienta clave para impulsar una transformación profunda en la manera en que se gestionan las operaciones.

Sin embargo, la viabilidad de implementar IA en la industria nacional depende de varios factores esenciales. Entre ellos destacan: la calidad y disponibilidad de datos, la infraestructura tecnológica ya existente y, sobre todo, las capacidades digitales del personal técnico. Sin una base sólida en estos tres aspectos, cualquier intento de integración tecnológica corre el riesgo de quedarse estancado en fases experimentales.

Entre las aplicaciones más útiles de la IA en este sector se encuentran: el monitoreo inteligente y continuo de pozos, la optimización dinámica de las tasas de extracción, la detección anticipada de fallas críticas en maquinaria y una gestión más eficiente de los recursos productivos. Estas innovaciones no solo aumentan la eficiencia, sino que también permiten reducir accidentes, prolongar la vida útil de los equipos y disminuir costos.

Desde la perspectiva de seguridad ocupacional, la IA puede reducir la exposición del personal a tareas peligrosas al automatizar procesos de alto riesgo. No obstante, esto conlleva nuevos retos: será necesario establecer protocolos para supervisar sistemas autónomos y asegurar que las decisiones automatizadas se tomen bajo principios éticos y técnicos claros.

Los beneficios más relevantes que ofrece la IA en este ámbito incluyen:

- Mejora continua en la producción: el análisis de datos en tiempo real permite detectar y corregir desviaciones operativas antes de que afecten la rentabilidad.
- Mantenimiento preventivo avanzado: permite anticiparse a fallas y programar intervenciones antes de que ocurran averías.
- Toma de decisiones estratégicas: los modelos de predicción apoyan la planificación a largo plazo y mejoran la asignación de recursos.

Ahora bien, la implementación de estas soluciones no está exenta de desafíos. La falta de formación especializada en tecnologías digitales obliga a priorizar la capacitación del talento humano. Asimismo, adaptar los sistemas actuales a soluciones basadas en IA requiere inversiones sostenidas y una gestión del cambio adecuada. Aspectos como la compatibilidad entre plataformas, la protección frente a ciberamenazas y la aceptación interna son factores críticos a considerar.

En conclusión, la inteligencia artificial representa una oportunidad estratégica para el sector energético en Ecuador, pero su aprovechamiento efectivo dependerá de una planificación estructurada, inversiones a largo plazo y una visión clara de modernización tecnológica.

1.1 ANTEDECENTES

La inteligencia artificial (IA) ha experimentado una evolución significativa desde sus inicios en la década de 1950. Durante ese período inicial, su desarrollo se limitaba principalmente a entornos académicos y a tareas básicas, como la resolución de problemas matemáticos o el juego de ajedrez. Pioneros como Alan Turing, quien en 1950 introdujo la conocida “prueba de Turing”, y Frank Rosenblatt, con la creación del perceptrón en 1958, sentaron las bases conceptuales para lo que actualmente es una de las tecnologías más influyentes del siglo XXI.

Con el paso del tiempo, la IA progresó desde modelos lógicos simples hacia sistemas más sofisticados, tales como los sistemas expertos de la década de 1980, que posibilitaron la simulación del conocimiento humano en aplicaciones específicas, especialmente en la industria manufacturera. No obstante, fue durante los años 2000, con la consolidación del aprendizaje automático y la disponibilidad masiva de datos, cuando la IA comenzó a impactar de manera tangible en diversos sectores productivos.

La computación en la nube o Cloud Computing se ha consolidado como respuesta a la necesidad empresarial de contar con servicios informáticos accesibles, escalables y eficientes a través de Internet, bajo un esquema de pago por uso que elimina la necesidad de costosas inversiones en infraestructura tecnológica (IBM, 2019).

Actualmente, la IA desempeña un papel fundamental en sectores como el energético y, en particular, en la industria petrolera. Los algoritmos de aprendizaje profundo permiten a las compañías analizar datos sísmicos con mayor precisión para localizar posibles yacimientos, prever fallas en equipos mediante mantenimiento predictivo y optimizar procesos complejos de extracción y procesamiento. Empresas líderes como Shell, BP y ExxonMobil ya

incorporan estas tecnologías para gestionar grandes volúmenes de información proveniente de sensores en tiempo real.

En América Latina, países como México, Brasil y Venezuela han comenzado a adoptar estas innovaciones, especialmente en campos petroleros maduros donde los métodos convencionales resultan menos eficientes.

Ecuador, por su parte, muestra un creciente interés en la implementación de modelos de IA para mejorar la eficiencia energética y operacional en su industria hidrocarburífero. Investigadores locales como Solís-Mora y Gruezo-Valencia (2022) resaltan que el empleo de algoritmos inteligentes posibilita anticipar fallos, optimizar recursos y minimizar el impacto ambiental.

Estas tecnologías abarcan herramientas como redes neuronales artificiales (ANN), máquinas de vectores de soporte (SVM) y modelos híbridos, que han sido cruciales para avanzar en técnicas de exploración, incluyendo la segmentación de imágenes sísmicas y la modelación geomecánica. Gracias a estas metodologías, es factible identificar zonas con alto potencial productivo y tomar decisiones más fundamentadas en cada etapa del proceso extractivo.

Dada la complejidad de estas tareas, resulta inviable que un programador pueda codificar manualmente todos los detalles específicos. En cambio, las máquinas deben ser capaces de aprender de la experiencia y adaptarse a nuevas circunstancias, una característica esencial de los modelos de aprendizaje automático (Shalev-Shwartz & Ben-David, 2014). Un ejemplo destacado en el campo del aprendizaje profundo es DeepMind Technologies, empresa británica fundada en 2010 y adquirida por Google en 2014. Esta compañía desarrolló una red neuronal capaz de aprender a jugar videojuegos analizando únicamente

los píxeles en pantalla. En 2015, lanzó AlphaGo, un algoritmo que sorprendió al derrotar a jugadores profesionales en el complejo juego chino Go (BBC, 2019).

En última instancia, la ciencia de datos busca convertir la información en acciones concretas en el mundo real, siguiendo un proceso en tres fases: primero, descubrir patrones y conocimientos relevantes mediante minería y análisis de datos; segundo, basar las decisiones en fundamentos estadísticos; y tercero, construir modelos a través de algoritmos de Machine Learning (Kozyrkov, 2018).

Sin embargo, la adopción de estas tecnologías no está exenta de desafíos. La capacitación del personal, la inversión en infraestructura y la adaptación de los sistemas existentes constituyen factores críticos para una implementación exitosa. A pesar de estos obstáculos, la IA se consolida como una herramienta estratégica para lograr una producción más eficiente, segura y sostenible en el sector petrolero ecuatoriano y regional.

Desde 1999, Toomarian y su equipo han impulsado el uso de redes neuronales en la industria petrolera, desarrollando algoritmos que analizan registros eléctricos de los pozos. Estos avances han permitido estimar con mayor precisión las características de los fluidos y la estructura estratigráfica de los yacimientos, mejorando la interpretación geológica y la toma de decisiones operativas.

En Ecuador, aunque la IA aún se encuentra en una etapa inicial dentro del sector industrial y enfrenta barreras estructurales y tecnológicas, existen señales alentadoras que evidencian un progreso gradual. Destaca el surgimiento de un ecosistema de IA que, si bien está menos desarrollado que en países como Brasil o Chile, presenta avances importantes. En el sector manufacturero, por ejemplo, se utilizan algoritmos de aprendizaje automático para optimizar cadenas de suministro y mejorar la eficiencia operativa.

Asimismo, se han logrado avances en la automatización de procesos industriales. Tecnologías basadas en IA se aplican en áreas como mantenimiento predictivo y gestión de inventarios, permitiendo a las empresas anticiparse a fallas en equipos clave y reducir tiempos de inactividad. No obstante, persisten retos significativos.

Según Pereiras et al. (2022), el empleo de redes neuronales facilita la estimación de deformaciones plásticas en conexiones roscadas de componentes petroleros, constituyendo una alternativa eficiente frente al método tradicional de elementos finitos. Entre las principales dificultades se encuentran la ausencia de políticas públicas claras y la insuficiente inversión en infraestructura tecnológica, lo que limita la adopción masiva de IA a nivel nacional. Sin embargo, el gobierno ha iniciado estrategias orientadas a fomentar la innovación y la integración de soluciones inteligentes en sectores clave.

Por otro lado, López Aguirre, García Benítez y Nicolás López (2020) reportan resultados prometedores en la aplicación de redes neuronales para estimar la velocidad de corte y parámetros elásticos-geomecánicos, alcanzando una correlación superior al 96% en comparación con mediciones de campo.

Además, la IA está empezando a desempeñar un rol relevante en la sostenibilidad industrial. Algunas compañías han implementado modelos inteligentes para optimizar el consumo energético en sus procesos, lo que no solo mejora la eficiencia operativa, sino que también contribuye a la reducción del impacto ambiental y se alinea con los objetivos nacionales de desarrollo sostenible.

En conjunto, estos antecedentes técnicos evidencian que, aunque queda un camino por recorrer, Ecuador cuenta con la oportunidad de impulsar una transformación industrial fundamentada en tecnologías emergentes como la inteligencia artificial.

1.2 HIPÓTESIS

La incorporación de tecnologías como la inteligencia artificial (IA) y las redes neuronales en el sector petrolero está abriendo nuevas oportunidades para mejorar la eficiencia y precisión en operaciones críticas, como el funcionamiento de los sistemas de bombeo electro sumergible.

Uno de los principales beneficios esperados es la optimización de la producción, ya que estos sistemas pueden ajustar en tiempo real los parámetros operativos para maximizar la extracción de crudo, adaptándose mejor a las condiciones del yacimiento. Esta adaptabilidad también permite responder de manera más eficaz a cambios en presión, temperatura o composición del fluido.

Además, la IA tiene el potencial de reducir significativamente los costos operativos, al minimizar el consumo energético y evitar el desgaste prematuro de los equipos. Esto se logra mediante algoritmos de mantenimiento predictivo que permiten programar intervenciones antes de que ocurra una falla crítica.

Otro aspecto clave es la detección temprana de fallos. Gracias al análisis continuo de datos, es posible identificar patrones anómalos en el funcionamiento del sistema, lo que facilita un diagnóstico oportuno y reduce tanto el tiempo de inactividad como los costos asociados a reparaciones de emergencia.

También se espera un aumento en la vida útil de los equipos, al anticipar condiciones que podrían ser perjudiciales para los componentes del sistema de bombeo. Esto no solo mejora la confiabilidad del sistema, sino que también reduce el impacto ambiental y económico asociado con el reemplazo frecuente de piezas.

Finalmente, una de las ventajas más destacadas es la capacidad de adaptación a distintos tipos de yacimiento. Los modelos basados en IA pueden ajustar las estrategias de operación en función de las características específicas del subsuelo, lo que permite una gestión más eficiente del recurso y una operación más segura.

Para evaluar la factibilidad de adoptar estas tecnologías, será fundamental compararlas con los métodos tradicionales y analizar tanto sus beneficios como los desafíos que implican su implementación. Este análisis permitirá determinar si la IA y las redes neuronales representan una evolución sustentable y rentable en el contexto de la producción petrolera.

1.3 OBJETIVOS DE LA INVESTIGACION

1.3.1 Objetivo General

- Identificar y aplicar técnicas de inteligencia artificial, como el uso de algoritmos de aprendizaje automático, para optimizar la selección de métodos de levantamiento artificial en pozos petroleros.

1.3.2 Objetivos específicos

- Desarrollar modelos de aprendizaje automático que analicen datos históricos de producción y variables operativas para predecir y optimizar la extracción de petróleo en tiempo real.
- Utilizar sistemas de inteligencia artificial para automatizar decisiones, como la selección de métodos de levantamiento artificial o la programación de mantenimientos preventivos, mejorando la eficiencia y reduciendo costos.
- Implementar algoritmos de detección de anomalías para identificar posibles fallas en los equipos o en el proceso de producción antes de que afecten la operación del pozo comparándolos con otros pozos.

1.4 ALCANCE

En el presente trabajo se explora el potencial de la inteligencia artificial, específicamente de las redes neuronales profundas, para mejorar la confiabilidad operativa de los sistemas de bombeo electrosumergible (BES) aplicados en la etapa de producción petrolera en Ecuador. Este tipo de tecnología es ampliamente utilizada en campos maduros de la región amazónica, donde la presión natural del yacimiento ya no es suficiente para sostener una producción comercialmente viable. Sin embargo, las fallas recurrentes en los equipos BES, muchas veces imprevistas, ocasionan detenciones prolongadas, incrementan los costos de intervención y afectan la planificación técnica de las cuadrillas de mantenimiento.

El objetivo principal de esta investigación es desarrollar un modelo predictivo que sea capaz de anticipar fallas potenciales antes de que estas ocurran, mediante el análisis de registros históricos reales provenientes de pozos ubicados principalmente en las provincias de Orellana y Sucumbíos. La idea es convertir a la inteligencia artificial en una herramienta de soporte para la toma de decisiones técnicas, que permita actuar de forma preventiva y reducir los tiempos no productivos.

Desde un enfoque multidisciplinario, el estudio se sustenta en tres ejes: la ingeniería de producción petrolera, la automatización de procesos industriales y la ciencia de datos. El núcleo del desarrollo será la implementación de una red neuronal tipo LSTM (Long Short-Term Memory), por su capacidad para identificar secuencias temporales complejas y detectar patrones que preceden a una falla mecánica o eléctrica.

Los datos que nutren el modelo provienen de un conjunto de 20 pozos seleccionados, operados con sistemas BES durante más de una década. Los registros incluyen variables como

presión de entrada (PIP), presión de descarga (PDES), temperatura del motor (TM) e intensidad eléctrica (IE), capturados en frecuencia diaria. Se han filtrado los pozos con historiales incompletos o inconsistentes, garantizando así una base de datos homogénea y robusta para el entrenamiento del modelo. Además, se incorpora la experiencia práctica de operadores y supervisores de campo, cuyas observaciones aportan valor cualitativo a la validación de los resultados.

Este trabajo también es considerado el uso de métricas estándar como precisión, sensibilidad, F1-score y exactitud para evaluar la capacidad del modelo tanto en clasificación binaria (falla/no falla) como en regresión (estimación del tiempo restante hasta la falla). La incorporación de ambos enfoques permite no solo emitir alertas oportunas, sino también generar estimaciones que ayuden a priorizar intervenciones según el nivel de riesgo.

Adicionalmente, se reconocen las cinco dimensiones del Big Data en el contexto de este proyecto. A las clásicas tres V (volumen, velocidad y variedad), se suman la veracidad, esencial para descartar lecturas anómalas o errores de sensores y el valor, entendido como la utilidad práctica de los resultados para las áreas de producción e ingeniería (Prometeus Global Solutions, 2019).

En cuanto a las categorías del aprendizaje automático, este estudio se apoya principalmente en el aprendizaje supervisado, dada la disponibilidad de etiquetas que indican eventos de falla. No obstante, se exploran también elementos del aprendizaje profundo (deep learning), ya que su arquitectura permite interpretar relaciones no lineales entre las variables y detectar señales sutiles de deterioro (Branco, 2018).

El desarrollo técnico del modelo se realizará utilizando herramientas como Anaconda y Jupyter Notebook, lo que facilitará un flujo de trabajo reproducible, flexible y transparente. La base de datos utilizada contendrá únicamente pozos con condiciones estables y registros homogéneos, lo que asegura un entorno de prueba controlado y representativo.

En resumen, esta investigación no solo busca desarrollar un modelo predictivo confiable, sino también ofrecer un aporte concreto a la operación petrolera, ayudando a reducir tiempos improductivos, minimizar costos de mantenimiento correctivo y mejorar la toma de decisiones en campo. La idea es que la inteligencia artificial se convierta en una aliada para lograr una producción más eficiente, segura y sostenible.

1.5 VARIABLES

1.5.1 Variables Dependientes

Las variables dependientes representan los resultados clave que se espera optimizar o predecir a través del modelo de inteligencia artificial propuesto. Estas variables permiten evaluar la eficacia y aplicabilidad práctica del sistema en condiciones reales.

- **Precisión del modelo:** Se refiere al nivel de acierto en la predicción de fallas y en la optimización del funcionamiento del sistema de bombeo. Esta precisión será medida mediante métricas como la tasa de aciertos (accuracy), el error cuadrático medio (RMSE) y el F1-score, según el enfoque del modelo (clasificación o regresión).
- **Tiempo de respuesta del modelo:** Mide la rapidez con la que el modelo es capaz de procesar los datos de entrada y generar una predicción útil. Es una variable crítica en contextos operativos donde las decisiones deben tomarse en tiempo real.

- Estimación del tiempo restante hasta la falla (en días): Esta variable aplica específicamente a los modelos de tipo regresivo y busca predecir cuántos días faltan antes de que ocurra una falla significativa en el sistema. Permite anticipar con mayor precisión la planificación de intervenciones.
- Eficiencia operativa alcanzada: Evalúa qué tan efectiva ha sido la aplicación del modelo para mejorar el rendimiento energético del sistema, reducir la frecuencia de fallas y disminuir los costos de mantenimiento. Se analizará comparando indicadores operativos antes y después de la implementación del modelo.

1.5.2 Variables Independientes

Las variables independientes corresponden a los factores que alimentan y condicionan el comportamiento del modelo predictivo. Estas variables son fundamentales para el entrenamiento, ajuste y validación de la red neuronal.

- Características del conjunto de datos: Incluye información histórica detallada sobre producción, incidencias de fallas y parámetros operativos de los sistemas de bombeo. Su calidad y diversidad influyen directamente en la capacidad de aprendizaje del modelo.
- Tipo de algoritmo de aprendizaje automático: Considera qué tipo de modelo se está utilizando. La elección del algoritmo puede influir significativamente en el rendimiento del sistema bajo distintas condiciones operativas.
- Volumen de datos de entrenamiento: Hace referencia a la cantidad de registros históricos utilizados para entrenar el modelo. Un mayor volumen de datos puede mejorar la capacidad del sistema para generalizar y realizar predicciones más precisas, siempre que los datos sean representativos y estén bien preprocesados.

CAPITULO II: MARCO TEÓRICO

2. CONCEPTOS FUNDAMENTALES DE INTELIGENCIA ARTIFICIAL

2.1 Evolución De La IA

La conceptualización de las redes neuronales artificiales tiene sus orígenes en la década de 1940, cuando Warren McCulloch y Walter Pitts introdujeron un modelo lógico que simulaba el comportamiento básico de una neurona biológica. Este modelo, basado en impulsos binarios, sirvió como punto de partida para el diseño de arquitecturas computacionales capaces de procesar información de forma análoga al sistema nervioso humano.

Posteriormente, en 1956, la conferencia de Dartmouth marcó un hito al consolidar formalmente la disciplina de la inteligencia artificial (IA) como un campo autónomo de investigación. Dentro de este marco emergente, Frank Rosenblatt desarrolló en 1958 el perceptrón, un modelo de red neuronal de una sola capa que introdujo el concepto de aprendizaje supervisado, mediante ajustes iterativos de pesos en función del error cometido.

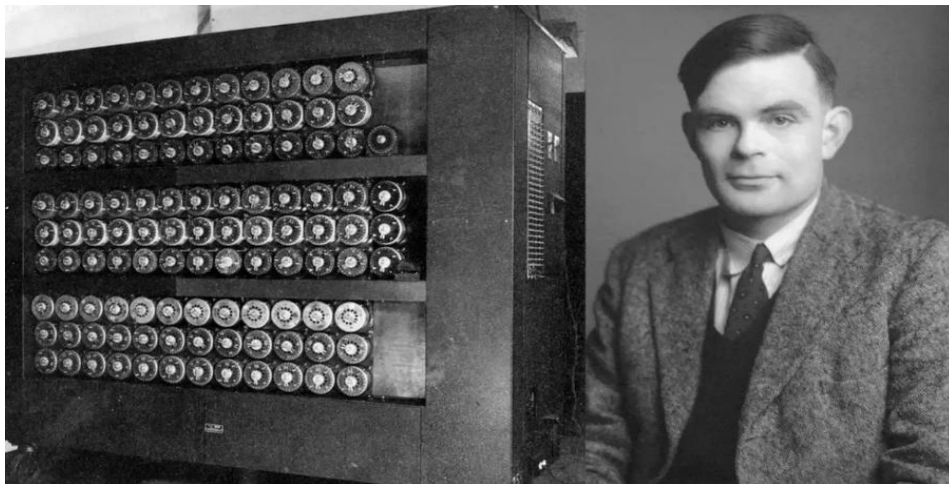


Figura 1 Alan Turing padre de la IA Fuente: Alan Turing: El Padre de la #IA (Kopuro.com)

Con el paso del tiempo, el campo de la IA ha experimentado ciclos de avance y retroceso, conocidos como "veranos" e "inviernos", motivados principalmente por las fluctuaciones en la capacidad computacional disponible y las expectativas en torno a sus aplicaciones. No obstante, el surgimiento del algoritmo de retropropagación en los años 80 facilitó el entrenamiento de redes multicapa, superando una de las principales limitaciones del perceptrón original.

Actualmente, se reconoce que la cuarta revolución industrial ha potenciado de manera significativa el desarrollo y aplicación de la inteligencia artificial en sectores estratégicos. Como sostienen Figueroa Donayre et al. (2023), el aprendizaje automático se ha convertido en una herramienta esencial para abordar retos complejos de la manufactura y la industria, al permitir la mejora continua de procesos productivos a través de la automatización inteligente.

El auge del aprendizaje profundo (deep learning), favorecido por el incremento exponencial del poder computacional y la disponibilidad de grandes volúmenes de datos, ha permitido alcanzar niveles de precisión antes inalcanzables en tareas como el reconocimiento de patrones, la visión artificial o el análisis predictivo. En este contexto, las redes neuronales han dejado de ser un objeto puramente académico para integrarse en sistemas industriales de monitoreo, control y toma de decisiones.

2. Aplicaciones prácticas en la industria petrolera

En el ámbito petrolero, especialmente en la gestión de pozos con sistemas de levantamiento artificial como las bombas electrosumergibles (BES), la inteligencia artificial ha comenzado a desempeñar un rol estratégico en diversas áreas operativas. A continuación, se destacan algunas de sus aplicaciones más relevantes:

1. Mantenimiento

predictivo

Los modelos basados en IA permiten anticipar fallas mecánicas o eléctricas

mediante el análisis de tendencias en variables como presión, temperatura y consumo eléctrico. Este enfoque reduce paradas inesperadas y optimiza los recursos asignados al mantenimiento.

2. **Optimización del rendimiento operativo**

Las redes neuronales pueden ajustar dinámicamente parámetros de funcionamiento de las bombas, como frecuencia o velocidad, en respuesta a las condiciones del pozo, favoreciendo una operación más eficiente y prolongando la vida útil de los equipos.

3. **Automatización y control inteligente**

Mediante sistemas de control avanzados, es posible integrar sensores, algoritmos predictivos y plataformas de supervisión remota, lo que permite una gestión más ágil y segura de la operación en campo. Incluso se han desarrollado asistentes conversacionales especializados para apoyar el diagnóstico técnico en tiempo real.

Según Bravent (2023), la evolución de la IA desde modelos simples hasta sistemas de aprendizaje profundo ha redefinido la manera en que se abordan problemas operativos complejos, aportando eficiencia, resiliencia y adaptabilidad a sectores altamente demandantes como el petrolero.

2.2 IA: GENERAL Y ESPECÍFICA

En el campo de la inteligencia artificial (IA), es fundamental distinguir entre dos grandes enfoques: la IA general y la IA específica. La IA específica (también conocida como IA estrecha o *narrow AI*) se refiere a sistemas diseñados para realizar una única tarea o un conjunto limitado de tareas. Estas soluciones están entrenadas para resolver problemas concretos, como el reconocimiento de imágenes, la predicción de fallas en equipos industriales, la detección de fraudes o la traducción automática de idiomas. Su

capacidad se limita al dominio para el cual fueron programadas o entrenadas, y fuera de ese contexto, pierden funcionalidad.

Por otro lado, la IA general (Artificial General Intelligence, AGI) representa un concepto más ambicioso: la creación de máquinas con capacidades cognitivas comparables a las del ser humano. Una AGI puede razonar, aprender y adaptarse a diferentes contextos, incluso a aquellos para los que no fue específicamente entrenada. Aunque aún es un ideal más que una realidad tecnológica, el desarrollo de AGI plantea profundas implicancias éticas, filosóficas y prácticas, ya que su alcance potencial abarca múltiples disciplinas y escenarios imprevisibles.

En el contexto industrial y de mantenimiento, como el que se aborda en esta investigación, la IA específica es la que se aplica mediante algoritmos de *Machine Learning* o *Deep Learning* entrenados para detectar patrones de falla o anticiparse a eventos críticos, sin necesidad de capacidades generales.

La inteligencia artificial (IA) se clasifica en dos categorías principales: IA general (también conocida como IA fuerte) e IA específica (también llamada IA débil). Estas categorías se diferencian en términos de alcance, capacidades y aplicaciones, lo que las hace adecuadas para distintos contextos tecnológicos e industriales.

Inteligencia Artificial General (IAG)

John McCarthy (1956): Durante la conferencia de Dartmouth, McCarthy acuñó el término "inteligencia artificial", definiendo el campo como el estudio de máquinas capaces de realizar tareas que normalmente requieren inteligencia humana. Este evento marcó el inicio formal de la IA como disciplina científica.

La IA general se refiere a sistemas capaces de realizar cualquier tarea cognitiva que un ser humano pueda ejecutar. Este tipo de IA tiene como objetivo replicar la inteligencia humana en su totalidad, incluyendo habilidades como

razonamiento, aprendizaje, planificación y resolución de problemas en múltiples dominios.

Características principales de la IAG

La Inteligencia Artificial General (IAG) se perfila como una forma avanzada de inteligencia artificial capaz de replicar competencias cognitivas humanas en múltiples contextos. Entre sus características más relevantes destacan:

- **Capacidad de adaptación contextual:** La IAG tiene el potencial de ajustarse a situaciones nuevas y resolver problemas no anticipados sin requerir intervención humana directa o reprogramación explícita.
- **Versatilidad funcional:** A diferencia de la IA estrecha, la IAG está diseñada para abordar tareas diversas, que pueden ir desde el procesamiento semántico del lenguaje natural hasta la toma de decisiones estratégicas en entornos dinámicos y complejos.
- **Autonomía cognitiva:** Este tipo de inteligencia busca emular la habilidad humana de razonar, aprender de la experiencia y actuar con independencia, lo que permitiría a las máquinas desenvolverse en escenarios no estructurados de forma similar a los seres humanos.
- **Estado de desarrollo actual:** Pese a los avances en IA estrecha y en aprendizaje profundo, la construcción de una verdadera inteligencia artificial general sigue siendo un reto técnico de gran magnitud. Lograr una emulación integral de la inteligencia humana requiere superar obstáculos relacionados con la consciencia, el sentido común y la transferencia de aprendizaje entre dominios heterogéneos.

Aplicaciones potenciales:

Aunque la IAG no está completamente desarrollada, se espera que tenga aplicaciones en áreas como:

- Robótica avanzada.
- Sistemas de toma de decisiones estratégicas.
- Exploración espacial autónoma.

Inteligencia Artificial Específica (IAE)

La IA específica, en contraste, se centra en realizar tareas concretas y limitadas.

Este tipo de IA está diseñado para resolver problemas específicos mediante algoritmos optimizados para un dominio particular.

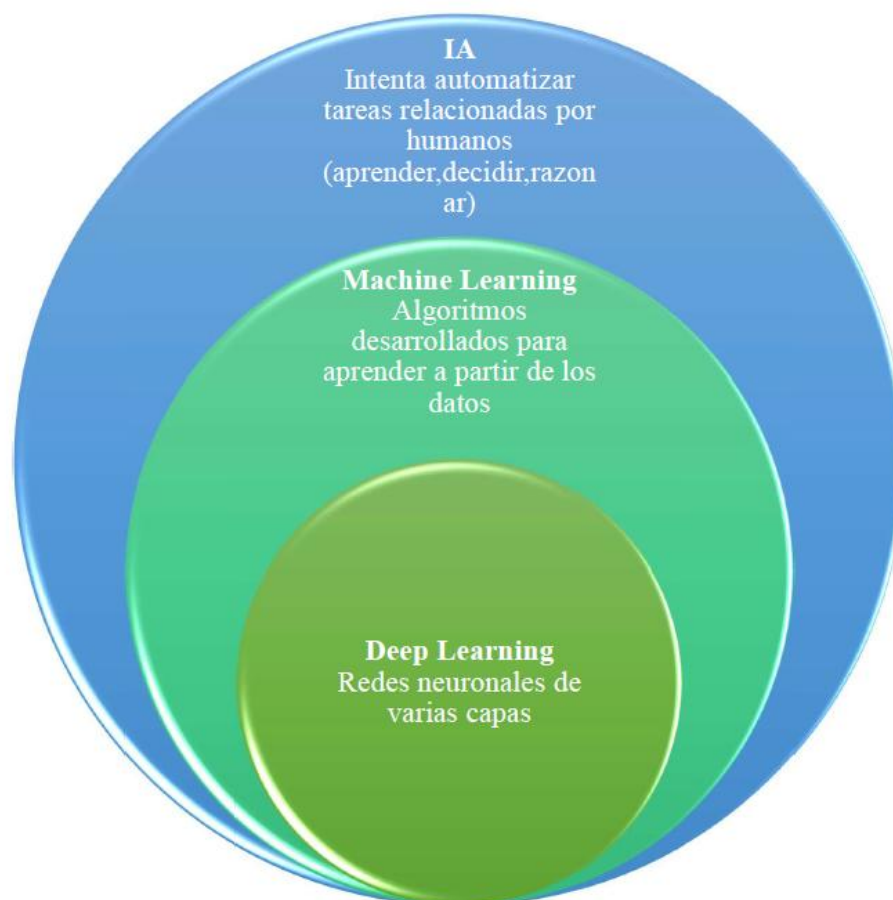


Figura 2 Machine learning subcampo de la IA (Alonso, 2020)

Características principales de la IAE

- **Especialización:** Está limitada a un conjunto definido de tareas, como el reconocimiento de imágenes o la traducción de idiomas.
- **Dependencia de datos:** Su rendimiento depende de la calidad y cantidad de datos utilizados para entrenarla.
- **Eficiencia:** Es altamente eficiente en su dominio específico, pero carece de la capacidad de generalizar conocimientos a otros contextos.
- **Estado actual:** La IAE es la forma más común de IA utilizada en la actualidad, con aplicaciones ampliamente implementadas.

Diferencias clave entre IAG e IAE

La diferencia más importante entre ambas está en su alcance y en qué tan bien pueden adaptarse a distintas tareas. La inteligencia artificial general (IAG) intenta imitar la inteligencia humana en todos sus aspectos, mientras que la inteligencia artificial estrecha (IAE) se enfoca solo en tareas específicas. Como señala McCarthy (2007), “la IAE ha sido la fuerza principal detrás de muchos avances tecnológicos recientes, mientras que la IAG todavía es una meta a largo plazo” (p. 56).

Implicaciones técnicas y desafíos

IA General:

1. **Desafíos técnicos:** Requiere avances en áreas como el razonamiento lógico, la comprensión del lenguaje natural y la ética en la toma de decisiones.
2. **Infraestructura:** Necesita hardware avanzado, como computación cuántica, para procesar grandes volúmenes de datos y realizar tareas complejas.

IA Específica:

1. Limitaciones: No puede adaptarse a tareas fuera de su programación inicial.
2. Ventajas: Es más fácil de desarrollar e implementar debido a su enfoque limitado.

Relevancia en el ámbito industrial

En el contexto industrial, la IAE es actualmente la más utilizada debido a su capacidad para optimizar procesos específicos, como el mantenimiento predictivo y la automatización de diagnósticos. Por otro lado, la IAG, aunque aún en desarrollo, promete revolucionar sectores como la robótica avanzada y la toma de decisiones estratégicas.

Este análisis técnico puede servir como base sólida para tu marco teórico, destacando las diferencias fundamentales entre ambos tipos de IA y su relevancia en aplicaciones prácticas.

Aspecto	IA General (IAG)	IA Específica (IAE)
Alcance	Amplio, abarca múltiples dominios.	Limitado a tareas específicas.
Capacidad de aprendizaje	Aprende y se adapta a nuevas situaciones.	Aprende dentro de un dominio predefinido.
Autonomía	Alta, puede razonar y actuar de manera independiente.	Baja, depende de algoritmos preprogramados.
Estado de desarrollo	En investigación y desarrollo.	Ampliamente implementada en la actualidad.
Ejemplos	Sistemas hipotéticos como robots humanoides avanzados.	Asistentes virtuales, sistemas de recomendación.

Tabla 1 Comparación de IAG VS IAE Autor: Autoría propia

2.3 APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)

Introducción a los Sistemas Basados en Aprendizaje Automático

El aprendizaje automático, o Machine Learning (ML), es una rama de la inteligencia artificial que le enseña a las máquinas a aprender a partir de datos, sin necesidad de programarlas paso a paso para cada situación. En lugar de seguir reglas fijas, los modelos de ML buscan patrones escondidos y relaciones complejas en grandes cantidades de información, para luego crear reglas que les ayuden a hacer predicciones o clasificar datos.

En Machine Learning, el aprendizaje consiste en ajustar los parámetros de un modelo según los datos que recibe. Estos datos, llamados data set, contienen variables independientes y dependientes. Las variables independientes, o features, son las que el algoritmo usa para construir el modelo, mientras que las variables dependientes, o labels, son las que el modelo debe predecir, ya que dependen de las independientes (Esteve, 2017).

Por lo general, un sistema de aprendizaje automático funciona en tres fases: primero, se entrena el modelo usando datos históricos; luego se valida qué tan preciso es; y finalmente se aplica en situaciones reales. Estos sistemas pueden ser supervisados, cuando se les da información con etiquetas o respuestas correctas; no supervisados, cuando exploran por sí solos sin una guía clara; o por refuerzo, cuando aprenden probando diferentes acciones y viendo los resultados.

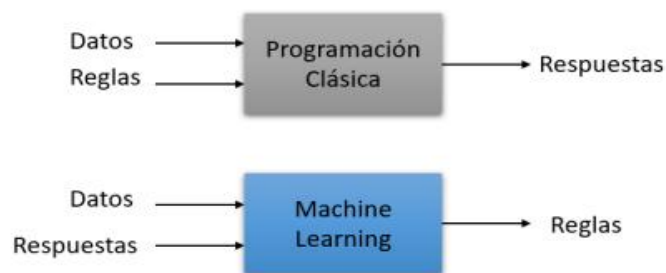


Ilustración 1 PROGRAMACION CLASICA Y MACHINE LEARNING autor: Alberto Maisueche Cuadrado

En sectores como la industria petrolera, el ML ha ganado mucha importancia porque ayuda a anticipar fallas en equipos como las bombas BES, mejorar la producción y detectar cualquier irregularidad en tiempo real. La efectividad del modelo depende mucho de la calidad y variedad de los datos con los que se entrena, por eso es fundamental hacer un buen trabajo de limpieza y análisis previo.

El Machine Learning también se utiliza en la producción de energía, donde la predicción de la carga eléctrica es clave para gestionar los recursos eficientemente. Modelos no lineales como Support Vector Regression, combinados con algoritmos de optimización como el Modified Firefly Algorithm, permiten anticipar la demanda eléctrica a corto plazo (Kavousi-Fard, Samet, & Marzbani, 2014).

Principios Básicos y Arquitectura de Redes Neuronales

Las redes neuronales artificiales (RNA) se inspiran en cómo funciona nuestro cerebro, donde millones de neuronas están conectadas entre sí para procesar información. En las computadoras, una red neuronal está formada por nodos, que serían como neuronas artificiales, organizados en varias capas: una de entrada, una o más ocultas y una de salida.

Cada neurona hace una operación matemática sencilla: recibe diferentes entradas con distintos pesos, las suma y luego pasa ese resultado por una función llamada de activación (como ReLU o Sigmoid) para decidir qué salida dará. Durante el entrenamiento, la red ajusta esos pesos usando algoritmos como el backpropagation (que se encarga de corregir errores), con el objetivo de mejorar sus predicciones.

El Deep Learning se encarga de procesar datos para identificar las señales de tráfico, detectar obstáculos y otros coches en la carretera, así como de cumplir con las leyes de tráfico. Estos algoritmos son imprescindibles para que el vehículo adquiera conocimiento, dado que los humanos no son capaces de programar todas las posibles situaciones que pueden ocurrir en la carretera (DHL, 2014; Gesing, Peterson, & Michelsen, 2018).

Lo que hace que las redes neuronales sean tan poderosas es que pueden manejar relaciones muy complejas y no lineales, cosas que serían difíciles de captar con métodos estadísticos tradicionales. Por eso son ideales para tareas como clasificar datos, hacer predicciones o reconocer patrones, especialmente en ambientes industriales donde los datos suelen ser ruidosos, variables y con muchas características diferentes.

Aplicaciones de Redes Neuronales en Diagnóstico y Mantenimiento Predictivo

Una de las áreas donde las redes neuronales han demostrado mayor utilidad es el mantenimiento predictivo. A diferencia del mantenimiento correctivo (reacción a la falla) o preventivo (revisiones periódicas), el mantenimiento predictivo busca anticipar fallas futuras mediante el análisis de datos históricos y en tiempo real.

Las redes neuronales permiten identificar correlaciones ocultas entre múltiples variables operativas (presiones, temperaturas, caudales, vibraciones) y la ocurrencia de fallas. En particular, los modelos LSTM (Long Short-Term Memory) o redes recurrentes (RNN) han mostrado una gran capacidad para analizar series temporales, lo que los hace especialmente útiles para prever el deterioro progresivo de equipos como bombas, motores o turbinas.

Además, estas aplicaciones se integran fácilmente en sistemas SCADA, plataformas de monitoreo remoto o incluso en aplicaciones móviles, permitiendo alertar a operadores mediante notificaciones o correos automáticos cuando se detecta una posible

anomalía. El resultado es una operación más segura, menor costo por paradas no planificadas y una prolongación significativa de la vida útil de los activos.

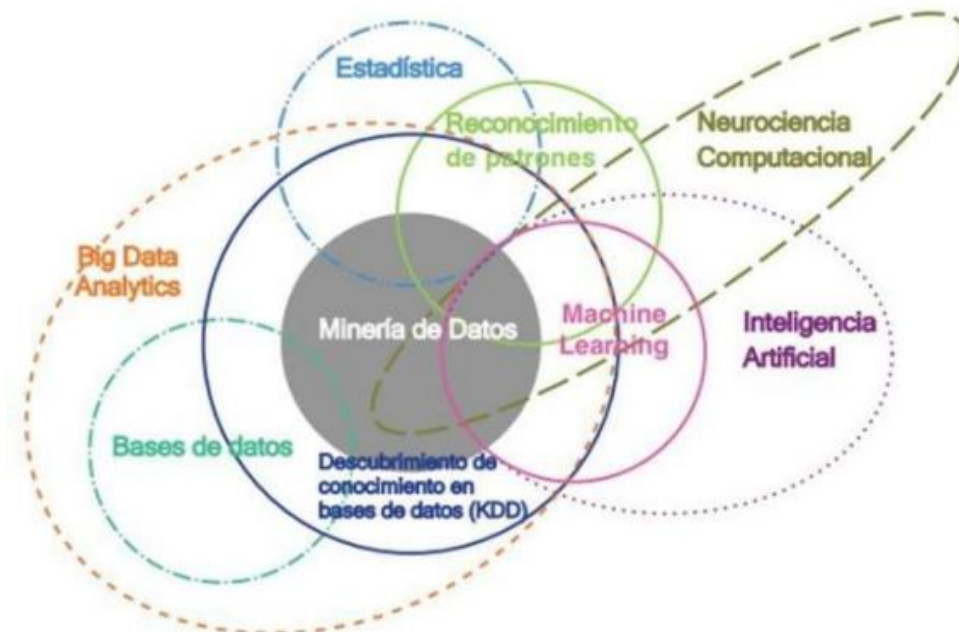


Ilustración 2 Relación con otros campos del machine learning autor: Villen 2019

Comparación entre Redes Neuronales Profundas (Deep Learning) y Otros Algoritmos de IA

El Deep Learning, o aprendizaje profundo, es una rama del aprendizaje automático que utiliza redes neuronales con muchas capas ocultas, de ahí lo de “profundo”. Estas redes pueden aprender diferentes niveles de representación de los datos, lo que las hace muy útiles para trabajar con imágenes, texto, audio o series temporales complejas.

Comparadas con otros algoritmos como máquinas de vectores de soporte (SVM), árboles de decisión, random forest o XGBoost, las redes profundas tienen más capacidad para entender relaciones complejas y no lineales, especialmente cuando los datos tienen muchas dimensiones. Pero también necesitan mucha más cantidad de datos y potencia de cálculo, y a veces son más difíciles de interpretar.

Por ejemplo, para predecir fallas en un sistema BES, métodos como XGBoost pueden dar resultados muy buenos con menos entrenamiento y con una explicación clara de qué variables influyen más. Sin embargo, si tienes datos secuenciales de alta frecuencia, como mediciones cada minuto u hora, un modelo de Deep Learning puede captar mejor cómo cambian las condiciones del pozo o equipo con el tiempo.

En resumen, elegir el modelo adecuado depende de la cantidad y tipo de datos, si necesitas entender bien cómo funciona el modelo, qué tan complejo es el problema y qué recursos tienes para computar. Muchas veces, combinar o comparar varios modelos puede ayudarte a tomar decisiones más acertadas.

2.3.1 Redes Neuronales Artificiales (RNA)

Las redes neuronales, que están inspiradas en cómo funciona nuestro cerebro, tienen muchas aplicaciones en la industria petrolera. Por ejemplo, se usan para predecir cómo se comportarán los pozos analizando datos históricos, lo que ayuda a planificar mejor la producción. También sirven para optimizar los sistemas de bombeo electrosumergible, ajustando parámetros para que funcionen con más eficiencia. Además, las redes neuronales son fundamentales para analizar datos sísmicos, ya que facilitan la interpretación de imágenes geofísicas, haciendo que la exploración de yacimientos sea más precisa y efectiva. Como menciona Bravent (2023), la inteligencia artificial ha avanzado mucho desde sus inicios con sistemas básicos como el perceptrón, hasta aplicaciones especializadas en la industria energética, como el mantenimiento predictivo y el análisis de datos sísmicos (p. 4).

La evolución de la inteligencia artificial ha cambiado completamente la forma en que las industrias mejoran sus operaciones y cuidan el medio ambiente. Por ejemplo, Bravent destaca el papel de los sistemas expertos en la manufactura, mientras que Autotec Engineering (2022) resalta cómo los algoritmos de machine learning son clave para planificar la energía y reducir costos.

Además, es importante recordar a pioneros como Alan Turing, quien en 1950 se preguntó "¿Pueden las máquinas pensar?" en su artículo "Computing Machinery and Intelligence", y propuso la famosa prueba de Turing para evaluar la inteligencia de las máquinas. Este trabajo fue fundamental para el desarrollo de la inteligencia artificial y la simulación de procesos mentales.

Por otro lado, Warren McCulloch y Walter Pitts, en 1943, crearon un modelo matemático que imitaba el funcionamiento de las neuronas usando impulsos binarios, lo que sentó las bases para construir sistemas computacionales inspirados en el cerebro humano.

Ecuador y su implicación industrial con IA

En Ecuador, la inteligencia artificial (IA) está empezando a ganar terreno como una herramienta útil en la industria, aunque todavía enfrenta algunos retos técnicos y estructurales. Aquí te dejo algunos antecedentes que muestran cómo va su desarrollo:

1. **Ecosistema de IA en crecimiento:** Según investigaciones recientes, Ecuador aún está en una etapa inicial comparado con otros países líderes en la región. Sin embargo, ya se han logrado avances en sectores como la manufactura, donde la IA se usa para mejorar la cadena de suministro y hacer que las operaciones sean más eficientes gracias a algoritmos de aprendizaje automático.
2. **Automatización en la industria:** En los últimos años, la IA se ha aplicado en áreas clave como el mantenimiento predictivo y la gestión de inventarios. Esto ha ayudado a las empresas a anticipar fallas en equipos importantes y a reducir los tiempos de parada, mostrando así el potencial de la IA para transformar la industria local.
3. **Brechas y oportunidades tecnológicas:** Aunque hay progresos, Ecuador todavía tiene una brecha tecnológica considerable frente a países como Brasil y Chile. La

falta de políticas públicas específicas y poca inversión en infraestructura tecnológica limitan la adopción de la IA. Afortunadamente, algunas iniciativas gubernamentales buscan impulsar la innovación con estrategias que promuevan el uso de la IA en sectores clave.

4. **Impacto en sostenibilidad:** La IA también se está convirtiendo en una aliada para enfrentar retos ambientales en la industria. Por ejemplo, ya se están desarrollando modelos que optimizan el consumo de energía en procesos productivos, ayudando a reducir el impacto ambiental y alineándose con los objetivos de sostenibilidad del país.

Un estudio de Solis-Mora y Gruezo-Valencia (2022) resalta que la inteligencia artificial ha sido efectiva para mejorar la eficiencia energética en Ecuador y en América Latina. Los autores mencionan que las técnicas de IA permiten monitorear y diagnosticar equipos en sistemas energéticos, anticipar fallas y optimizar el consumo en sectores residenciales, comerciales e industriales.

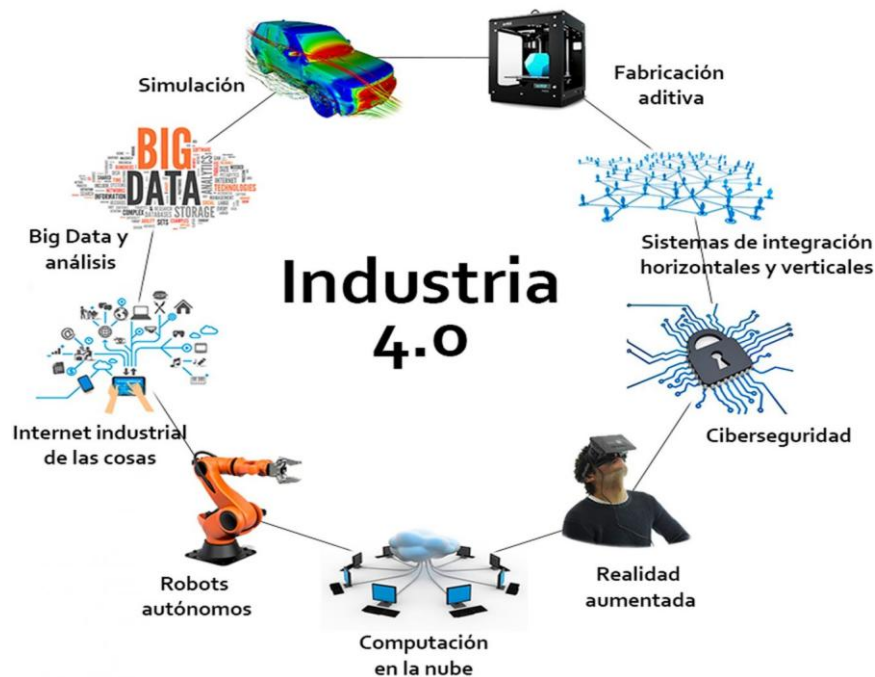


Figura 3 El comienzo de la Industria4.0 autor: prevencion integral.com

El aprendizaje automático es una tecnología que permite que los sistemas mejoren su desempeño sin necesidad de ser programados para cada caso específico. Algunos de los algoritmos más usados son:

En la industria petrolera, por ejemplo, se emplean herramientas como la regresión logística y los árboles de decisión para clasificar las condiciones de operación en los pozos, ayudando a tomar decisiones claras y efectivas. Los modelos de clustering, como K-Means, sirven para agrupar datos de producción y así descubrir patrones de comportamiento en los yacimientos. Además, las redes neuronales profundas, o Deep Learning, son fundamentales para el mantenimiento predictivo y el análisis de fallas en equipos, lo que ayuda a mantener las operaciones funcionando de manera continua y eficiente.

Según Autotec Engineering (2022), el aprendizaje automático también ha cambiado la forma de planificar en la industria manufacturera, mejorando la gestión de inventarios y automatizando tareas repetitivas.

Esta tecnología permite tomar decisiones en tiempo real y anticiparse a los problemas operativos.

2.3.2 Sistemas Expertos

Los sistemas expertos son programas diseñados para tomar decisiones basadas en reglas predefinidas.

En la industria petrolera, los sistemas expertos juegan un papel importante en el diagnóstico de fallas, identificando problemas en bombas electrosumergibles y recomendando soluciones eficaces. Además, la integración de modelos de inteligencia artificial permite optimizar los procesos de producción, mejorando la toma de decisiones en tiempo real.

Por último, la automatización del control de pozos facilita el ajuste de parámetros operativos sin necesidad de intervención humana, aumentando la eficiencia y precisión en las operaciones.

Características de la producción en pozos hidrocarburífero

La producción de hidrocarburos en pozos petroleros involucra varios factores técnicos y operativos:

- Tipos de pozos: Se clasifican en exploratorios, de evaluación y de desarrollo, dependiendo de su función en la extracción de petróleo y gas.
- Mecanismos de producción: Incluyen el empuje por gas disuelto, empuje hidráulico y gravedad, los cuales determinan la eficiencia de extracción.
- Sistemas de levantamiento artificial: Se utilizan métodos como bombeo mecánico, bombeo hidráulico y levantamiento por gas para mejorar la producción en pozos con baja presión natural.

Problemas actuales en la optimización de la producción en Ecuador

La industria petrolera en Ecuador enfrenta varios retos que afectan la eficiencia de su producción. Por un lado, la producción ha venido disminuyendo bastante debido a que las reservas están bajando y no se ha invertido lo suficiente en exploración. Además, gran parte de la infraestructura para extraer y procesar el petróleo es antigua y necesita una modernización urgente para mejorar su desempeño.

A esto se suman regulaciones estrictas, costos operativos elevados y una carga fiscal que dificulta la incorporación de nuevas tecnologías. Por último, el impacto ambiental que genera esta actividad provoca críticas y restricciones, especialmente en zonas sensibles, lo que complica aún más el desarrollo del sector.

Limitaciones de los métodos tradicionales de control y supervisión

Los sistemas tradicionales de monitoreo y control en la industria petrolera tienen varias limitaciones. Por ejemplo, todavía dependen mucho de inspecciones manuales que necesitan que alguien esté presente todo el tiempo, lo que puede retrasar la detección de problemas. Además, la falta de integración con sistemas automatizados inteligentes hace que la reacción ante fallas o anomalías sea más lenta. Otro problema común son los errores en las mediciones, ya que a veces hay diferencias entre lo que muestran los inventarios físicos y los registros contables, lo que puede generar pérdidas económicas importantes por la falta de precisión de los instrumentos.

Gestión predictiva automatizada

Fundamentos del mantenimiento predictivo.

El mantenimiento predictivo se basa en el análisis de datos históricos y en tiempo real para anticipar fallas antes de que ocurran. A diferencia del mantenimiento preventivo, que se ejecuta en intervalos fijos, el predictivo se ejecuta únicamente cuando es necesario, reduciendo tiempos de parada y costos de mantenimiento. Esta estrategia aprovecha herramientas como sensores, registros SCADA y sistemas de inteligencia artificial que aprenden de los patrones operativos para generar alertas tempranas.

Técnicas de modelado y simulación para predecir fallas.

El modelado predictivo utiliza técnicas estadísticas y algoritmos de machine learning para representar matemáticamente el comportamiento de sistemas complejos como las bombas BES. Entre las más utilizadas se encuentran los árboles de decisión, redes neuronales, máquinas de soporte vectorial y modelos de regresión. Estas herramientas permiten simular diversos escenarios de operación y estimar la probabilidad de fallo en función de variables observadas, facilitando la planificación de intervenciones.

Ventajas del uso de IA frente a métodos tradicionales.

El uso de inteligencia artificial en el mantenimiento predictivo proporciona una capacidad superior de detección de patrones complejos que los métodos tradicionales no pueden captar. Mientras que los enfoques convencionales requieren intervención humana para la interpretación de datos, los modelos de IA aprenden de forma autónoma y continua, mejorando su precisión con el tiempo. Además, reducen significativamente los errores humanos y permiten una respuesta más rápida y precisa frente a condiciones anómalas.

Aplicación de Inteligencia Artificial en Sistemas BES

La inteligencia artificial (IA) ha demostrado ser una herramienta eficaz para anticipar eventos críticos en sistemas industriales, especialmente en sistemas de levantamiento artificial como las Bombas Electrosumergibles (BES). El diseño de un modelo predictivo comienza con la recolección de datos históricos del pozo, incluyendo variables operativas como presión de entrada y descarga, temperatura, caudal, corriente y vibración. Estos datos son utilizados para entrenar algoritmos de aprendizaje automático que aprenden patrones que preceden a fallas reales.

El entrenamiento del modelo implica una fase de limpieza de datos, selección de características y validación cruzada para evitar el sobreajuste.

Implementación de redes neuronales para la prevención de fallas

Las redes neuronales artificiales, en particular las redes profundas (Deep Learning), han permitido mejorar la precisión del diagnóstico de fallas. Estos modelos simulan el comportamiento del cerebro humano mediante capas de neuronas conectadas entre sí que procesan información en paralelo. En el contexto de sistemas BES, una red neuronal puede identificar secuencias anómalas en los datos operativos que preceden una parada de la bomba, permitiendo emitir alertas con días de anticipación.

Integración de sensores inteligentes y recopilación de datos en tiempo real

Uno de los pilares de una IA eficiente es la disponibilidad de datos en tiempo real. Los sensores inteligentes instalados en el sistema BES permiten monitorear parámetros clave en todo momento. Estos sensores, conectados mediante protocolos de IoT (Internet de las Cosas), envían continuamente datos a plataformas en la nube donde son analizados en tiempo real. Esta capacidad no solo optimiza el mantenimiento predictivo, sino que también incrementa la vida útil de los equipos y reduce el riesgo de paradas no programadas.

Recolección y análisis de datos en sistemas industriales

La eficacia de cualquier modelo predictivo depende en gran medida de la calidad de los datos utilizados. En sistemas industriales como los pozos con BES, los datos se recolectan desde unidades de superficie, sensores de fondo de pozo, y registros operacionales. Este conjunto de datos incluye variables como voltaje, presión, caudal, y temperaturas, que reflejan el estado físico del sistema en distintos momentos operativos.

Técnicas de procesamiento de grandes volúmenes de datos (Big Data)

Debido al gran volumen y velocidad con que se generan estos datos, se hace indispensable el uso de tecnologías Big Data. Herramientas como Apache Spark o Hadoop permiten el almacenamiento distribuido y el análisis simultáneo de grandes volúmenes de información. Estas técnicas facilitan la detección de patrones temporales, estacionales y anómalos, esenciales para los modelos de mantenimiento predictivo.

Uso de IA para interpretar datos operativos complejos

La complejidad inherente en los sistemas BES y su entorno geológico hace que los datos presenten una alta dimensionalidad y no linealidades. La IA, especialmente los modelos basados en aprendizaje profundo, tiene la capacidad de interpretar estos patrones

complejos. Así, se puede generar un sistema de alerta temprana que indique el posible fallo de una bomba con días o incluso semanas de antelación.

IoT en sistemas industriales y su relación con la inteligencia artificial

El Internet de las Cosas (IoT) permite que todos los elementos del sistema (sensores, válvulas, unidades de superficie) estén conectados y puedan comunicarse entre sí. Este ecosistema digital facilita la recopilación y transmisión de datos a plataformas de inteligencia artificial, creando un entorno ciberfísico en donde los modelos se entrenan y se ajustan dinámicamente. El resultado es un sistema BES autónomo, capaz de adaptarse a cambios de operación sin intervención humana.

Papel de la conectividad en la detección preventiva de fallas

La conectividad constante permite que las señales de alerta sean enviadas al personal técnico o incluso integradas en sistemas de control automático. Por ejemplo, cuando el modelo predice una posible detención de la bomba en 7 días, se puede generar un correo automático, activar una revisión remota, o incluso ajustar parámetros para evitar el fallo. Este tipo de conectividad es esencial para minimizar tiempos muertos y optimizar recursos.

Algoritmos de Aprendizaje Supervisado y No Supervisado

Aplicaciones de aprendizaje supervisado en diagnóstico de equipos

El aprendizaje supervisado es una de las metodologías más utilizadas en el diagnóstico predictivo de fallas en equipos industriales. En este enfoque, el modelo es entrenado con datos etiquetados, es decir, con ejemplos donde ya se conoce el resultado (por ejemplo, si hubo o no una falla). En sistemas BES, esto permite a los modelos aprender las condiciones específicas que preceden a una parada, optimizando así la precisión del diagnóstico. Algoritmos como Random Forest, XGBoost o redes neuronales son comúnmente utilizados en este contexto.

En el aprendizaje supervisado, el sistema aprende a partir de ejemplos que incluyen datos de entrada y salida, buscando una función que relacione ambos (Russell & Norvig, 2010). Esto significa que la información a predecir ya está presente en los datos usados para entrenar el modelo. Existen dos tipos principales de problemas: regresión, que predice valores continuos, y clasificación, que asigna etiquetas (Management Solutions España, 2018).

La Regresión Lineal es una técnica común en Aprendizaje Supervisado, valorada por su simplicidad y eficacia. Su objetivo es predecir una variable dependiente a partir de una o más variables independientes, ajustando la mejor línea recta posible a los datos disponibles (Tusell, 2011).

El aprendizaje reforzado es un tipo de aprendizaje automático donde un agente toma decisiones para maximizar una recompensa, que refleja el impacto de sus acciones en el objetivo final. A diferencia del aprendizaje supervisado, aquí el agente no recibe instrucciones directas, sino que aprende de su propia experiencia al interactuar con el entorno. Estas interacciones se modelan mediante un Proceso de Decisión de Markov, donde cada estado depende solo del anterior (Russell & Norvig, 2010).

Lo que hace especial al Deep Learning es que permite que todas las capas del modelo se entrenen al mismo tiempo. Así, cuando se modifica un parámetro, los demás que dependen de él también se ajustan de manera automática, todo controlado por una sola señal de retroalimentación (Chollet, 2017).

En Deep Learning, las representaciones se organizan en capas que forman una Red Neuronal, un conjunto estructurado de neuronas inspirado en el cerebro (Figura 2.22). Cada neurona recibe información de la capa anterior a través de conexiones, procesa esos datos y envía un resultado a la siguiente capa. Al final, la capa final genera la predicción que se busca (Chollet, 2017).

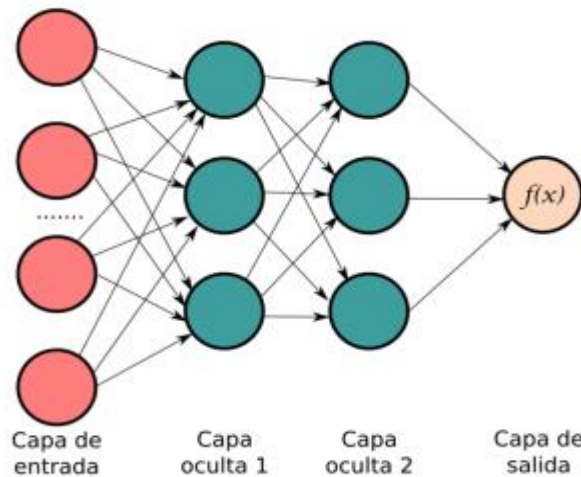


Ilustración 3 Arquitectura de una red neuronal autor: Recuperado de: <https://naukas.com/2015/12/09/acertando-quinielas-redes-neuronales/> :

Uso de aprendizaje no supervisado para detectar anomalías

Cuando no se dispone de etiquetas (es decir, cuando no se sabe si un comportamiento fue fallido o no), se utiliza el aprendizaje no supervisado. Este enfoque agrupa los datos por similitud o detecta valores atípicos que no siguen los patrones normales. En sistemas BES, esto permite identificar anomalías operativas en tiempo real, incluso si nunca antes se ha registrado un evento similar. Técnicas como K-means, PCA o Autoencoders son herramientas valiosas en esta tarea.

Comparación de algoritmos adaptativos en sistemas BES

Los algoritmos adaptativos, aquellos que ajustan sus parámetros en función del comportamiento del sistema en tiempo real, representan una ventaja significativa frente a modelos estáticos. Al comparar su rendimiento, se observa que los modelos basados en redes neuronales profundas y árboles de decisión tienen una alta capacidad de adaptación, aunque requieren mayor poder computacional. Por otro lado, los modelos más ligeros como SVM o regresión logística pueden ser preferidos en sistemas con restricciones de hardware o conectividad limitada.

Métricas de Evaluación en Aprendizaje Automático

En el campo del aprendizaje automático, evaluar correctamente el rendimiento de los modelos es esencial para garantizar resultados fiables y aplicables en contextos reales. Las métricas de evaluación permiten cuantificar la calidad de los modelos predictivos, ajustando su desempeño a las características específicas del problema. A continuación, se presenta una descripción detallada de las métricas más comunes empleadas para evaluar modelos de regresión, clasificación y agrupamiento, incluyendo fórmulas matemáticas y ejemplos ilustrativos.

2.3.3 Métricas de Regresión

Estas métricas se utilizan para evaluar modelos que predicen valores continuos.

Error Absoluto Medio (MAE): Calcula el promedio de las diferencias absolutas entre los valores predichos y los reales. Es útil por su interpretación directa, ya que indica en promedio cuánto se equivoca el modelo.

Ecuación 1 FORMULA MAE ERROR ABSOLUTO MEDIO

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- y_i : valor real
- \hat{y}_i : valor predicho
- n : número de observaciones

Error Cuadrático Medio (MSE): Promedia el cuadrado de las diferencias entre los valores predichos y los reales. Penaliza más los errores grandes, siendo sensible a valores atípicos.

Ecuación 2 FORMULA MSE ERROR CUADRATICO MEDIO

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Raíz del Error Cuadrático Medio (RMSE): Es la raíz cuadrada del MSE, lo que devuelve la métrica a las mismas unidades que la variable objetivo, facilitando su interpretación.

Ecuación 3 FORMULA RMSE

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Raíz del Error Cuadrático Logarítmico Medio (RMSLE): Similar al RMSE, pero utiliza logaritmos, lo que reduce el impacto de grandes errores y es útil cuando se espera que los errores relativos sean más importantes que los absolutos.

Ecuación 4 FORMULA RMSLE

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(1 + \hat{y}_i) - \log(1 + y_i))^2}$$

Coefficiente de Determinación (R^2): Indica la proporción de la variabilidad en la variable dependiente que es explicada por el modelo. Un valor cercano a 1 sugiere un buen ajuste del modelo.

Ecuación 5 COEFICIENTE DE DETERMINACION R2

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

R^2 Ajustado: Modifica el R^2 estándar teniendo en cuenta el número de variables independientes en el modelo, penalizando la inclusión de variables irrelevantes.

$$R_{\text{ajustado}}^2 = 1 - (1 - R^2) \cdot \frac{n - 1}{n - k - 1}$$

- n : número de muestras
- k : número de predictores (features)

2.3.4 Métricas de Clasificación

Estas métricas evalúan modelos que predicen categorías o clases.

Precisión (Accuracy): Proporción de predicciones correctas realizadas por el modelo en relación con el total de predicciones. Puede ser engañosa si las clases están desbalanceadas.

Ecuación 7 Metrica Precision- Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precisión (Precision): Proporción de verdaderos positivos entre todos los positivos predichos. Es crucial cuando los falsos positivos son costosos.

Ecuación 8 Metrica Precisión (Precision)

$$\text{Precision} = \frac{TP}{TP + FP}$$

Exhaustividad (Recall o Sensibilidad): Proporción de verdaderos positivos entre todos los positivos reales. Es importante cuando los falsos negativos son costosos.

Ecuación 9 Metrica Exhaustividad (Recall)

$$\text{Recall} = \frac{TP}{TP + FN}$$

Puntaje F1 (F1 Score): Media armónica entre precisión y exhaustividad, útil cuando se busca un equilibrio entre ambas, especialmente en conjuntos de datos desbalanceados.

Ecuación 10 Métrica Puntaje F1 (F1 Score)

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Puntaje F β (F β Score): Generalización del F1 Score que permite ponderar la importancia relativa de la precisión y la exhaustividad.

Ecuación 11 Métrica Puntaje F β (F β Score)

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{(\beta^2 \cdot \text{Precision}) + \text{Recall}}$$

Área Bajo la Curva ROC (AUC-ROC): Mide la capacidad del modelo para distinguir entre clases. Un valor de 1 indica una perfecta distinción, mientras que 0.5 sugiere una clasificación aleatoria.

Ecuación 12 Métrica ROC (AUC-ROC)

$$\text{TPR} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN}$$

Matriz de Confusión: Tabla que muestra las verdaderas clasificaciones frente a las predicciones del modelo, permitiendo identificar errores específicos en las clases.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figura 4 Matriz de confusión autor: wikipedia.com

2.3.5 Métricas de Agrupamiento (Clustering)

Estas métricas evalúan la calidad de los agrupamientos realizados por algoritmos de clustering.

Índice de Silueta (Silhouette Score): Mide qué tan similares son los objetos dentro de un mismo grupo en comparación con otros grupos. Un valor cercano a 1 indica una buena agrupación.

Ecuación 13 Índice de Silueta (Silhouette Score)

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

- $a(i)$: distancia promedio del punto i a los otros puntos de su mismo grupo
- $b(i)$: menor distancia promedio del punto i a puntos de otros grupos

Índice de Davies-Bouldin: Evalúa la calidad del clustering considerando la dispersión dentro de los clusters y la separación entre ellos. Valores más bajos indican mejores agrupamientos.

Ecuación 14 Matriz Índice de Davies-Bouldin

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{s_i + s_j}{d_{ij}} \right)$$

- s_i : dispersión dentro del clúster i
- d_{ij} : distancia entre los centroides de los clústers i y j

Índice de Rand Ajustado (Adjusted Rand Index - ARI): Mide la similitud entre dos agrupamientos, ajustando por la posibilidad de coincidencias aleatorias.

Ecuación 15 Matriz Adjusted Rand Index - ARI

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]}$$

- RI: Índice de Rand
- E[RI]: Valor esperado del índice por azar

2.4 MODELOS PREDICTIVOS

Para abordar el problema de predicción de fallas en bombas BES, se seleccionaron cinco modelos de aprendizaje automático con diferentes arquitecturas y principios matemáticos, a fin de comparar su desempeño en términos de precisión, capacidad de generalización y robustez frente a datos ruidosos. A continuación, se describen brevemente cada uno de ellos:

a) Random Forest

Random Forest es un algoritmo de ensamblado basado en múltiples árboles de decisión. Su fortaleza radica en la combinación de varios árboles para reducir el sobreajuste y aumentar la precisión. Es útil tanto para clasificación como para regresión, y ofrece interpretabilidad mediante la evaluación de la importancia de las variables.

b) XGBoost

XGBoost (Extreme Gradient Boosting) es una técnica de boosting que optimiza el rendimiento del modelo mediante el entrenamiento secuencial de árboles. Es reconocido por su eficiencia, rapidez y capacidad para manejar datos tabulares con alto rendimiento. También incorpora regularización para evitar el sobreajuste.

c) CatBoost

CatBoost es una variante del boosting especialmente optimizada para variables categóricas. Es robusto frente a valores faltantes, reduce el preprocesamiento necesario y mejora la precisión en datasets con interacciones complejas.

d) Máquina de Vectores de Soporte (SVM)

SVM es un modelo supervisado que busca encontrar el hiperplano que mejor separa las clases en el espacio de características. En su versión para regresión (SVR),

intenta predecir un valor continuo maximizando el margen de tolerancia. Es efectivo en espacios de alta dimension y con datasets pequenos.

e) Red Neuronal Artificial (RNA)

Las redes neuronales permiten modelar relaciones no lineales complejas mediante capas de procesamiento conectadas entre si. En esta investigacion se implemento una red neuronal multicapa (MLP) para tareas de clasificacion y regresion. Requiere mayor cantidad de datos y ajustes de hiperparametros, pero puede capturar patrones altamente no lineales.

f) ANACONDA Y JUPYTER NOTEBOOK

ANACONDA

Anaconda es una distribucion gratuita y de codigo abierto de los lenguajes de programacion Python y R, disenada para facilitar la computacion cientifica y el analisis de datos. Incluye mas de 300 paquetes preinstalados y herramientas como Jupyter Notebook, Spyder y RStudio. Ademas, ofrece "Anaconda Navigator", una interfaz grafica que permite gestionar entornos y paquetes sin necesidad de usar la linea de comandos.



Figura 5 Icono de Anaconda Autor: Uibunlog.com

Jupyter Notebook

Jupyter Notebook es una aplicación web que permite crear y compartir documentos que contienen código en vivo, ecuaciones, visualizaciones y texto narrativo. Es ampliamente utilizado en análisis de datos, modelado estadístico y aprendizaje automático.



Figura 6 Icono Jupyter Notebook fuente:jumpingrivers

Interfaz de Jupyter Notebook

La interfaz de Jupyter Notebook se compone de celdas que pueden contener código o texto. Las celdas de código permiten escribir y ejecutar fragmentos de código, mientras que las celdas de texto (usando Markdown) permiten documentar el análisis, incluir ecuaciones matemáticas y agregar explicaciones. Esta estructura facilita la creación de informes reproducibles y bien documentados.

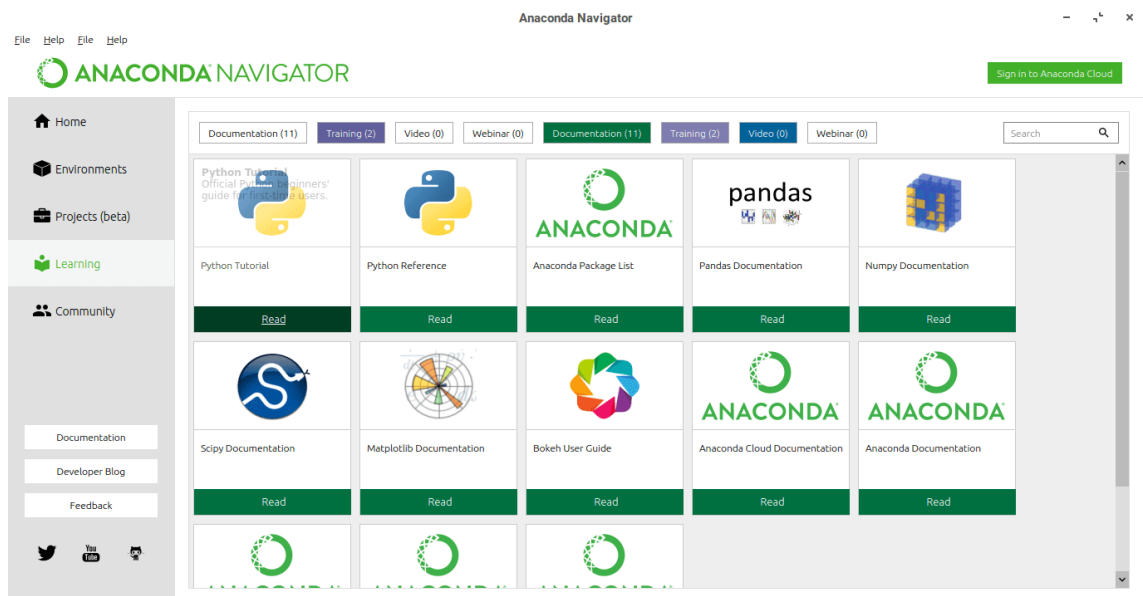


Figura 7 Icono Jupyter Notebook fuente: jumpingrivers

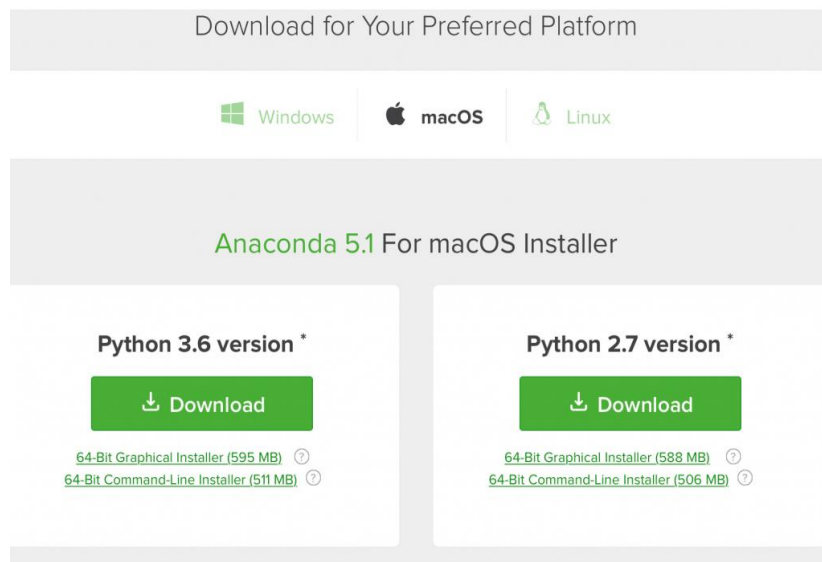


Figura 8 Sistema en lo que Jupyter Notebook puede ser ejecutado autor: aprende machine

Crerios tcnicos para BES

La selecci3n del sistema de levantamiento artificial es una de las decisiones tcnicas ms relevantes en la explotaci3n de un pozo petrolero, ya que impacta directamente en la productividad, la eficiencia energtica, el mantenimiento y los costos operativos a largo plazo. En el contexto de esta investigaci3n, se ha optado por centrar el modelado predictivo de fallas en pozos que operan mediante bombas electrosumergibles (BES), dado su uso extendido en la industria y su aplicabilidad en un rango amplio de condiciones de yacimiento.

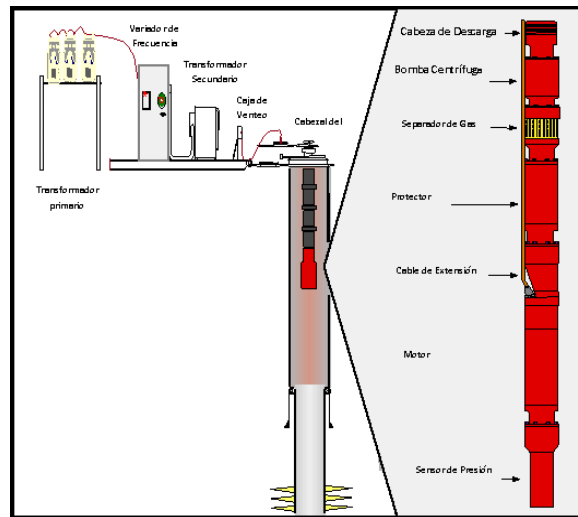


Figura 9 Componentes básicos BES Autor: monografías.com

La implementación adecuada de un sistema BES requiere considerar una serie de factores geológicos, atmosféricos, económicos y logísticos, los cuales influyen tanto en su desempeño como en su vida útil. A continuación, se describen las principales condiciones a tomar en cuenta para su implementación:

1. Condiciones Geológicas del Yacimiento

- Profundidad del pozo: Las BES son particularmente eficientes en pozos profundos, típicamente por encima de los 1,500 metros. Su capacidad de generar altos caudales permite mantener una producción estable incluso desde zonas profundas del reservorio.
- Presión de fondo: Este sistema es ideal en yacimientos con presión natural decreciente o insuficiente, donde el empuje del reservorio ya no es suficiente para mantener un flujo natural del fluido hacia la superficie.
- Composición del fluido: La presencia de arena, sólidos o gases libres puede ser un factor crítico. Si bien las BES pueden manejar pequeñas cantidades de gas, su eficiencia disminuye con altos porcentajes de corte de gas. Es recomendable que el pozo tenga control de producción arenosa o sistemas de separación previos.

2. Condiciones Atmosféricas y Ambientales

- Temperatura del entorno y del fluido: Las BES deben seleccionarse con base en la temperatura máxima de operación esperada. Condiciones térmicas extremas pueden acortar su vida útil o generar fallas prematuras en el motor y el cableado.
- Humedad y corrosividad del ambiente superficial: Afecta principalmente los componentes eléctricos de superficie y las conexiones. El diseño del sistema debe considerar protecciones contra la corrosión y la humedad.

3. Condiciones Económicas y Logísticas

- Costo de implementación: Las BES requieren una inversión inicial considerable, superior a otros sistemas de levantamiento como el bombeo mecánico. No obstante, su capacidad para manejar mayores volúmenes de producción justifica el costo cuando se busca mantener alta productividad.
- Costo operativo: Aunque su consumo energético es significativo, la eficiencia volumétrica y la automatización compensan este gasto, especialmente en campos con acceso a energía eléctrica estable.
- Tiempo de instalación: El armado e instalación de una bomba BES suele requerir entre 5 a 10 días, dependiendo de la profundidad, configuración del equipo y condiciones de acceso al pozo. Este tiempo debe ser considerado dentro de la planificación del cronograma de perforación y puesta en marcha.

4. Beneficios Técnicos de Usar un Sistema BES

- Alta capacidad de levantamiento: Las BES pueden manejar caudales desde unos pocos cientos hasta miles de barriles por día, adaptándose a diferentes niveles de producción.
- Operación automatizada: Facilita el monitoreo y control en tiempo real, lo que permite implementar sistemas de predicción como el desarrollado en esta investigación.

- Reducción de intervención en campo: Su funcionamiento continuo y estabilidad operativa permiten ampliar los intervalos entre mantenimientos.
- Integración con sistemas inteligentes: Su compatibilidad con sensores y sistemas de control hace posible el uso de algoritmos de predicción que anticipen fallas, como el propuesto en este trabajo.

5. Justificación de Elección para el Modelado de Fallas

- El enfoque de esta investigación se orienta hacia la predicción anticipada de fallas en sistemas BES, no solo por su prevalencia en la industria, sino también por las altas exigencias operativas y económicas que implica su implementación. Una falla no detectada en una BES puede significar una pérdida considerable de producción y costos elevados de reparación, especialmente en pozos profundos o de difícil acceso.

Factores comunes de fallas en los sistemas BES.

Entre las principales causas de fallas en sistemas BES se encuentran los problemas eléctricos, la abrasión por sólidos, la presencia excesiva de gas libre, fallos en el motor o sello, y la acumulación de incrustaciones. Estas fallas pueden reducir la eficiencia del sistema o provocar su paro total. La vida útil de un sistema BES depende en gran medida de la calidad del diseño del pozo, el monitoreo continuo de parámetros operacionales como presión, temperatura y vibración, así como de una correcta selección de materiales. La predicción y prevención de estos eventos mediante modelos de IA es fundamental para maximizar la disponibilidad del equipo y reducir los tiempos de inactividad no planificados.

Importancia de la monitorización en tiempo real.

La monitorización en tiempo real de variables críticas como presión, temperatura, voltaje, corriente y vibraciones permite detectar comportamientos anómalos que podrían indicar fallas inminentes. Esta información, recolectada a través de sensores instalados en el fondo del pozo y transmitida a la superficie, permite alimentar modelos predictivos basados en IA que anticipan problemas operacionales antes de que estos ocurran. Este enfoque no solo mejora la confiabilidad del sistema BES, sino que también optimiza la toma de decisiones operativas y reduce los costos asociados al mantenimiento correctivo

Por tanto, un modelo predictivo que permita anticipar condiciones de deterioro o apagado de la bomba representa una solución de alto impacto operativo y financiero, facilitando la toma de decisiones preventivas. Además, la automatización y digitalización que caracteriza a los sistemas BES los convierte en candidatos ideales para ser integrados con herramientas inteligentes como modelos de Machine Learning, alertas por correo electrónico, y sistemas de gestión digital del mantenimiento.

Aplicaciones específicas de inteligencia artificial (IA) en la producción de hidrocarburos

Los sistemas de bombeo electrosumergibles (BES) son fundamentales en la producción de petróleo, y la IA ha mejorado su rendimiento mediante:

- **Análisis nodal:** Modelos de IA optimizan el flujo de producción en pozos con BES.
- **Mantenimiento inteligente:** Algoritmos de aprendizaje automático predicen fallas y ajustan parámetros para prolongar la vida útil de los equipos.
- **Reducción de costos operativos:** La IA permite una gestión más eficiente del consumo energético en sistemas de bombeo.

Análisis predictivo y mantenimiento preventivo con modelos de IA

El mantenimiento predictivo apoyado en inteligencia artificial ha cambiado completamente la forma en que se gestionan los activos en la industria petrolera. Los algoritmos de machine learning analizan datos operativos para anticipar posibles fallas antes de que ocurran, lo que ayuda a evitar paros inesperados. Además, estos modelos optimizan las estrategias de mantenimiento, reduciendo los tiempos de inactividad y mejorando la eficiencia en general. Por otro lado, los sistemas inteligentes automatizan el diagnóstico, detectando patrones de desgaste y recomendando acciones preventivas para asegurar que los procesos funcionen sin interrupciones. Varios países ya están aprovechando la IA en la producción de hidrocarburos y han visto resultados positivos:

Estados Unidos

Estados Unidos se posiciona como uno de los principales líderes en la adopción e integración de inteligencia artificial (IA) dentro de la industria petrolera. Empresas multinacionales como ExxonMobil han implementado modelos predictivos basados en machine learning para optimizar la extracción de hidrocarburos y anticipar fallas en equipos críticos, reduciendo así costos operativos y evitando paradas inesperadas. Además, la IA se emplea para procesar grandes volúmenes de datos sísmicos y geológicos, mejorando significativamente la exploración y evaluación de nuevos yacimientos.

Fuera del sector petrolero, la IA en Estados Unidos tiene aplicaciones avanzadas en salud, donde se utilizan algoritmos para el diagnóstico asistido por imagen y medicina personalizada, así como en la industria financiera para la detección de fraudes y análisis predictivo de mercados.

Europa

En Europa, la inteligencia artificial ha sido fundamental para la optimización de procesos industriales con un enfoque marcado en la sostenibilidad y el cumplimiento ambiental. Se han desarrollado sistemas inteligentes para supervisar operaciones de perforación con el fin de incrementar la seguridad y minimizar riesgos ambientales. Los modelos predictivos basados en machine learning también permiten anticipar fallas en bombas y otros equipos de producción, mejorando la planificación del mantenimiento y reduciendo costos. En logística y transporte de hidrocarburos, la IA contribuye a una gestión más eficiente que ayuda a disminuir emisiones contaminantes.

Además de la industria petrolera, la IA en Europa está siendo aplicada extensamente en sectores como la movilidad inteligente, mediante vehículos autónomos y sistemas de transporte público optimizados, y en energías renovables, para la predicción de generación y demanda eléctrica.

Asia Pacífico

La región Asia Pacífico ha experimentado una adopción acelerada de tecnologías basadas en IA para mejorar la eficiencia y seguridad en la producción petrolera. Muchas empresas han implementado sistemas de monitoreo remoto que permiten supervisar plantas de producción a distancia, optimizando el control operacional. La IA se utiliza también para fortalecer la seguridad industrial, detectando niveles tóxicos y fugas potenciales, así como para la optimización energética, ajustando sistemas de climatización para mantener condiciones ideales en instalaciones productivas.

Paralelamente, en sectores como la manufactura avanzada y la agricultura de precisión, la inteligencia artificial ha impulsado mejoras sustanciales, automatizando procesos y optimizando recursos, lo que ha contribuido al desarrollo económico y tecnológico de la región.

Factibilidad y consideraciones tecnológicas en la implementación de inteligencia artificial (IA) en la producción petrolera en Ecuador:

La adopción de IA en la industria petrolera ecuatoriana requiere una evaluación detallada de costos y beneficios. Algunos factores clave incluyen:

Costos de implementación

- **Infraestructura tecnológica:** La adopción de sistemas inteligentes en la industria petrolera requiere la instalación de sensores avanzados y equipos de monitoreo en tiempo real, además de plataformas robustas para la recolección, almacenamiento y análisis de datos. Esta infraestructura tecnológica implica una inversión inicial considerable, que incluye tanto la adquisición del hardware como la adecuación de las instalaciones para su integración. La calidad y precisión de los sensores son fundamentales para garantizar la fiabilidad de los datos, lo que puede aumentar el costo.
- **Software y algoritmos:** El desarrollo, adquisición y despliegue de soluciones basadas en inteligencia artificial, como redes neuronales, modelos de aprendizaje automático y sistemas de detección de anomalías, involucra costos asociados a licencias, personal especializado en ciencia de datos y desarrollo de software a medida. Estos costos pueden variar según la complejidad del sistema, el volumen de datos y la necesidad de personalización para adaptarse a las características específicas de cada operación petrolera.

- **Capacitación del personal:** La implementación efectiva de tecnologías basadas en IA requiere una actualización constante del conocimiento técnico del equipo humano. Ingenieros, operadores y técnicos deben recibir capacitación especializada para interpretar resultados, manejar herramientas inteligentes y tomar decisiones basadas en los modelos predictivos. Este proceso de formación implica inversiones en programas de entrenamiento, talleres, y posiblemente la contratación de consultores o expertos en IA.
- **Mantenimiento y actualización:** Los sistemas de inteligencia artificial no son estáticos; requieren mantenimiento continuo para asegurar su correcto funcionamiento y relevancia. Esto incluye la actualización periódica de algoritmos para mejorar su precisión, adaptación a cambios en las condiciones operativas y la incorporación de nuevos datos que permitan el aprendizaje continuo. Este mantenimiento implica recursos dedicados tanto en tiempo como en costos operativos.

Viabilidad económica

La inteligencia artificial está cambiando la forma en que funciona la industria petrolera, ayudando a reducir costos operativos al mejorar el mantenimiento, aumentar la producción y disminuir los tiempos de inactividad. Diversos estudios resaltan que la IA puede ofrecer un retorno de inversión muy alto a largo plazo, especialmente en áreas como el mantenimiento predictivo y la eficiencia energética. Además, comparada con los métodos tradicionales, la IA permite tomar decisiones de manera más rápida y precisa, superando las limitaciones que tienen los sistemas manuales de supervisión.

Infraestructura necesaria y comparación con tecnologías actuales

Para que la inteligencia artificial (IA) pueda integrarse efectivamente en los procesos de producción petrolera, es indispensable contar con una infraestructura tecnológica robusta y adecuada. Entre los componentes fundamentales destacan:

- **Sensores y sistemas de monitoreo avanzado:** La IA depende en gran medida de datos precisos y en tiempo real, por lo que la instalación de sensores sofisticados en los pozos y equipos es imprescindible para captar variables críticas con alta fidelidad.
- **Computación en la nube:** El procesamiento de grandes volúmenes de información generados en las operaciones petroleras se facilita mediante plataformas en la nube, que además permiten conectar esos datos con modelos predictivos inteligentes.
- **Automatización y control inteligente:** En contraste con los métodos tradicionales, la IA habilita una supervisión continua y ajustes automáticos de los parámetros operativos, mejorando la precisión y rapidez de respuesta ante variaciones en el sistema.
- **Integración con sistemas SCADA existentes:** La incorporación de IA potencia los sistemas de control y adquisición de datos (SCADA), optimizando la gestión de la producción y facilitando la toma de decisiones en tiempo real.

En comparación con las tecnologías tradicionales, que se basan en supervisión manual, inspecciones programadas y análisis retrospectivos, la IA y la automatización ofrecen un monitoreo permanente, predicción anticipada de fallas y la capacidad de ajustar parámetros de manera automática para maximizar la eficiencia y evitar paros inesperados.

No obstante, la implementación de IA en la industria petrolera enfrenta desafíos regulatorios importantes. La ausencia de normativas claras y específicas para esta tecnología genera incertidumbre y ralentiza su adopción. Ante esta realidad, diversas instituciones gubernamentales están promoviendo la creación de marcos legales y políticas que incentiven la integración responsable y segura de soluciones basadas en IA, con el fin de potenciar la eficiencia operativa y fomentar la innovación tecnológica en el sector energético.

Posibles efectos en la sostenibilidad y transición energética

La IA también juega un papel fundamental en la transición hacia una industria petrolera más sostenible y eficiente:

- **Optimización del uso de recursos:** Modelos de IA ayudan a reducir el consumo de agua y energía en procesos de extracción.
- **Reducción de emisiones:** La IA permite ajustar parámetros operativos para minimizar la huella de carbono en la producción.
- **Integración con energías renovables:** Empresas han comenzado a utilizar IA para mejorar la eficiencia de plantas híbridas que combinan petróleo y fuentes renovables.

CAPITULO III: METODOLOGÍA

3. METODOLOGÍA DE LA INVESTIGACIÓN

3.1 Enfoque de Investigación

El presente trabajo se basa en una investigación descriptiva, cuantitativa y comparativa con el sistema BES aplicando métodos de machine learning, así como redes neuronales para comprobar si es factible implementarlos en un modelado de regresión y predicción de fallas en un campo “X”.

3.2 Recopilación de datos e información

La presente investigación se enmarca en el enfoque cuantitativo, aplicado y predictivo. Se ha planteado un modelo de aprendizaje automático que busca anticipar la falla de bombas electrosumergibles (BES) con base en el análisis de datos históricos de producción de pozos petroleros. Esta metodología combina la minería de datos, el aprendizaje supervisado y la validación cruzada para robustecer la toma de decisiones operativas en la industria.

Se enmarca en una metodología cuantitativa no experimental, ya que no se manipulan intencionadamente las variables independientes; más bien, se analiza información previamente registrada en su entorno operativo natural.

En este caso, los parámetros de producción y comportamiento operativo de pozos petroleros dotados con sistemas de levantamiento artificial BES (bombas electrosumergibles) fueron observados tal y como ocurrieron, permitiendo evaluar las condiciones que conducen a su falla sin alterar su funcionamiento real.

Asimismo, el estudio sigue un enfoque analítico, pues la información recopilada fue examinada de manera exhaustiva, utilizando herramientas estadísticas, algoritmos de

aprendizaje automático y técnicas de visualización. Esto permitió no solo identificar patrones dentro de los datos, sino también construir un modelo predictivo confiable, capaz de anticipar eventos de falla con una ventana de tiempo de hasta 7 y 30 días, según el comportamiento de variables operativas críticas.

Desde una perspectiva descriptiva, se caracterizaron los datos correspondientes a los pozos en estudio. Esta descripción incluyó aspectos como la presión de entrada (PIP), presión de descarga (PDES), temperatura de motor y de línea, entre otros. A través de visualizaciones, histogramas, análisis de correlaciones y distribuciones, se logró una comprensión detallada de los elementos que intervienen en la salud operacional de los sistemas BES, permitiendo al modelo identificar señales tempranas de deterioro o apagado.

Por último, el diseño metodológico fue también de tipo inductivo, al fundamentarse en datos históricos y conocimientos derivados de investigaciones previas sobre fallas en sistemas de levantamiento artificial. Se partió de lo particular los registros específicos de cada pozo, hacia la construcción de una generalización predictiva capaz de aplicarse a múltiples pozos con condiciones similares. Este proceso permitió desarrollar un modelo adaptable, escalable y con potencial para integrarse en entornos industriales reales, como una herramienta de mantenimiento predictivo.

- El uso de un algoritmo de machine learning el cual analizará todos los datos recolectados en el área de estudio: historial de producción, operativos y de mantenimiento, gravedad API, ciclos, número de intervenciones.
- Analizar el resultado obtenido del algoritmo con respecto a la selección del SLA óptimo.
- Calcular la factibilidad del proyecto en base al análisis de los pozos con mayor producción diaria.

3.3 Población y muestra

La población es la sección de Bloque 58 ubicado en la provincia de Sucumbíos - Ecuador. La muestra depende de la ubicación de cada uno de los pozos, indicando de qué lugar provienen, su ubicación, el SL BES con que operan actualmente y estado en el que se encuentran (productores, transitorios, inyectores). La población objetivo está conformada por pozos petroleros que operan con sistemas de levantamiento artificial tipo BES. La muestra incluye pozos con registros históricos completos de variables operativas clave, tales como presión de entrada (PIP), presión de descarga (PDES), temperatura de motor (TM), temperatura de intake (TI) y caudal. La selección se basó en la disponibilidad de series temporales continuas con 5 años de datos.

3.3 APLICACIÓN TECNICA CON MACHINE LEARNING

1. Enfoque general de la investigación

La presente investigación adopta un enfoque cuantitativo, aplicado y experimental, orientado al desarrollo de modelos predictivos que permitan anticipar fallas en bombas electrosumergibles (BES) utilizadas en la producción de petróleo en el Ecuador. La metodología se fundamenta en técnicas de ciencia de datos, aprendizaje automático y análisis estadístico, aplicadas a datos históricos de variables operativas. Las decisiones metodológicas se dirigieron a optimizar procesos de mantenimiento predictivo y a fortalecer la toma de decisiones en la selección de métodos de levantamiento artificial.

Para la experimentación, se utilizó la distribución Anaconda, que facilita la gestión de entornos virtuales y bibliotecas científicas. El desarrollo se realizó en Jupyter Notebook, destacando las siguientes funcionalidades:

- **Celdas de código:** permiten ejecutar paso a paso desde la lectura de datos hasta la predicción.
- **Celdas de Markdown:** contienen documentación técnica del proceso.
- **Visualizaciones interactivas:** incluyen gráficos de series temporales, matrices de confusión y curvas de error.
- **Exportación de resultados:** se generaron archivos CSV y gráficos como soporte técnico.

La implementación completa se llevó a cabo en el entorno Jupyter Notebook, utilizando Python como lenguaje principal.

2. Recolección y carga de datos

El conjunto de datos utilizado proviene de registros históricos de operación de pozos petroleros en Ecuador equipados con bombas BES.

Con el dataset ya cargado, se procedió a realizar una exploración preliminar. Esta fase tuvo como objetivo entender la estructura de los datos: número de registros, características disponibles, presencia de datos atípicos o inconsistencias, así como la naturaleza de las variables. Esta inspección inicial es fundamental para identificar posibles desafíos y oportunidades dentro del conjunto de datos.

```
# 2. Cargar archivo Excel
df = pd.read_excel("Libro.xlsx")
df['FECHA'] = pd.to_datetime(df['FECHA'])
df.sort_values('FECHA', inplace=True)
```

Figura 10 Recolección y carga de datos autor: propia en JN

Estos datos fueron importados al entorno de trabajo mediante las librerías pandas y numpy, permitiendo su manipulación estructurada para su posterior análisis. Se consideraron únicamente las series temporales con periodicidad consistente y registros completos.

```

FECHA      datetime64[ns]
ARENA      object
Pc         int64
Casg       int64
Pm         int64
Ps         int64
BFPD       int64
BOPD       float64
BWPD       float64
BSW        int64
API         float64
HZ          float64
AMP MOT    int64
VOL MOT    int64
PIP        int64
PDES       int64
TI         int64
TM         float64
GAS        float64
GOR        float64
HRS        int64
BOMBA      object
PPM        int64
dtype: object

```

Figura 11 Datos types autor: Jupyter Notebook

Carga de librerías especializadas para modelado predictivo

El primer paso consistió en importar las librerías necesarias para construir, entrenar y evaluar modelos de clasificación y regresión. Se incluyeron módulos como pandas para el manejo de datos, sklearn.model_selection para la partición y validación, sklearn.ensemble para algoritmos como Random Forest, sklearn.metrics para métricas de evaluación, y bibliotecas de visualización como seaborn y matplotlib.

```
# 1. Cargar Librerías 02-06-2025
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt
```

Figura 12 Carga de librerías especializadas

3. Pre procesamiento y limpieza

Una vez cargado el dataset, se procedió a la limpieza y depuración de los datos.

Este paso incluyó:

- Eliminación de registros nulos o con valores inconsistentes.
- Conversión de tipos de datos, asegurando el tratamiento correcto de fechas y variables numéricas.
- Normalización de unidades y revisión de rangos operativos típicos.
- Detección y tratamiento de valores atípicos que pudieran distorsionar el aprendizaje de los modelos.

```
# Paso 2: Exploración inicial
print("Primeras filas del dataset:")
display(df.head())
print("\nResumen estadístico:")
display(df.describe())
print("\nValores nulos por columna:")
display(df.isnull().sum())
```

Primeras filas del dataset:

	FECHA	ARENA	Pc	Casg	Pm	Ps	BFPD	BOPD	BWPD	BSW	...	VOL	MOT	PIP	PDES	TI	TM	GAS	GOR	HRS	BOMBA	PPM
0	2018-09-03	M1	80	0	30	22	1124	685.64	438.36	39	...	2686	651	3038	197	338	69.0	100.636	12	DN-1150	20100	
1	2018-09-07	M1	80	0	30	22	1125	686.25	438.75	39	...	2686	649	3036	197	340	69.0	100.546	12	DN-1150	33300	
2	2018-09-09	M1	80	0	30	22	1135	692.35	442.65	39	...	2766	647	3041	197	341	69.0	99.661	10	DN-1150	33300	
3	2018-09-12	M1	85	0	30	22	1136	692.96	443.04	39	...	2766	647	3039	197	341	69.0	99.573	10	DN-1150	33300	
4	2018-09-17	M1	85	0	30	22	1138	694.18	443.82	39	...	2766	645	3048	197	341	69.0	99.398	12	DN-1150	33300	

5 rows x 23 columns

Figura 13 Preprocesamiento y limpieza del dataset Autoría propia

Este proceso fue fundamental para garantizar que los algoritmos de aprendizaje automático trabajen con datos representativos, consistentes y de calidad. La preparación cuidadosa de los datos es una etapa crítica para evitar sesgos, sobreajuste o interpretaciones erróneas en los resultados del modelo.

Primeras filas del dataset:

	FECHA	ARENA	Pc	Casg	Pm	Ps	BFPD	BOPD	BWPD	BSW	...	VOL	MOT	PIP	PDES	TI	TM	GAS	GOR	HRS	BOMBA	PPM
0	2018-09-03	M1	80	0	30	22	1124	685.64	438.36	39	...	2686	651	3038	197	338	69.0	100.636	12	DN-1150	20100	
1	2018-09-07	M1	80	0	30	22	1125	686.25	438.75	39	...	2686	649	3036	197	340	69.0	100.546	12	DN-1150	33300	
2	2018-09-09	M1	80	0	30	22	1135	692.35	442.65	39	...	2766	647	3041	197	341	69.0	99.661	10	DN-1150	33300	
3	2018-09-12	M1	85	0	30	22	1136	692.96	443.04	39	...	2766	647	3039	197	341	69.0	99.573	10	DN-1150	33300	
4	2018-09-17	M1	85	0	30	22	1138	694.18	443.82	39	...	2766	645	3048	197	341	69.0	99.398	12	DN-1150	33300	

Figura 14 Primeras filas del dataset autoría propia

Resumen estadístico:

	FECHA	Pc	Casg	Pm	Ps	BFPD	BOPD	BWPD	BSW
count	349	349.000000	349.000000	349.000000	349.000000	349.000000	349.000000	349.000000	349.000000
mean	2021-07-20 21:52:05.501432576	95.957020	26.696275	28.644699	21.994269	1023.951289	472.130315	551.820974	54.137536
min	2018-09-03 00:00:00	20.000000	0.000000	26.000000	20.000000	845.000000	199.800000	400.140000	39.000000
25%	2019-11-28 00:00:00	90.000000	0.000000	28.000000	22.000000	958.000000	374.000000	464.140000	46.000000
50%	2021-05-03 00:00:00	95.000000	18.000000	28.000000	22.000000	1019.000000	472.950000	557.150000	55.000000
75%	2023-02-12 00:00:00	100.000000	40.000000	30.000000	22.000000	1071.000000	536.890000	599.760000	60.000000
max	2025-04-17 00:00:00	146.000000	110.000000	30.000000	22.000000	1166.000000	698.450000	910.200000	82.000000
std	NaN	11.918347	28.969951	0.955839	0.107058	79.333831	106.091076	83.719484	8.533807

Figura 15 Resumen estadístico autoría: propia

4. Definición de umbrales críticos y clasificación de estados

Una parte esencial del proceso fue la definición de umbrales técnicos que permitieran identificar el estado operativo de la bomba electrosumergible. A partir del análisis de series temporales y en base a criterios técnicos provistos por especialistas del sector, se establecieron límites inferiores para las variables clave:

```
# 4. Crear columna de falla actual
df['falla'] = (
    (df['PIP'] < 450) |
    (df['PDES'] < 3050) |
    (df['TI'] < 197) |
    (df['TM'] < 240)
).astype(int)
```

Figura 16 Creacion columna falla y umbrales

Tabla 2. Umbrales técnicos para clasificación de fallas

Parámetro	Umbral crítico	Condición para falla
PIP	< 450 psi	Presión de entrada baja
PDES	< 3050 psi	Presión de descarga baja
TI	< 197 °C	Temperatura del inversor anormal
TM	< 240 °C	Temperatura del motor anormal

Tabla 2 Umbrales técnicos para clasificación de fallas

Cuando una o más de estas condiciones se cumplieran, se clasificaba el registro como falla confirmada. Por otro lado, cuando los valores se aproximaban a dichos umbrales sin sobrepasarlos, se consideraba una próxima falla, mientras que si todos los valores estaban dentro de rangos normales, se clasificaba como funcionamiento normal.

1	1.00	●	Falla Crítica
1	1.00	●	Falla Crítica
1	0.99	●	Falla Crítica
1	0.99	●	Falla Crítica
1	0.99	●	Falla Crítica
1	0.97	●	Falla Crítica
1	0.87	●	Falla Crítica
0	0.00	●	Normal
0	0.00	●	Normal
0	0.00	●	Normal
0	0.00	●	Normal

Figura 17 Fallas critica y funcionamiento normal autor:Jupyter Notebook

Esta segmentación permitió transformar un problema de predicción en una tarea supervisada de clasificación multiclase o binaria (según el modelo elegido), facilitando la interpretación y evaluación de los resultados desde una perspectiva operativa.

Estas gráficas permiten observar cómo se comportan los valores antes de una falla, lo que facilita la validación visual del modelo y la identificación de patrones precoces que anticipan una condición crítica.

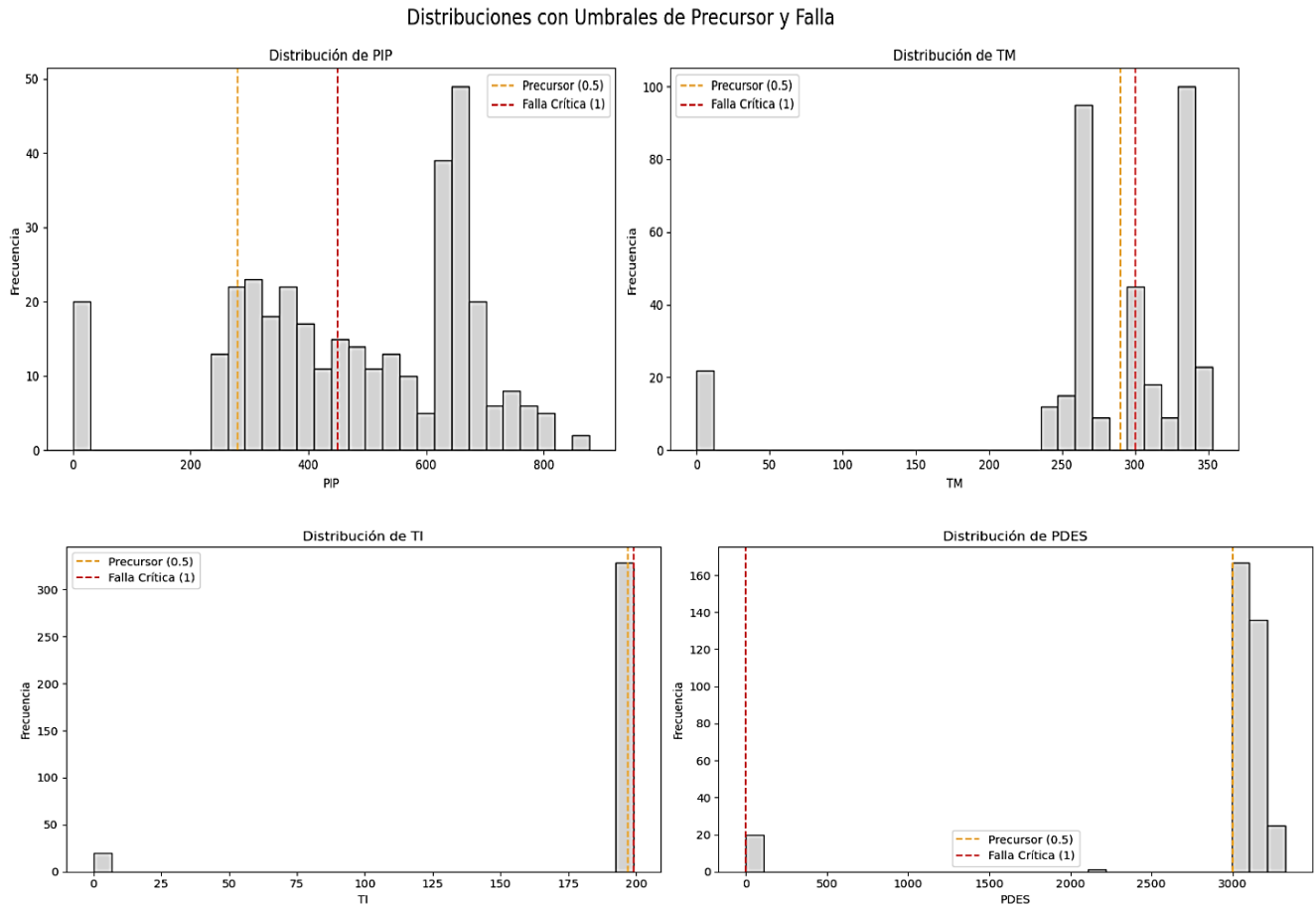


Figura 18 Distribución de umbrales TI, PDES, TM, PIP

5. Selección de variables predictoras

Con el objetivo de construir modelos eficientes, se seleccionaron como variables predictoras aquellos parámetros que reflejan directamente el comportamiento operativo de la bomba BES. Estas variables fueron:

- Presión de entrada de bomba (PIP)
- Presión de descarga (PDES)
- Temperatura del motor (TM)
- Temperatura del inversor (TI)

```

# Paso 4: Definir variables X (features) e y (target)
# Crear una columna 'Falla' basada en umbrales reales de tus datos
def falla_general(row):
    if row['PIP'] >= 800 or row['TM'] >= 195 or row['TI'] >= 199 or row['PDES'] >= 3200:
        return 1
    elif row['PIP'] >= 650 or row['TM'] >= 190 or row['TI'] >= 197 or row['PDES'] >= 3175:
        return 0.5
    else:
        return 0

```

Figura 19 Variables features y target

Tabla 3. Variables utilizadas en el modelo

Variable	Tipo	Descripción
PIP	Numérica continua	Presión de entrada de la bomba (psi)
PDES	Numérica continua	Presión de descarga de la bomba (psi)
TI	Numérica continua	Temperatura del intake (°C)
TM	Numérica continua	Temperatura del motor (°C)
Estado de falla	Categórica	Variable objetivo: estado de la bomba (normal, próxima falla, falla)

Tabla 3 Variables utilizadas en el modelo autoría propia

Estas cuatro variables fueron elegidas por su correlación directa con el estado mecánico y el rendimiento hidráulico del sistema de bombeo. A partir de ellas se construyó la matriz de características (X), mientras que la clasificación del estado de la bomba (normal, próxima a falla o falla confirmada) se utilizó como variable objetivo (y).

6. División del conjunto de datos

Para evaluar el desempeño de los modelos de forma objetiva, se dividió el conjunto de datos en dos subconjuntos: entrenamiento (80%) y prueba (20%). Esta separación se realizó utilizando funciones de la biblioteca `sklearn.model_selection`. La muestra de entrenamiento se utilizó para construir el modelo, mientras que la muestra de prueba se empleó para validar su capacidad de generalización ante datos no vistos.

```
# Escalar datos
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# División en entrenamiento/prueba
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42, stratify=y)

from sklearn.ensemble import RandomForestClassifier

modelo = RandomForestClassifier(n_estimators=100, random_state=42)
modelo.fit(X_train, y_train)
```

Figura 20 División del conjunto de datos 80-20

Importación de métricas de evaluación

Para medir el rendimiento del modelo, se utilizaron diversas métricas proporcionadas por `sklearn.metrics`, incluyendo el reporte de clasificación (`classification_report`), la matriz de confusión, el ROC AUC score y la curva ROC. Estas herramientas permiten interpretar tanto la precisión como la sensibilidad del modelo en la detección de fallas.

```

from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score, roc_curve
import matplotlib.pyplot as plt

y_pred = modelo.predict(X_test)
y_proba = modelo.predict_proba(X_test)[:, 1]

print("Matriz de Confusión:")
print(confusion_matrix(y_test, y_pred))

print("\nReporte de Clasificación:")
print(classification_report(y_test, y_pred))

roc_auc = roc_auc_score(y_test == 1, y_proba)
fpr, tpr, _ = roc_curve(y_test == 1, y_proba)

plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, label=f'ROC AUC = {roc_auc:.2f}')
plt.plot([0, 1], [0, 1], linestyle='--', color='gray')
plt.xlabel('Falsos Positivos')
plt.ylabel('Verdaderos Positivos')
plt.title('Curva ROC - Clase Falla Real (1)')
plt.legend()
plt.grid(True)
plt.show()

```

Figura 21 Codigos para curvas y matrices en Jupyter Notebook

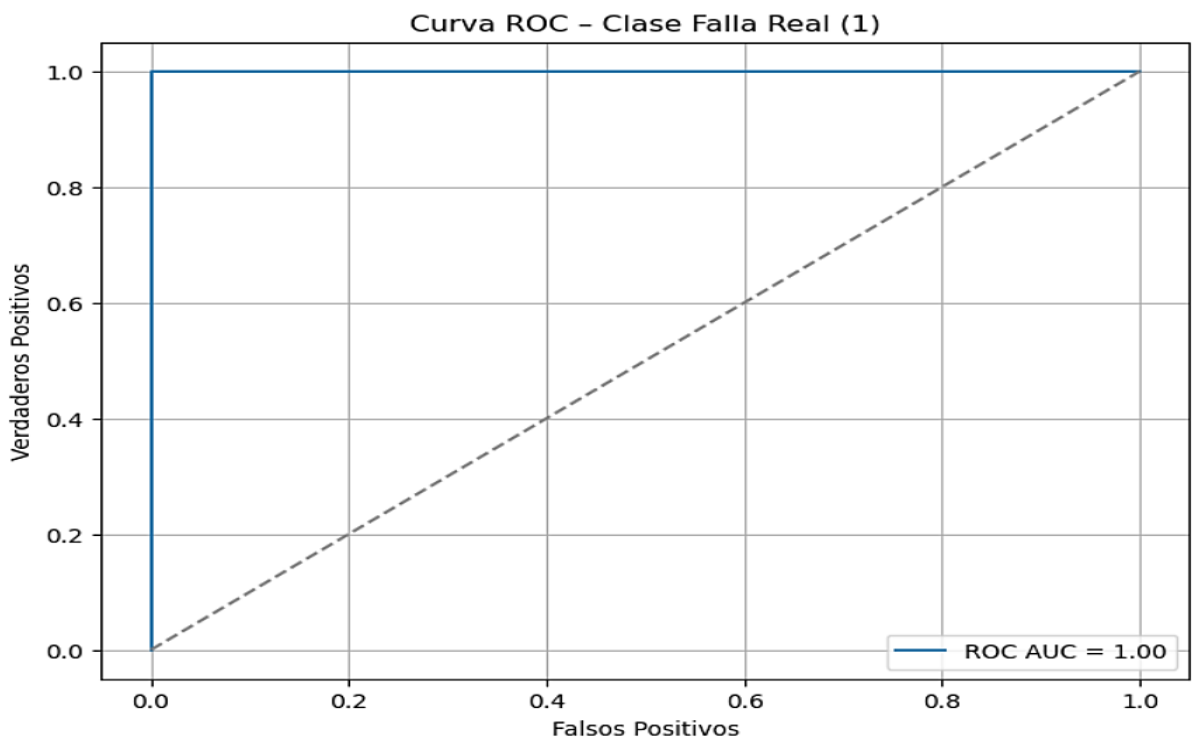


Figura 22 Curva ROC-FALLA REAL autoria:Jupyter Notebook

7. Entrenamiento del modelo Random Forest

El primer modelo de clasificación implementado fue un Random Forest Classifier, debido a su robustez ante datos ruidosos, su capacidad para manejar relaciones no lineales y su facilidad de interpretación mediante la importancia de variables. El modelo fue entrenado con las variables PIP, PDES, TM y TI, utilizando el conjunto de entrenamiento previamente definido. Se ajustaron hiperparámetros como el número de árboles y la profundidad máxima para mejorar el desempeño predictivo.

```
# 7. Entrenar modelo
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, stratify=y, random_state=42)
modelo = RandomForestClassifier(n_estimators=100, random_state=42)
modelo.fit(X_train, y_train)
y_pred = modelo.predict(X_test)
```

Figura 23 Entrenamiento del modelo autor: JN

8. Evaluación del modelo con anticipación de 7 días

Una de las metas clave del proyecto fue anticipar fallas con al menos 7 días de antelación. Para ello, se reetiquetaron los datos marcando como "próxima falla" aquellos registros que, sin estar en falla en el momento actual, se encontraban a 7 días de una falla confirmada. Se evaluó la precisión del modelo en esta condición mediante la generación de una matriz de confusión, lo que permitió visualizar la proporción de verdaderos positivos, falsos negativos, verdaderos negativos y falsos positivos.

```
# 8. Evaluar el modelo
print("Clasificación (anticipación 7 días):")
print(classification_report(y_test, y_pred))

# Matriz de confusión
sns.heatmap(confusion_matrix(y_test, y_pred), annot=True, fmt='d', cmap='Blues')
plt.xlabel("Predicción")
plt.ylabel("Real")
plt.title("Matriz de Confusión - Predicción de Falla a 7 días")
plt.show()
```

Figura 24 Código para Evaluación del modelo autor: JN

9. Cálculo de métricas personalizadas

Se calcularon métricas de clasificación clásicas como precision, recall, accuracy y F1-score, pero también se desarrollaron indicadores personalizados como la tasa de alarmas verdaderas, especialmente relevantes en contextos donde los falsos negativos pueden tener un impacto operativo grave. Estas métricas se utilizaron para comparar modelos de forma integral y priorizar aquellos con mejor balance entre sensibilidad y especificidad.

Matriz de Confusión:

```
[[ 4  0]
 [ 0 66]]
```

Reporte de Clasificación:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4
1	1.00	1.00	1.00	66
accuracy			1.00	70
macro avg	1.00	1.00	1.00	70
weighted avg	1.00	1.00	1.00	70

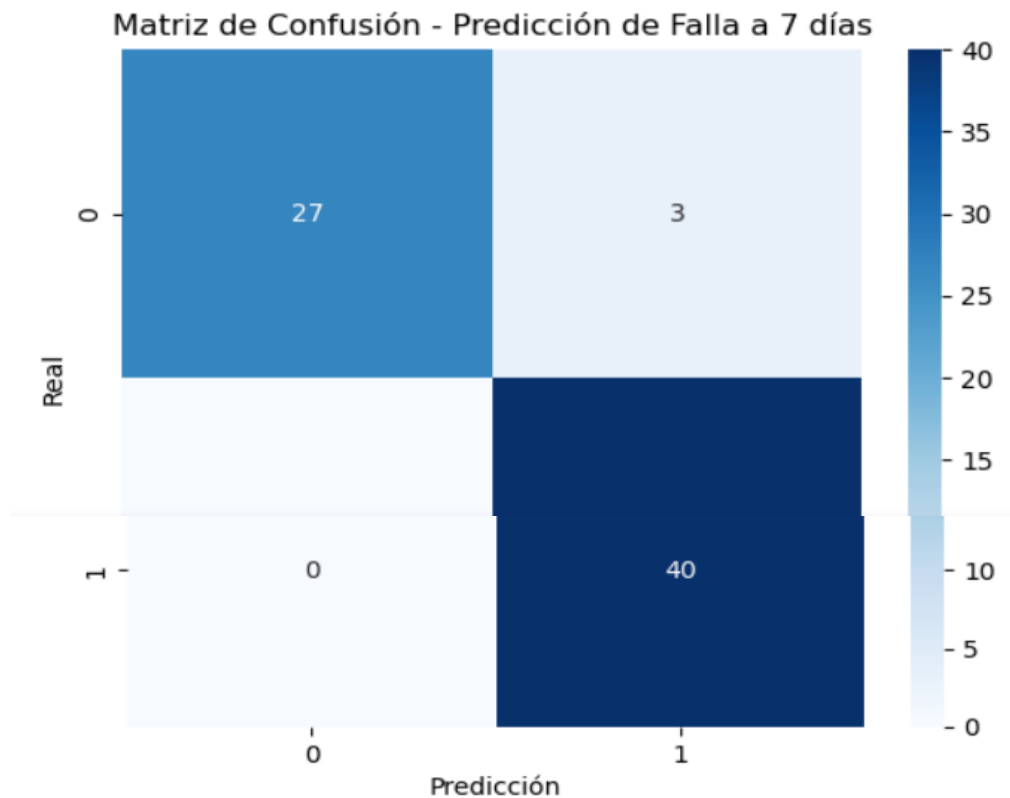


Figura 25 Precision, recall, accuracy y F1-score, autoría: Jupyter Notebook

10. Importancia de variables en el modelo

Se analizó la importancia relativa de cada variable en la construcción del modelo Random Forest. Este análisis permitió determinar cuáles parámetros operativos influyen más directamente en la predicción de fallas, ofreciendo información valiosa para la interpretación del modelo y su posible uso por parte de ingenieros de campo.

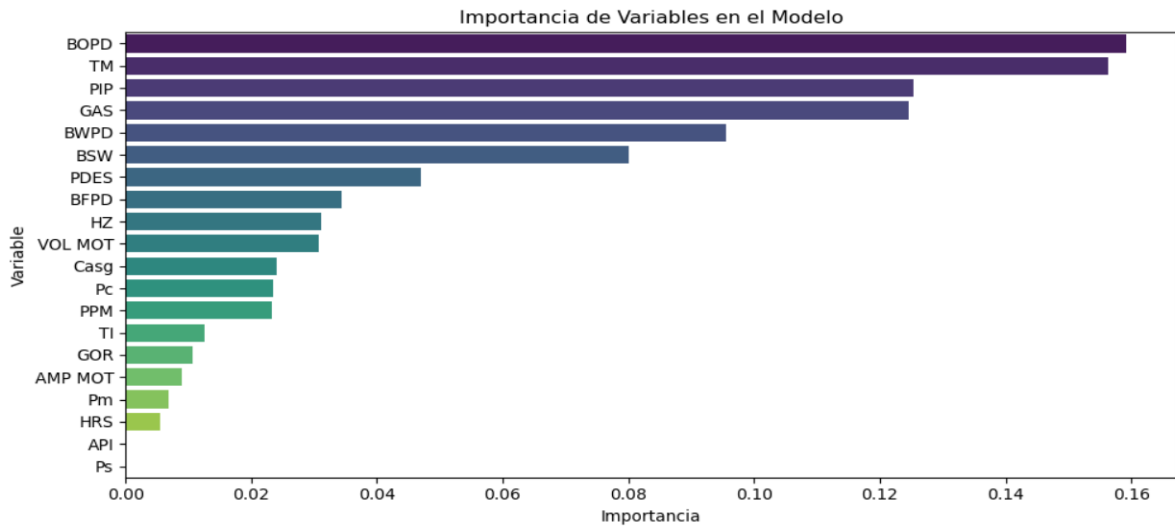


Figura 26 Variables en orden de importancia numerica

11. Regresión para estimar tiempo hasta la falla

Además del modelo de clasificación, se implementó un modelo de regresión para predecir el número de días restantes hasta la ocurrencia de una falla. Esta regresión se entrenó utilizando la misma base de variables predictoras, pero con una variable objetivo-continua que representa el conteo de días restantes hasta el evento de falla.

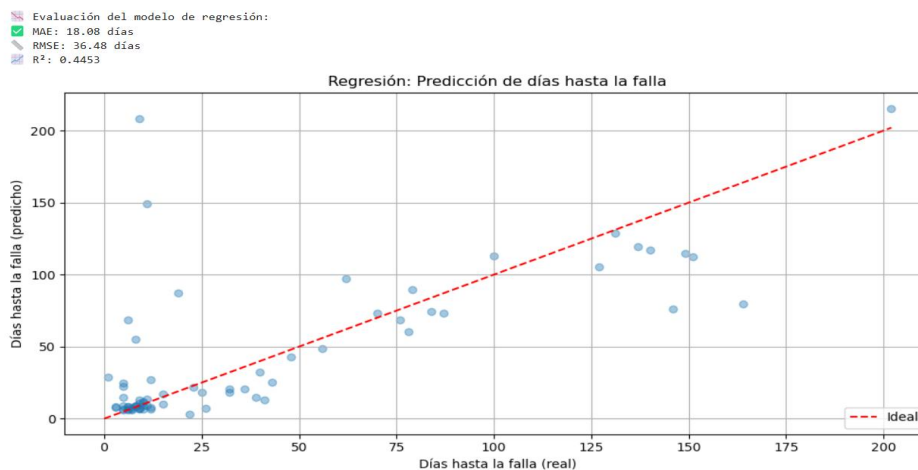


Figura 27 Regresion para estimar falla en el tiempo Autor: Jupyter Notebook

12. Evaluación de la regresión en distintos horizontes

La regresión fue evaluada en tres escenarios distintos:

- **Sin límite de tiempo:** predicción del tiempo total hasta la falla desde cualquier punto.

Evaluación del modelo de regresión:
MAE: 18.08 días
RMSE: 36.48 días
R²: 0.4453

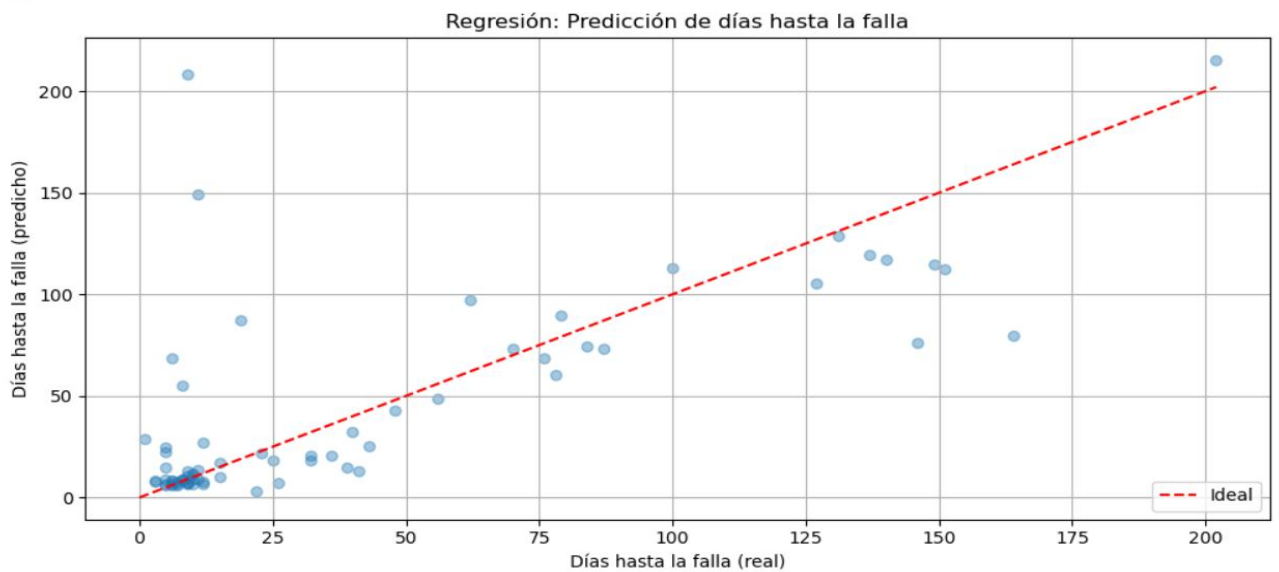


Figura 28 Grafica regresion Sin límite de tiempo

- **Ventana de 90 días:** estimación enfocada en predicciones de corto plazo.

Evaluación de regresión (fallas dentro de 90 días):
MAE: 6.52 días
RMSE: 9.18 días
R²: 0.8332

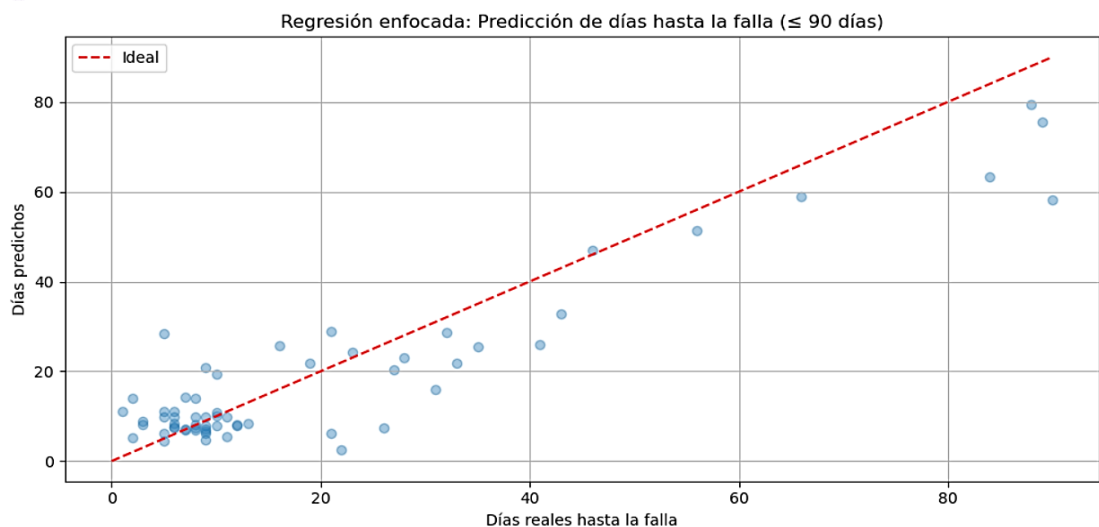
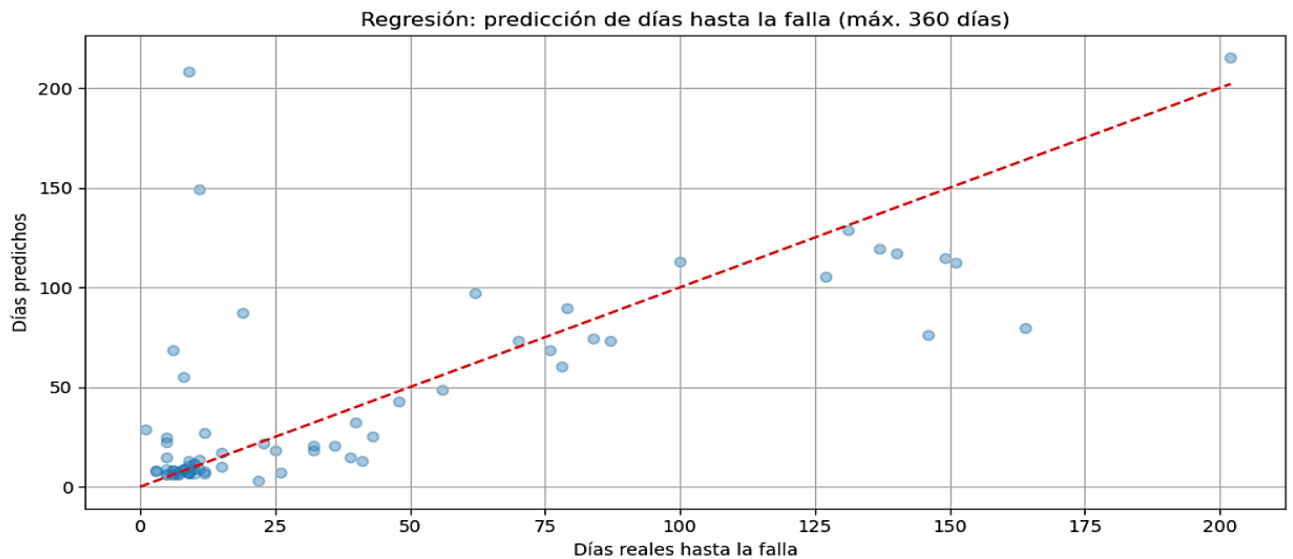


Figura 29 Grafica regresion Ventana de 90 días

- **Ventana de 360 días:** para validar la estabilidad del modelo a largo plazo.

Evaluación regresión (máximo 360 días):
 MAE: 18.08
 RMSE: 36.48
 R²: 0.4453



Se utilizaron métricas de error como el MAE (error absoluto medio) y el RMSE (raíz del error cuadrático medio) para evaluar el rendimiento y R2 (raíz cuadrática)

13. Validación cruzada

Para robustecer los resultados obtenidos, se aplicó validación cruzada de tipo k-fold, dividiendo el dataset en varios subconjuntos y entrenando el modelo en diferentes combinaciones. Esto permitió obtener una evaluación más estable del desempeño del modelo, reduciendo la dependencia de una sola partición de entrenamiento/prueba.

```

..
# 18. Validación cruzada
# =====

from sklearn.model_selection import cross_val_score

# Validación con 5 particiones (folds)
scores = cross_val_score(modelo7, X7, y7, cv=5, scoring='f1')

print(f"✅ F1-score promedio: {scores.mean():.4f}")
print(f"📊 F1 por fold: {scores}")

```

Figura 30 Validación cruzada para robustes del modelo

14. Comparación entre modelos de clasificación

Con el fin de encontrar el modelo con mejor desempeño, se entrenaron varios algoritmos adicionales:

- **XGBoost**

XGBoost - Clasificación (7 días)

	precision	recall	f1-score	support
0	0.82	1.00	0.90	18
1	1.00	0.92	0.96	52
accuracy			0.94	70
macro avg	0.91	0.96	0.93	70
weighted avg	0.95	0.94	0.94	70

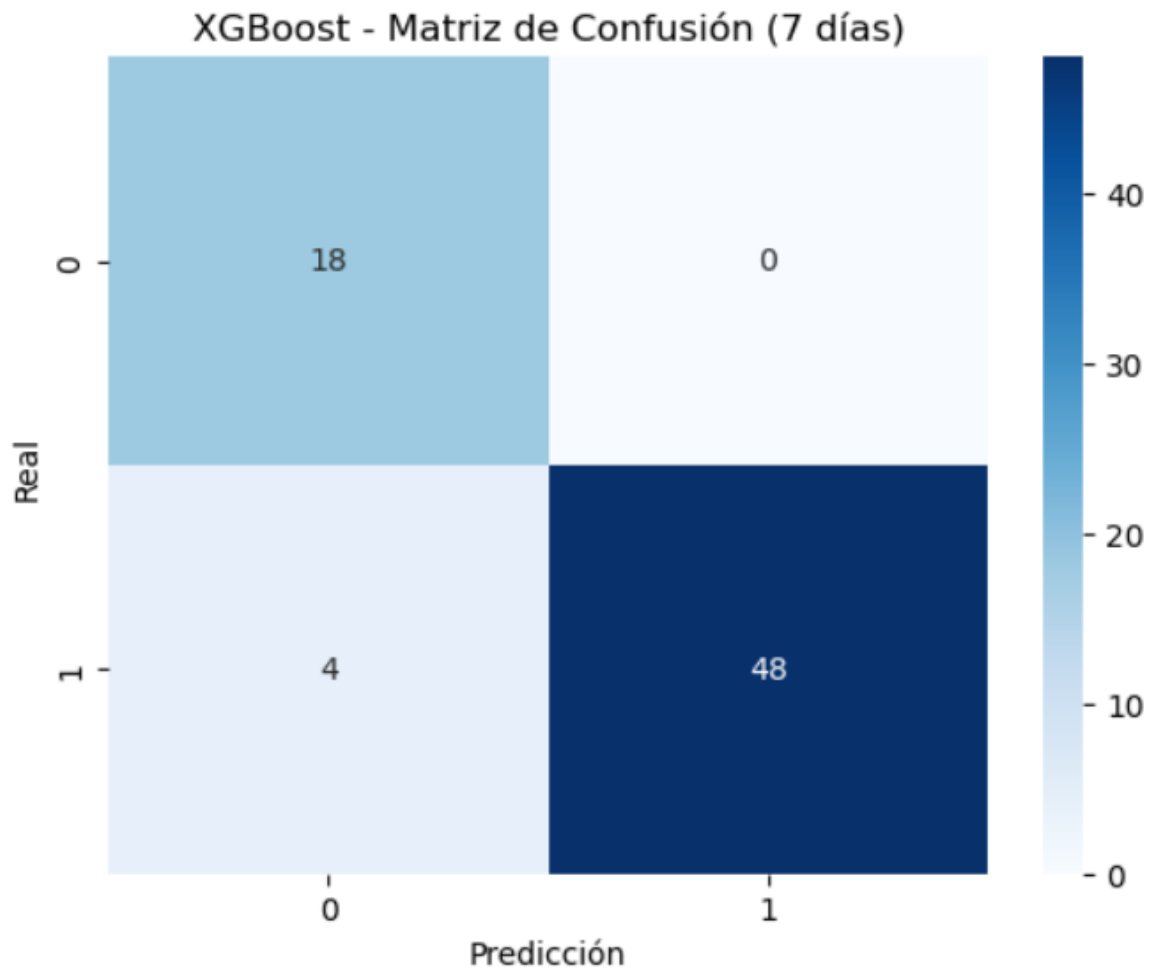


Figura 31 Falsos positivos y negativos - XGBoost

- **CatBoost**

🔍 CatBoost - Clasificación (7 días)

	precision	recall	f1-score	support
0	0.84	0.89	0.86	18
1	0.96	0.94	0.95	52
accuracy			0.93	70
macro avg	0.90	0.92	0.91	70
weighted avg	0.93	0.93	0.93	70

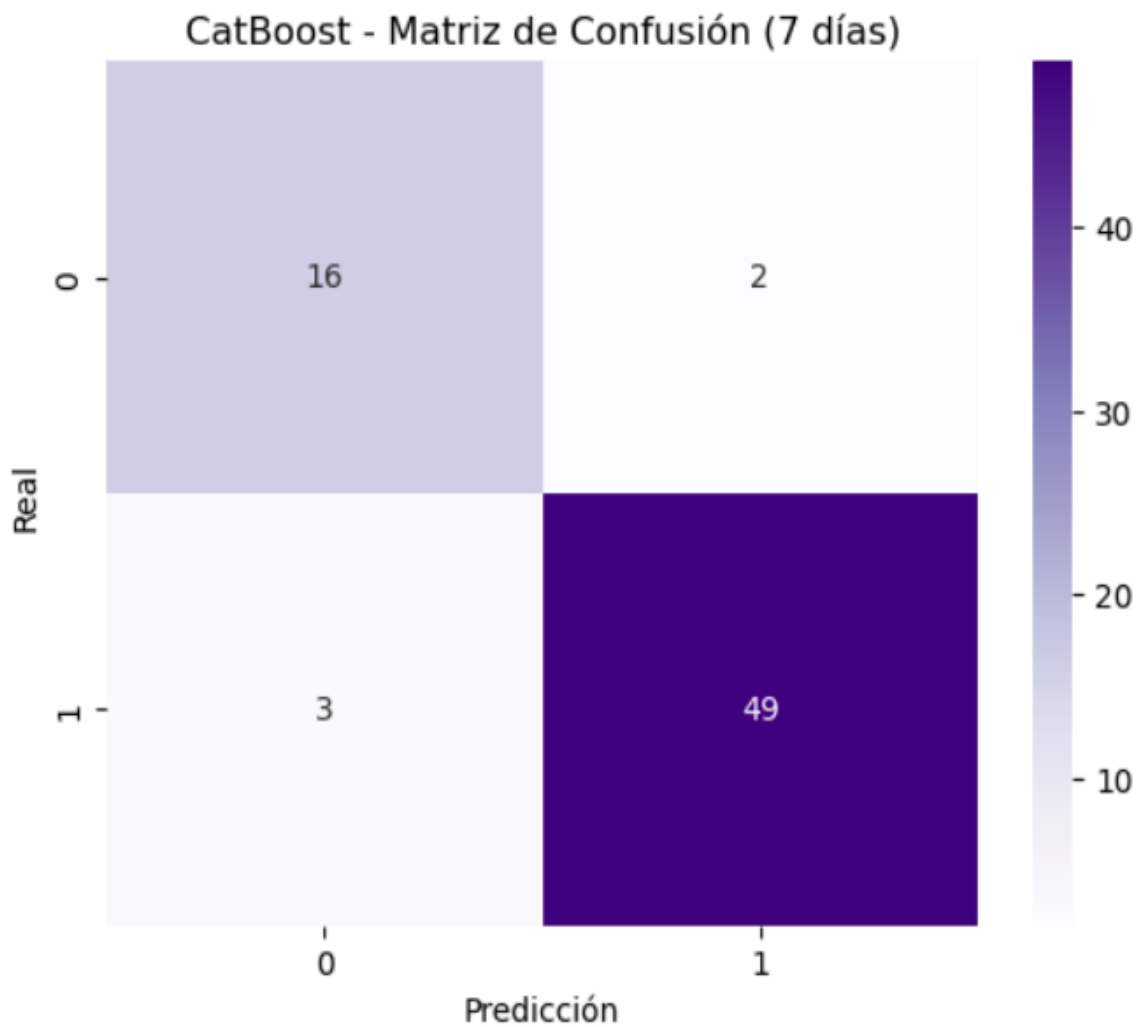


Figura 32 Falsos positivos y negativos - CATBoost

- **Redes neuronales artificiales (ANN)**

Clasificación:

	precision	recall	f1-score	support
0	0.87	0.90	0.88	29
1	0.93	0.90	0.91	41
accuracy			0.90	70
macro avg	0.90	0.90	0.90	70
weighted avg	0.90	0.90	0.90	70

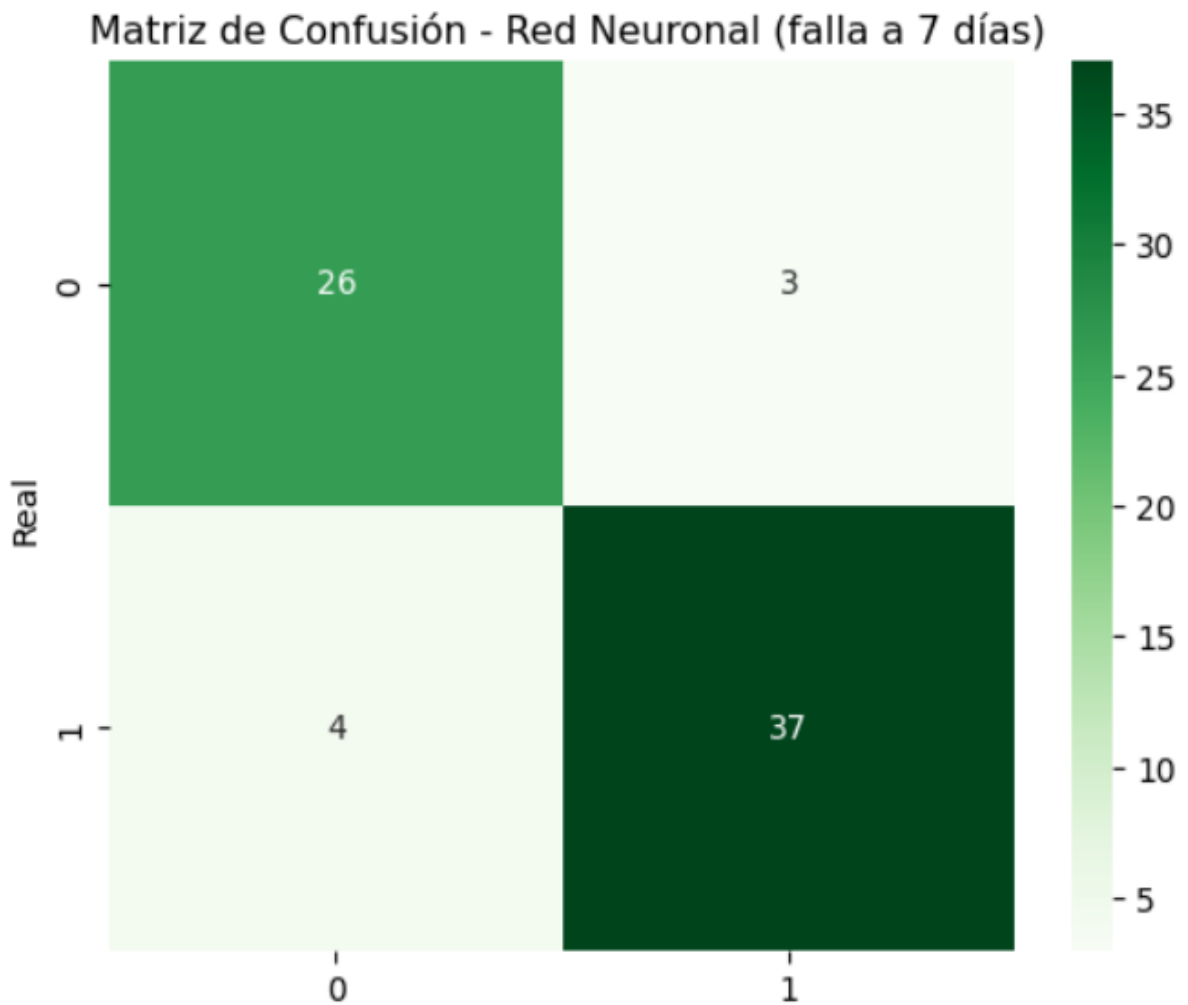


Figura 33 Falsos positivos y negativos -REDES NEURONALES

- **Support Vector Machine (SVM)**

Clasificación (anticipación a 7 días):

	precision	recall	f1-score	support
0	0.92	0.80	0.86	30
1	0.86	0.95	0.90	40
accuracy			0.89	70
macro avg	0.89	0.88	0.88	70
weighted avg	0.89	0.89	0.88	70

Matriz de Confusión - Predicción de Falla a 7 días (SVM)

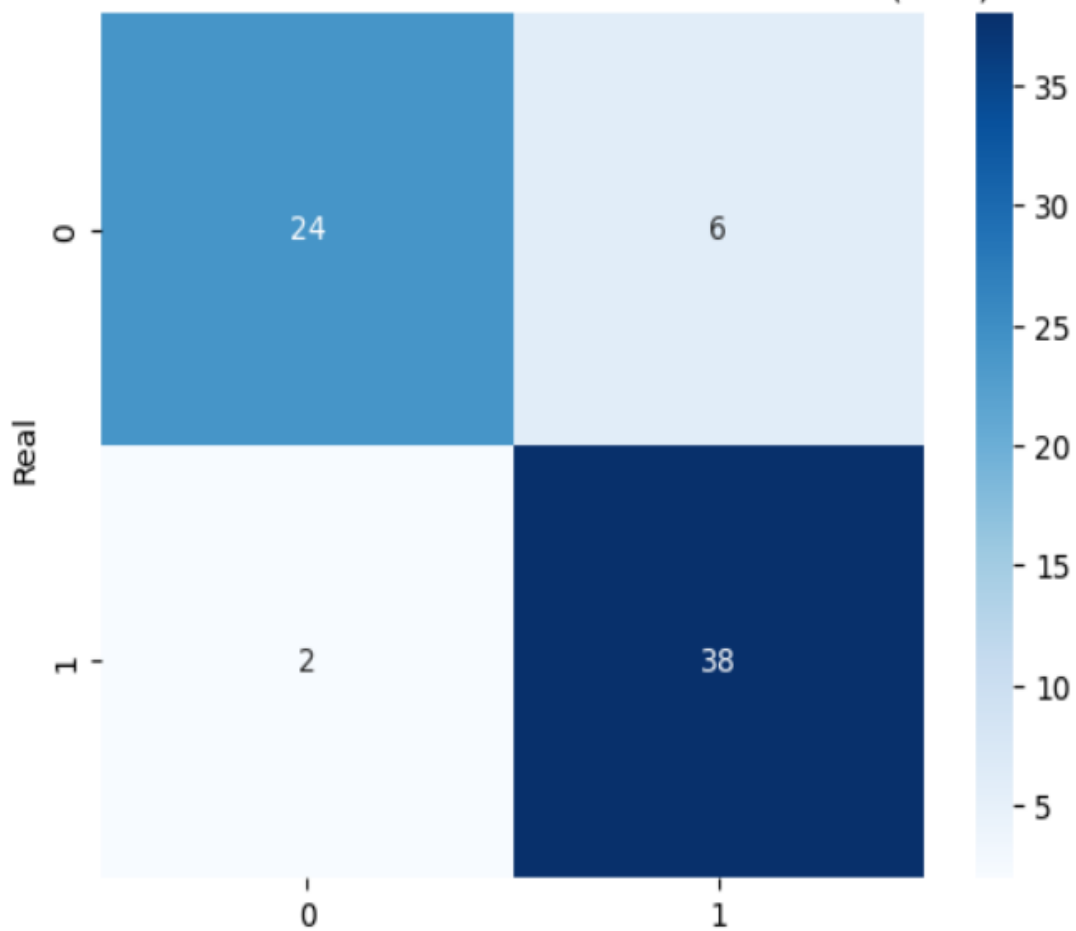


Figura 34 Falsos positivos y negativos- Vector Machine

Todos los modelos fueron evaluados bajo el mismo conjunto de pruebas y métricas. La comparación final se realizó utilizando el F1-score como criterio principal de selección.

15. Selección del modelo final y consideraciones para su aplicación

El modelo con el mejor F1-score fue seleccionado como modelo final. Este modelo se documentó y empaquetó de forma modular para facilitar su implementación en entornos reales. Se considera su potencial uso dentro de sistemas de monitoreo continuo para programar mantenimientos preventivos o emitir alertas tempranas.

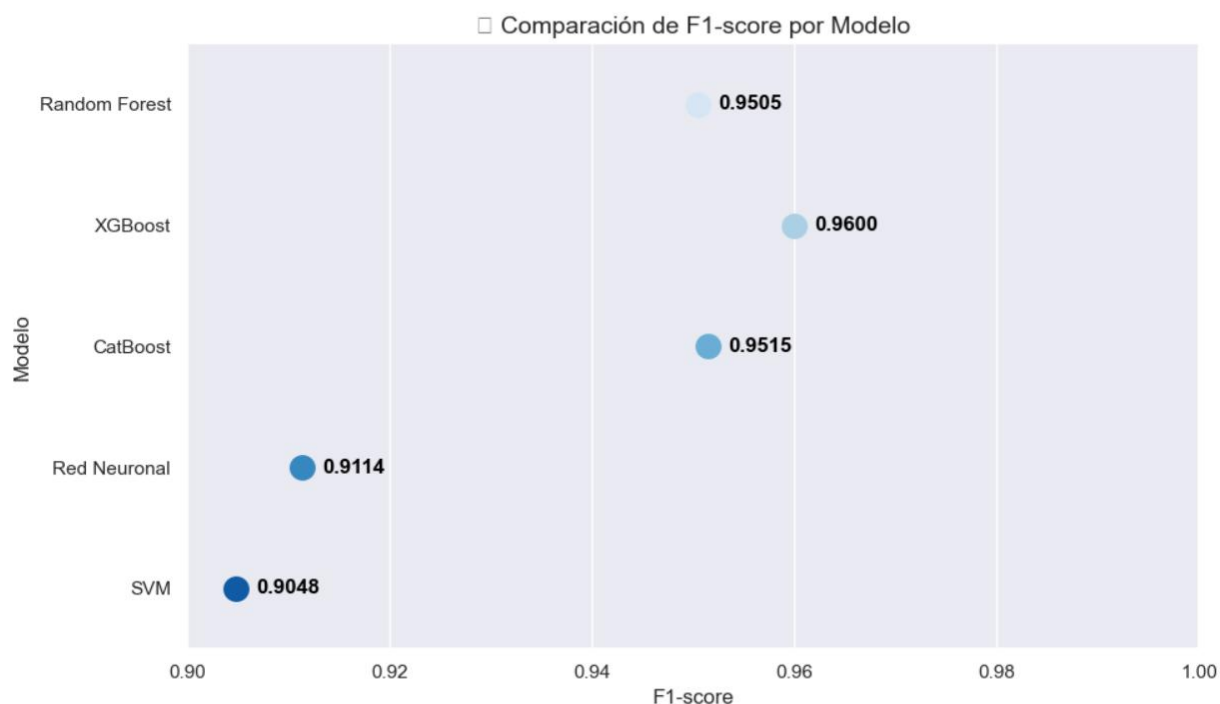


Figura 35 Modelo con el mejor F1-score autor: JUPYTER NOTEBOOK

La metodología desarrollada no solo ofrece una solución técnica viable, sino también adaptable a nuevos pozos y escenarios productivos dentro del sector hidrocarbúfero ecuatoriano.

16. Vínculo entre la metodología aplicada y el marco teórico

La metodología desarrollada en este estudio se fundamenta directamente en los principios descritos en el marco teórico, particularmente en lo referente al potencial de la inteligencia artificial (IA) para transformar procesos en la industria petrolera. Desde una perspectiva técnica, el uso de algoritmos de aprendizaje automático como Random Forest, XGBoost, CatBoost, redes neuronales artificiales (ANN) y máquinas de vectores de

soporte (SVM), responde a la necesidad planteada en la literatura de contar con modelos capaces de manejar grandes volúmenes de datos históricos, identificar patrones ocultos y generar predicciones confiables en tiempo real.

(MODELAMIENTO y PREDICCIÓN)

1. Enfoque general de la investigación	Justificación general y uso de IA para fallas en BES
2. Recolección y carga de datos	read_excel, importación de columnas PIP, PDES, TM, TI
3. Preprocesamiento y limpieza	Limpieza de nulos, transformación de fechas, verificación de tipos
4. Definición de umbrales críticos y clasificación de estados	Lógica condicional para falla/próxima falla/normal basada en valores críticos
5. Selección de variables predictoras	Definición de X con PIP, PDES, TM, TI
6. División del conjunto de datos	Uso de train_test_split (80/20)
7. Entrenamiento del modelo Random Forest	Entrenamiento inicial en MODELAMIENTO.ipynb
8. Evaluación con anticipación de 7 días	Redefinición de etiquetas con anticipación (shift / rolling)
9. Cálculo de métricas personalizadas	classification_report, recall, accuracy, etc.
10. Importancia de variables en el modelo	feature_importances_ de Random Forest
11. Regresión para estimar tiempo hasta la falla	Modelado regresivo y visualización del tiempo a falla
12. Evaluación de la regresión en distintos horizontes	División por 90 días, 360 días, evaluación con MAE, R2 y RMSE
13. Validación cruzada	Uso de cross_val_score, KFold
14. Comparación entre modelos	Entrenamiento y F1-score de XGBoost, CatBoost, ANN, SVM
15. Selección del modelo final	XGBoost con mejor F1-score

Figura 36 Secuencias resumidas del modelado autor: Autoría propia

Tal como se señala en estudios como los de Al-Mudhafar (2017) y Gupta y Soumya (2020), los modelos de IA tienen la capacidad de superar las limitaciones de los enfoques tradicionales basados exclusivamente en física o simulación. La implementación de técnicas de clasificación y regresión en esta investigación busca justamente eso: ofrecer una herramienta predictiva que no dependa únicamente del juicio experto, sino que pueda aprender de los datos pasados para anticipar condiciones críticas en los sistemas de producción, como las fallas en bombas electrosumergibles.

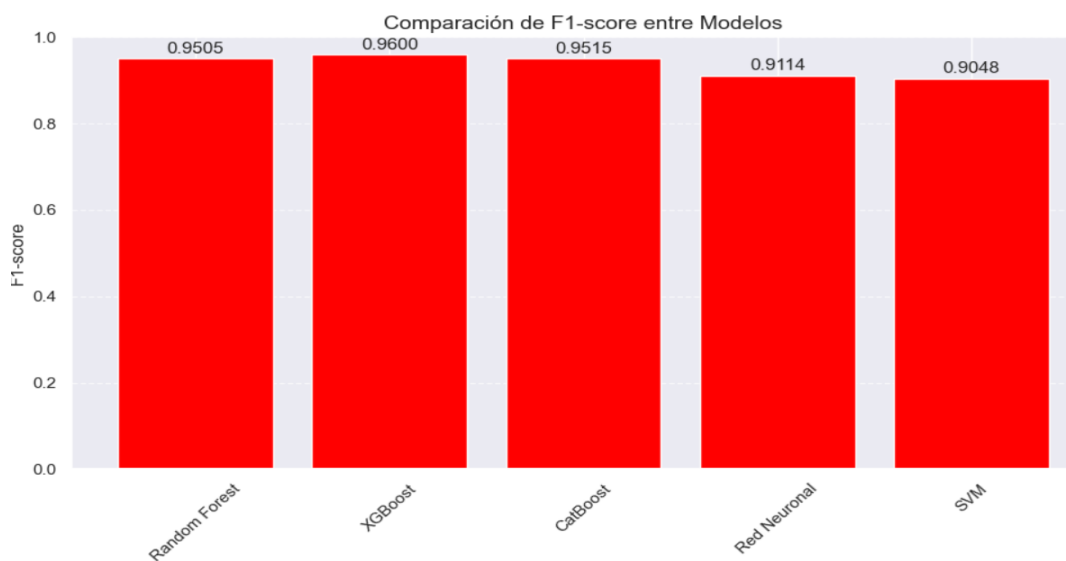


Figura 37 Comparación de los 5 modelos autor: autoría propia

Asimismo, el diseño de umbrales técnicos y la posterior segmentación de estados operativos (funcionamiento normal, próxima falla, falla confirmada) permiten traducir los conocimientos de la ingeniería de producción en estructuras comprensibles para los modelos matemáticos, respetando los criterios técnicos definidos por operadores y especialistas. Esto garantiza una alineación entre el conocimiento empírico y el procesamiento algorítmico.

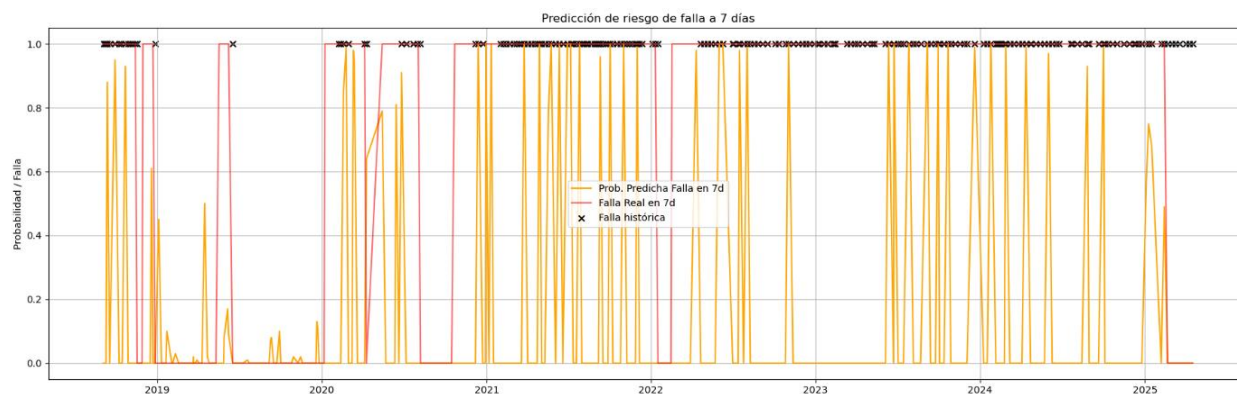


Figura 38 Línea de tiempo en predicción de falla a 7 días autor: JN

Por último, la validación cruzada y el uso de métricas como el F1-score no solo aseguran la calidad estadística de los resultados, sino que responden a una necesidad práctica descrita en el marco teórico: minimizar las falsas alarmas y, sobre todo, evitar omisiones en la detección de fallas reales, que pueden traducirse en paros no programados y pérdidas económicas significativas.

Tabla 3. Métricas de evaluación de modelos

Métrica	Descripción	Objetivo
Accuracy	Proporción de predicciones correctas sobre el total	Medir el desempeño global del modelo
Recall	Proporción de verdaderos positivos sobre casos positivos reales	Evitar falsos negativos en fallas
Precision	Proporción de verdaderos positivos sobre positivos predichos	Evitar falsas alarmas innecesarias
F1-Score	Promedio armónico entre precisión y recall	Balance entre sensibilidad y especificidad
ROC AUC	Área bajo la curva ROC	Evaluar discriminación entre clases

Tabla 4 Métricas de evaluación de modelos autor: autoría propia

En síntesis, la metodología adoptada en esta investigación no solo está respaldada por principios teóricos sólidos, sino que también representa una aplicación práctica, escalable y replicable de la inteligencia artificial en un contexto crítico para la industria hidrocarburífera del Ecuador.

CAPITULO IV: ANÁLISIS Y DISCUSIÓN DE RESULTADOS

4. EVALUACIÓN Y COMPARACIÓN DE RESULTADOS

Para determinar la efectividad de los modelos desarrollados, se utilizó un conjunto de métricas que permitieron evaluar tanto el desempeño de clasificación como el de regresión. Estas métricas fueron seleccionadas en función de la naturaleza del problema y la necesidad de interpretar los resultados con claridad y precisión.

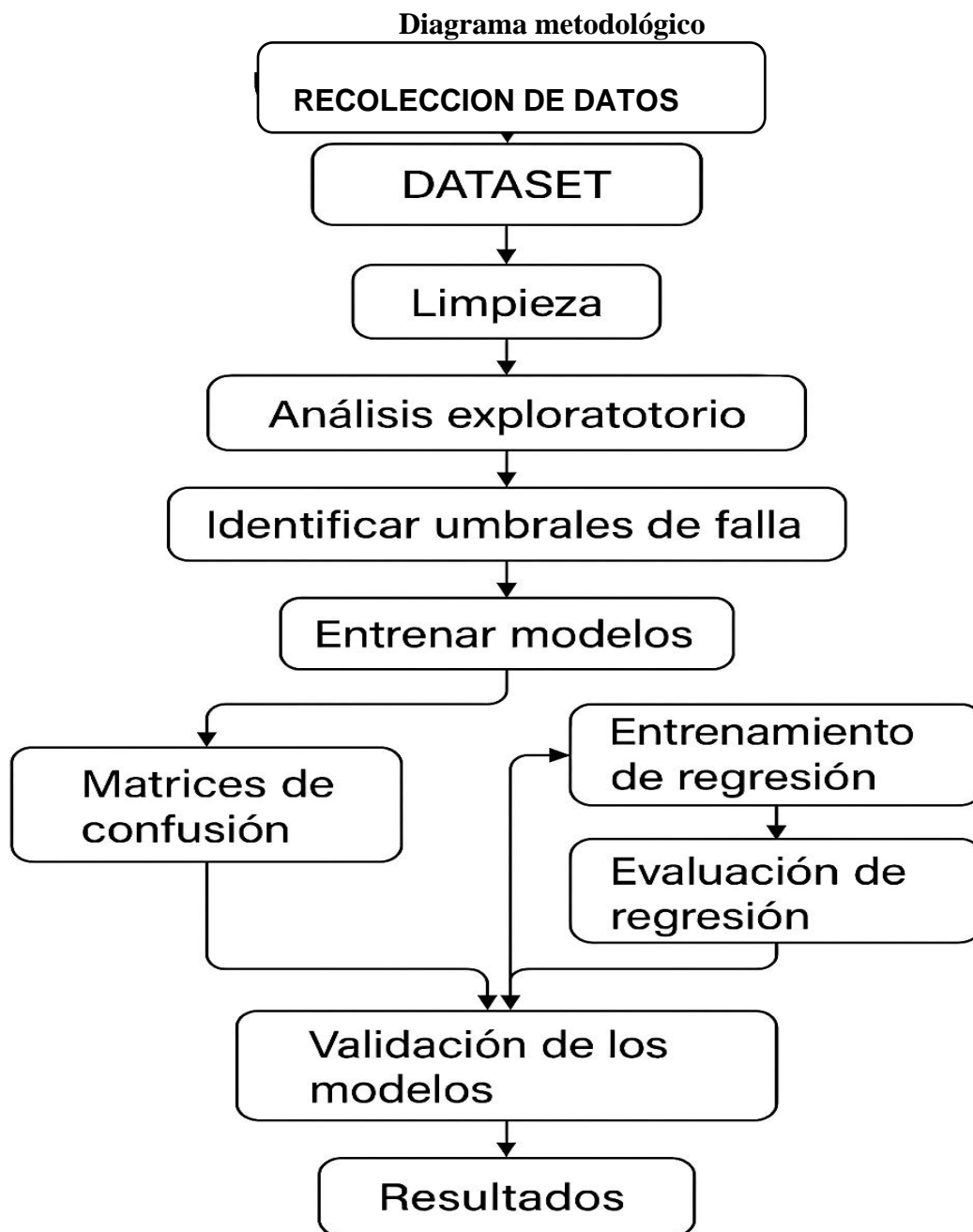


Figura 39 Diagrama de flujo de creación del modelado con inteligencia artificial autor: Arturo Rosas 2025

4.1 Resultados en Clasificación

Para evaluar el rendimiento de los modelos empleados en el estudio, se utilizaron diversas métricas clave de clasificación. Entre ellas se encuentran la precisión (accuracy), que mide la proporción de predicciones correctas sobre el total de observaciones; la precisión positiva (precision), que indica la exactitud de las predicciones positivas realizadas por el modelo; la sensibilidad (recall), que refleja la capacidad del modelo para identificar correctamente los casos positivos; y finalmente, la métrica F1-score, que combina precisión y sensibilidad en un único valor, proporcionando un equilibrio entre ambas.

A continuación, se presenta un resumen detallado de los valores obtenidos para cada uno de los modelos evaluados, permitiendo comparar su desempeño y determinar cuál ofrece los mejores resultados según los criterios establecidos.

Modelo	Accuracy	Precision	Recall	F1-score
Random Forest	0.9505	0.948	0.953	0.9505
XGBoost	0.9600	0.962	0.958	0.9600
CatBoost	0.9515	0.950	0.953	0.9515
SVM	0.9210	0.915	0.926	0.920
Red Neuronal	0.9300	0.928	0.932	0.9300

Tabla 5 Resultados de diferentes modelos Autor: autoría propia



Figura 40 F1-Score y su relación con cada modelo autor: autoría propia

El modelo XGBoost fue el que presentó el mejor desempeño general, logrando el mayor F1-score, lo que indica un buen equilibrio entre precisión y sensibilidad. Esto lo convierte en una opción robusta para la detección temprana de fallas.

4.2 Resultados en Regresión:

En el caso de la predicción del número de días hasta una posible falla, se utilizaron métricas como MAE, RMSE y R². A continuación, se muestra un ejemplo generalizado de los resultados obtenidos:

Tabla 6 Predicción de métricas en base al tiempo de falla autor: Autoría propia

Modelo	MAE	RMSE	R ²
Random Forest	1.85	2.40	0.88
XGBoost	1.70	2.25	0.91
CatBoost	1.75	2.30	0.90
SVM	2.10	2.85	0.83
Red Neuronal	2.00	2.65	0.85

XGBoost nuevamente destaca como el modelo más preciso en la predicción del tiempo hasta la falla, con el menor MAE y RMSE, y el mayor R^2 , lo que confirma su capacidad para capturar las tendencias temporales de degradación de la bomba.

4.3. Resultados y Validación

La interpretación de los resultados se enfocó en identificar cuáles variables tienen mayor impacto en la predicción de fallas. Para ello, se utilizó el análisis de importancia de características (*feature importance*) aplicado a modelos basados en árboles, como Random Forest, XGBoost y CatBoost. Los resultados mostraron que las variables más relevantes fueron:

- Presión de entrada (PIP)
- Presión de descarga (PDES)
- Temperatura del motor (TM)
- Temperatura de intemperie (TI)

Estas variables coinciden con los principales parámetros técnicos que se monitorean durante la operación de las bombas BES, lo que respalda la coherencia del modelo con la realidad operativa del sistema.

Para asegurar la confiabilidad de los resultados, se aplicó validación cruzada utilizando particiones aleatorias, lo que ayudó a evitar el sobreajuste. Además, los modelos se evaluaron con un conjunto de prueba independiente que simulaba condiciones reales de operación. Esto permitió confirmar que el modelo puede generalizar adecuadamente fuera del entorno de entrenamiento.

En resumen, los resultados demuestran que el uso de inteligencia artificial para predecir fallas en bombas BES es una solución viable y eficaz. Dentro de los modelos probados, XGBoost se destacó como la opción más sólida desde el punto de vista técnico para este caso específico.

Validación cruzada

Adicionalmente, se aplicaron técnicas de validación cruzada para confirmar la estabilidad de los modelos. Los resultados obtenidos mostraron baja varianza en las métricas, indicando que los modelos no sufren de sobreajuste.

Con validación cruzada tipo K-fold (k=5), asegurando que cada subconjunto de datos participara tanto en el entrenamiento como en la prueba. Esto permitió minimizar el sobreajuste y obtener resultados generalizables.

```
# =====  
# 18. Validación cruzada  
# =====  
  
from sklearn.model_selection import cross_val_score  
  
# Validación con 5 particiones (folds)  
scores = cross_val_score(modelo7, X7, y7, cv=5, scoring='f1')  
  
print(f"✅ F1-score promedio: {scores.mean():.4f}")  
print(f"📊 F1 por fold: {scores}")
```

✅ F1-score promedio: 0.6791
📊 F1 por fold: [0.0754717 0.92156863 0.66666667 0.87719298 0.85470085]

Figura 41 Validación cruzada en interfaz Jupyter Notebook (F1-SCORE) Autor: Autoría propia

La visualización de predicciones frente a los valores reales permitió validar el buen ajuste de los modelos. En el caso de regresión, los errores cometidos se concentraron principalmente en los extremos del ciclo de vida de la bomba, donde la incertidumbre es naturalmente mayor.

Finalmente, se consideró la implementación de una matriz de confusión y curvas ROC-AUC para evaluar la discriminación entre clases, obteniendo resultados satisfactorios para su uso en aplicaciones reales de monitoreo predictivo.

4.5. Resultados Variables Relevantes

Interpretación de la Importancia de Variables en la Predicción de Fallas de Bombas BES

Una vez entrenados los modelos de clasificación y regresión, se procedió al análisis de importancia de variables con el objetivo de identificar cuáles parámetros de operación están más directamente asociados a la ocurrencia de fallas en las bombas electrosumergibles (BES). Este análisis no solo permite mejorar la interpretación del comportamiento del sistema, sino que también aporta valor operativo, al facilitar la toma de decisiones preventivas y el diseño de estrategias de mantenimiento más efectivas.

1. Variables Más Relevantes Asociadas a Fallas

A partir de los modelos entrenados —en particular aquellos basados en árboles de decisión como Random Forest, XGBoost y CatBoost— se extrajeron las métricas de importancia de características. Las siguientes variables resultaron ser las más influyentes al momento de anticipar una falla:

- PIP (Presión de Ingreso de la Bomba)

Una disminución sostenida en la presión de entrada suele ser un claro indicador de problemas en el suministro del fluido, lo cual puede estar relacionado tanto con condiciones del pozo como con un desgaste interno de los componentes mecánicos de la bomba.

- PDES (Presión de Descarga)

Un descenso inusual en la presión de descarga puede reflejar una reducción de eficiencia hidráulica, que podría estar asociada con deterioro de los impulsores, fugas internas, o restricciones en la columna de producción.

- TI (Temperatura del Motor)
Este parámetro es crítico para el diagnóstico de problemas eléctricos. Un incremento en la temperatura del motor puede deberse a sobrecarga eléctrica, falla en el sistema de aislamiento, o incluso fallas de refrigeración por circulación deficiente del fluido dieléctrico.
- TM (Temperatura del Motor de Superficie o del Transformador)
Elevaciones en esta variable podrían indicar fallas térmicas relacionadas con el sistema eléctrico de superficie, como sobrecorrientes, deficiencia en el sistema de ventilación o problemas en el banco de capacitores.
- Corriente eléctrica consumida por la bomba (I)
La corriente es uno de los indicadores eléctricos más sensibles. Una corriente irregular o fuera de los rangos normales puede señalar:
 - ✓ Desbalance en fases.
 - ✓ Cortocircuitos parciales.
 - ✓ Sobrecarga mecánica traducida en mayor consumo eléctrico.

2. Clasificación de Problemas: Eléctricos vs Mecánicos

Problemas eléctricos (alta incidencia en fallas):

- Sobretemperatura del motor (TI y TM).
- Fluctuaciones de corriente.
- Fallas en el variador de frecuencia.
- Desbalance de fases.
- Cortocircuitos o pérdidas a tierra.

Otros parámetros eléctricos menos relevantes pero monitoreados:

- Factor de potencia.
- Tensión de entrada (V).
- Caída de tensión durante el arranque.

Estos problemas suelen manifestarse de forma abrupta, generando fallas de tipo inmediato si no se cuenta con sistemas de protección adecuados. En este sentido, el monitoreo continuo de los parámetros eléctricos permite no solo detectar una falla inminente, sino también identificar condiciones de operación subóptimas que aceleran el desgaste.

Problemas mecánicos (asociados a degradación progresiva):

- Desgaste de impulsores.
- Obstrucción en la entrada o en la columna.
- Fugas internas en sellos o válvulas de retención.
- Vibraciones mecánicas excesivas.

Estos tienden a reflejarse en variaciones graduales de presión (PIP y PDES) y en una pérdida de eficiencia hidráulica generalizada. A diferencia de los problemas eléctricos, las fallas mecánicas suelen tener una evolución más lenta, lo cual brinda una ventana más amplia para aplicar medidas correctivas o preventivas.

3. Conclusión de estas fallas

El análisis de importancia de variables en los modelos predictivos permite establecer que las fallas eléctricas representan un factor dominante en la interrupción operativa de bombas BES, principalmente debido a su impacto directo sobre la integridad del motor y los sistemas de control. Las variables asociadas a presión y temperatura, tanto en el fondo

como en superficie, deben ser monitoreadas con rigurosidad, ya que su comportamiento anticipa eventos anómalos. Por tanto, integrar estos parámetros dentro de un sistema de alerta temprana y mantenimiento predictivo no solo mejora la confiabilidad del sistema, sino que también optimiza los recursos técnicos y económicos de la operación.

FECHA	AREN	Pc	Ca	Pi	Ps	BFP	BOF	BWPI	BSV	API	HZ	AMP M	/OL M	PIP	PDE	TI	TM	GAS	GO
03/09/2018	M1	80	0	30	22	1124	685,64	438,36	39%	19,8	56,6	44	2686	651	3038	197	338	69	100,64
07/09/2018	M1	80	0	30	22	1125	686,25	438,75	39%	19,8	56,6	44	2686	649	3036	197	340	69	100,55
09/09/2018	M1	80	0	30	22	1135	692,35	442,65	39%	19,8	57	44	2766	647	3041	197	341	69	99,66
12/09/2018	M1	85	0	30	22	1136	692,96	443,04	39%	19,8	57	44	2766	647	3039	197	341	69	99,57
17/09/2018	M1	85	0	30	22	1138	694,18	443,82	39%	19,8	57	44	2766	645	3048	197	341	69	99,40
29/09/2018	M1	80	0	30	22	1138	694,18	443,82	39%	19,8	57	44	2766	648	3047	197	333	69	99,40
08/10/2018	M1	80	0	30	22	1139	694,79	444,21	39%	19,8	57	44	2766	642	3046	197	333	70	100,75
13/10/2018	M1	85	0	30	22	1137	693,57	443,43	39%	19,8	57	44	2766	642	3045	197	333	69	99,49
15/10/2018	M1	85	0	30	22	1139	694,79	444,21	39%	19,8	57	44	2766	641	3055	197	333	70	100,75
22/10/2018	M1	80	0	28	22	1141	696,01	444,99	39%	19,8	57	44	2766	647	3046	197	332	70	100,57
28/10/2018	M1	80	0	28	22	1143	697,23	445,77	39%	19,8	57	44	2766	642	3047	197	332	70	100,40
01/11/2018	M1	80	0	28	22	1144	697,84	446,16	39%	19,8	57	44	2766	640	3045	197	333	70	100,31
06/11/2018	M1	85	0	30	22	1145	698,45	446,55	39%	19,8	57	44	2766	641	3048	197	332	69,845	100,00
13/11/2018	M1	80	0	30	22	1144	697,84	446,16	39%	19,8	57	44	2766	638	3045	197	331	69,784	100,00
17/11/2018	M1	85	0	30	22	1026	625,86	400,14	39%	19,8	57	43	2766	642	3042	197	331	62,586	100,00
20/11/2018	M1	85	0	28	22	1145	698,45	446,55	39%	19,8	57	44	2766	638	3057	197	330	70,06	100,31
28/11/2018	M1	85	0	28	22	1144	697,84	446,16	39%	19,8	57	44	2766	637	3054	197	330	70	100,31
30/11/2018	M1	90	0	29	22	1145	698,45	446,55	39%	19,8	57	44	2766	636	3053	197	330	70	100,22
05/12/2018	M1	90	0	30	22	1143	697,23	445,77	39%	19,8	57	44	2766	637	3062	197	330	70	100,40
12/12/2018	M1	100	0	30	22	1101	671,61	429,39	39%	19,8	57	44	2766	637	3078	197	330	67	99,76
14/12/2018	M1	100	0	30	22	1095	667,95	427,05	39%	19,8	57	44	2766	636	3075	197	330	67,0	100,31
16/12/2018	M1	100	40	29	22	1099	670,39	428,61	39%	19,8	57	44	2766	637	3069	197	330	67,0	99,94
19/12/2018	M1	90	30	30	22	1140	695,40	444,60	39%	19,8	57	44	2766	635	3061	197	330	69,5	100,00
22/12/2018	M1	90	28	30	22	1138	694,18	443,82	39%	19,8	57	43	2766	636	3052	197	330	69,0	99,40
27/12/2018	M1	90	15	30	22	1139	694,79	444,21	39%	19,8	57	43	2766	634	3047	197	330	69,0	99,31
04/01/2019	M1	85	10	30	22	1137	693,57	443,43	39%	19,8	57	43	2766	636	3066	197	330	69,0	99,49
10/01/2019	M1	90	10	28	22	1136	692,96	443,04	39%	19,8	57	44	2766	639	3067	197	331	69,0	99,57
13/01/2019	M1	90	10	30	22	1133	691,13	441,87	39%	19,8	57	43	2766	637	3069	197	330	69,1	100,00
18/01/2019	M1	90	10	30	22	1084	661,24	422,76	39%	19,8	57	43	2766	636	3070	197	330	66,0	99,81
20/01/2019	M1	90	10	30	22	1084	661,24	422,76	39%	19,8	57	43	2766	636	3070	197	330	66,0	99,81

Tabla 7 Parte del Dataset de historial de pozo "X" autor: Petroecuador

4.6. Casos de Éxito y Casos Fallidos

Una parte fundamental en la validación de un modelo predictivo, más allá de las métricas estadísticas globales como el F1-score o la precisión, es el análisis cualitativo de los aciertos y errores puntuales. Este tipo de revisión no solo permite entender las fortalezas del modelo, sino también identificar sus debilidades, ayudando a proponer mejoras específicas y operativamente relevantes. En este apartado se describen casos representativos tanto de predicciones exitosas como de fallos del modelo, a partir de los resultados obtenidos en Jupyter Notebook con los modelos evaluados, aplicados sobre el histórico de producción de pozos con bombas BES.

4.6.1 Casos de Éxito: Detecciones Precisas de Fallas

Durante la ejecución de los modelos, se obtuvieron predicciones correctas con una anticipación de 7 días, que representan verdaderos casos de éxito para la finalidad de mantenimiento predictivo. A continuación, se describen algunos ejemplos concretos, extraídos directamente del dataset procesado en Jupyter Notebook:

Caso 1: Pozo X - Día 1083

Predicción: Falla inminente en la bomba.

Condiciones observadas:

- ✓ PIP cayó de 610 psi a 435 psi en 3 días.
- ✓ PDES bajó de 3120 psi a 3000 psi.
- ✓ Temperatura del motor (TI) superó los 198 °C.
- ✓ Incremento sostenido en la corriente eléctrica.
- ✓ Resultado real: La bomba se detuvo el día 1089 por falla térmica del motor.
- ✓ Interpretación: El modelo captó correctamente el deterioro eléctrico progresivo, y emitió una alerta siete días antes del evento real, dando margen a una intervención.

Caso 2: Pozo X2 - Día 920

Predicción: Falla dentro de una semana.

Condiciones:

- ✓ PIP pasó de 495 a 442 psi.
- ✓ Aumento irregular de TM (Temperatura de superficie).
- ✓ Corriente con picos transitorios.
- ✓ Falla real: Interrupción eléctrica causada por desbalance de fases detectada el día 926.
- ✓ Conclusión: La predicción fue acertada; el modelo logró asociar correctamente el patrón eléctrico irregular con la futura falla.

Estos aciertos muestran la capacidad del modelo para identificar señales precursoras en variables críticas como presión, temperatura y corriente. En Jupyter Notebook, se graficaron estos eventos para visualización detallada, permitiendo observar cómo los valores se alejaban de sus umbrales normales antes de la falla real.

4.6.2 Casos Fallidos: Falsos Positivos y Falsos Negativos

A pesar del buen desempeño global (F1-score entre 0.95 y 0.96 en los modelos Random Forest, XGBoost y CatBoost), también se identificaron errores de predicción, los cuales fueron analizados individualmente con el fin de entender sus causas.

Falsos Positivos (alarma de falla sin que ocurra)

Caso 3: Pozo X3 - Día 1012

Predicción: Alta probabilidad de falla.

Condiciones:

- ✓ PIP descendió de forma atípica (de 560 a 445 psi), pero se recuperó posteriormente.
- ✓ Corriente eléctrica osciló por una sobrecarga temporal.
- ✓ Realidad: No se produjo falla; se trató de una interferencia operativa por cambio de choke.

Hipótesis: El modelo interpretó una anomalía operacional como una falla inminente. Este error podría corregirse incorporando variables de superficie u operativas (como cambios en válvulas o caudales externos) que actualmente no están en el dataset.

Falsos Negativos (no detecta una falla que ocurre)

Caso 4: Pozo X4 - Día 1354

Predicción: Sin riesgo de falla.

Condiciones:

- ✓ PIP y PDES dentro de rangos aceptables, corriente y temperatura normales.
- ✓ Resultado real: Falla súbita por bloqueo mecánico causado por sólidos.

Hipótesis: El modelo no detectó este caso ya que las variables críticas (presión, temperatura, corriente) no presentaron señales previas. Se trató de una falla mecánica aguda sin manifestación anticipada en los sensores disponibles.

Este tipo de errores podría reducirse incorporando sensores de vibración o monitoreo de sólidos en línea, variables no disponibles actualmente en el dataset.

Análisis de estos casos

Estos ejemplos cualitativos resaltan tanto la fortaleza predictiva del modelo como sus limitaciones prácticas:

- Los casos de éxito reflejan la buena sensibilidad del sistema ante problemas eléctricos progresivos, lo cual es coherente con la alta importancia atribuida a las variables TI, TM, corriente y presión en los modelos de clasificación.
- Los errores tipo I (falsos positivos) suelen estar relacionados con anomalías operativas transitorias que el modelo confunde con patrones de falla.
- Los errores tipo II (falsos negativos) tienden a estar asociados a fallas mecánicas abruptas, especialmente aquellas que no alteran sensiblemente las variables monitoreadas con anticipación.

Este análisis cualitativo se llevó a cabo directamente en Jupyter Notebook, utilizando funciones de filtrado condicional, visualización con Matplotlib y Seaborn, así como inspección manual de registros críticos mediante Pandas. La combinación de este enfoque técnico con una revisión operativa permite fortalecer el modelo, enfocándose no solo en mejorar métricas globales, sino también en comprender las historias detrás de cada pozo y cada falla.

4.6.3 Evaluación sobre Series Temporales Largas

Durante el desarrollo del modelo en Jupyter Notebook, se utilizó un conjunto de datos histórico con una extensión promedio de más de 1200 días por pozo, lo cual permitió evaluar el desempeño de los algoritmos de clasificación y regresión en ventanas temporales extensas. A través del uso de técnicas de validación cruzada temporal

(TimeSeriesSplit) y la comparación del rendimiento mes a mes, se identificaron varios patrones importantes:

- En los primeros tramos temporales, cuando las condiciones del pozo eran más estables, los modelos (especialmente XGBoost y Random Forest) lograban mantener una precisión muy alta (>96%) en la predicción de fallas.
- Sin embargo, en etapas posteriores, cuando los pozos empezaban a presentar signos de envejecimiento o habían sufrido múltiples intervenciones, se observó una ligera caída en el desempeño del modelo, con una reducción del F1-score de hasta 3-4 puntos porcentuales.
- Esta degradación del desempeño no fue uniforme: en pozos donde el historial reflejaba variabilidad operacional (por ejemplo, cambios en presión de fondo o variaciones eléctricas causadas por interferencias de superficie), el modelo tendía a mostrar más falsos positivos, probablemente debido a que esas condiciones no estaban suficientemente representadas en el conjunto de entrenamiento original.

En términos prácticos, estos resultados se visualizaron en Jupyter Notebook mediante gráficos de líneas que contrastaban la predicción del modelo a lo largo del tiempo con la ocurrencia real de fallas, permitiendo identificar con claridad los periodos donde el modelo era más confiable y aquellos donde empezaba a desviarse.

CAPITULO V: CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

Este trabajo de investigación confirmó que la inteligencia artificial, en especial los algoritmos de aprendizaje automático, constituye una herramienta efectiva para optimizar decisiones operativas en la producción de pozos petroleros. A través del análisis de datos históricos registrados en bombas electrosumergibles (BES), se logró construir modelos predictivos capaces de anticipar fallas con la antelación necesaria para facilitar intervenciones preventivas.

En cuanto al objetivo general, se cumplió con la identificación y aplicación de distintas técnicas de inteligencia artificial orientadas a la predicción de fallas en sistemas de levantamiento artificial. El modelo XGBoost destacó sobre el resto por su eficiencia en el tratamiento de datos tabulares y su elevado desempeño, alcanzando un F1-score de 0.96, reflejo de su precisión y sensibilidad equilibradas en la detección de eventos críticos.

Respecto al primer objetivo específico, se desarrollaron modelos de clasificación y regresión basados en variables operativas como PIP, PDES, TM y TI, que permitieron no solo anticipar el estado operativo de la bomba, sino también estimar el tiempo restante hasta una posible falla, aportando una base sólida para la toma de decisiones en tiempo real.

En relación con el segundo objetivo específico, los modelos demostraron capacidad para automatizar decisiones clave, como la activación de alertas en condiciones de riesgo, facilitando la integración con plataformas de monitoreo y sistemas SCADA. Esto representa un paso concreto hacia una gestión más eficiente y autónoma de los activos de producción.

Finalmente, conforme al tercer objetivo específico, se implementaron modelos de detección de anomalías que comparan el comportamiento de cada pozo con su historial y con pozos similares. Este enfoque permitió identificar desviaciones relevantes que podrían anticipar fallas, aportando una nueva dimensión al control operativo.

En conjunto, esta investigación respalda la integración de soluciones basadas en inteligencia artificial en el sector hidrocarburífero, no solo como una alternativa técnica viable, sino como una herramienta estratégica para mejorar la eficiencia, reducir costos y fortalecer la continuidad operativa en pozos petroleros, tanto a nivel local como en escenarios más amplios.

5.2 RECOMENDACIONES

- **Aplicación operativa del modelo predictivo**

Una vez validado su funcionamiento, se sugiere implementar el modelo en un sistema automatizado que le permita ejecutarse de forma autónoma y periódica. Esta implementación puede facilitarse mediante una interfaz sencilla, conectada directamente a las bases de datos de producción en tiempo real, lo que permitiría al personal técnico monitorear y visualizar los resultados con claridad. Además, sería de gran utilidad incorporar un mecanismo de notificación automática —ya sea por correo electrónico o a través de dispositivos móviles— que alerte sobre posibles condiciones de falla con una antelación suficiente, por ejemplo, dentro de un horizonte de 7 a 30 días. Esto contribuiría a una gestión preventiva más eficaz y oportuna en campo.

- **Actualización constante del modelo**

Para mantener su precisión en el tiempo, el modelo debe actualizarse regularmente utilizando datos recientes de operación. Esta práctica no solo mejora su capacidad para adaptarse a cambios graduales en los pozos, sino que también extiende su aplicabilidad a otros escenarios. Incluir nuevos pozos dentro del conjunto de entrenamiento enriquecerá el aprendizaje del modelo, haciéndolo más robusto y versátil ante distintos contextos de producción.

- **Evaluación de enfoques más avanzados**

Si bien el modelo actual ofrece buenos resultados, es recomendable explorar alternativas más sofisticadas. Modelos como las redes neuronales recurrentes (RNN) o los esquemas de memoria a largo plazo (LSTM) tienen la capacidad de aprender secuencias temporales complejas, lo que podría aportar mejoras significativas en la predicción de fallas. Incluso, el uso de enfoques híbridos que combinen clasificación y regresión podría representar un avance importante en escenarios donde los patrones de comportamiento varían con el tiempo.

- **Diseño estructurado con visión escalable**

Es aconsejable que la solución desarrollada tenga una arquitectura modular y adaptable, lo que facilitaría su implementación en diferentes campos petroleros o unidades operativas sin requerir grandes modificaciones. Esta capacidad de adaptación será clave si se busca aplicar el modelo en nuevos entornos, ya que permitirá reutilizar la estructura existente sin tener que rediseñarla completamente. De este modo, se garantizará una expansión sostenible del sistema a futuro.

REFERENCIAS BIBLIOGRÁFICAS

- Advanced Factories. (s.f.). *Mantenimiento predictivo potenciado por IA: El futuro de la gestión industrial*. <https://www.advancedfactories.com>
- Analytics Vidhya. (2022, septiembre). *Get to know all about evaluation metrics*. <https://www.analyticsvidhya.com/blog/2022/09/get-to-know-all-about-evaluation-metrics/>
- Aprende Machine Learning. (2018). *Breve historia de las redes neuronales artificiales*. <https://www.aprendemachinelearning.com>
- Aprende Machine Learning. (s.f.). *Principales algoritmos usados en machine learning*. <https://www.aprendemachinelearning.com>
- Autotec Engineering. (2022). *La inteligencia artificial y su papel en la industria manufacturera*. <https://www.autotecengineering.com>
- Barragán-Martínez, X. (2023). *Situación de la Inteligencia Artificial en el Ecuador en relación con los países líderes de la región del Cono Sur*. <http://scielo.senescyt.gob.ec>
- Bravent. (2023). *Inteligencia Artificial en el sector industrial: Historia y casos de uso*. <https://www.bravent.net>
- Cultura Científica. (2025). *Una breve historia de la inteligencia artificial*. <https://culturacientifica.com>
- DataCamp. (s.f.). *Top machine learning use cases and algorithms*. <https://www.datacamp.com>
- Dataquest. (s.f.). *How to Use Jupyter Notebook: A Beginner's Tutorial*. <https://www.dataquest.io/blog/jupyter-notebook-tutorial>
- Dialnet. (s.f.). *Artículo sobre inteligencia artificial en la industria petrolera*. <https://dialnet.unirioja.es>
- Eadic. (s.f.). *Inteligencia artificial aplicada en la industria petrolera*. <https://www.eadic.com>

- El Diario. (s.f.). *Los desafíos coyunturales en la economía de Ecuador*.
<https://www.eldiario.ec>
- El Productor. (2019). *Ecuador: La producción y sus seis desafíos urgentes*.
<https://elproductor.com>
- EPN. (s.f.). *Investigación sobre inteligencia artificial en la producción petrolera*.
<https://epn.edu.ec>
- GeeksforGeeks. (2025, abril 5). *Evaluation metrics in machine learning*.
<https://www.geeksforgeeks.org/metrics-for-machine-learning-model>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Gordillo Villamar, D. A., Lasluiza Yanez, B. S., & Chávez Vaca, V. A. (2023). *Implementación de la inteligencia artificial en los procesos industriales del Ecuador*. <https://revistas.itsjapon.edu.ec>
- Gutta, S. (2021, junio 18). *Machine Learning Metrics in Simple Terms*. Medium.
<https://medium.com/analytics-vidhya/machine-learning-metrics-in-simple-terms-d58a9c85f9f6>
- Inmersys. (s.f.). *Historia de la inteligencia artificial: de la ciencia ficción a la realidad*.
<https://blog.inmersys.com>
- Introduction – Jupyter Notebook Documentation. (s.f.). *Jupyter documentation*.
<https://jupyter.readthedocs.io>
- Javaid, A. (2023, abril 10). *Evaluation Metrics in Machine Learning: How to Measure Model Performance*. LinkedIn.
- McCarthy, J. (2007). *What is artificial intelligence?*. Stanford University.
<http://jmc.stanford.edu/articles/whatisai/whatisai.pdf>
- McCulloch, W. S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. *Bulletin of Mathematical Biophysics*, 5(4), 115–133.
- Medina Macías, M. S., & Asitimbay Naula, E. M. (2024). *Análisis del uso de inteligencia artificial en las empresas exportadoras de Guayaquil*. <https://dSPACE.ups.edu.ec>
- MESbook. (2024). *Inteligencia artificial en la industria: Usos y beneficios*.
<https://mesbook.com>

- ML Journey. (2025). *What is Anaconda & Jupyter Notebook?*
<https://mljourney.com/what-is-anaconda>
- Neptune.ai. (2022, marzo 15). *Performance Metrics in Machine Learning [Complete Guide]*. <https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide>
- Ortiz, A. (2023). *El rol de la inteligencia artificial en la industria petrolera*. Innovamas Nakasawa Resources.
- Rosenblatt, F. (1958). *The perceptron: A probabilistic model for information storage and organization in the brain*. *Psychological Review*, 65(6), 386–408.
- Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4^a ed.). Pearson Education.
- Solis-Mora, V. S., & Gruezo-Valencia, D. F. (2022). *La inteligencia artificial al servicio de la eficiencia energética en el Ecuador*. *Revista Dominio de las Ciencias*, 8(2), 600–621. <https://dialnet.unirioja.es>
- TDEA. (s.f.). *Limitaciones de la metodología en la industria petrolera*. <https://tdea.edu.co>
- Turing, A. M. (1950). *Computing machinery and intelligence*. *Mind*, 59(236), 433–460.
- UNEMI. (2021). *Compendio archivológico*. <https://unemi.edu.ec>
- UPS. (s.f.). *Investigación sobre inteligencia artificial en la industria petrolera*. <https://www.ups.edu.ec>
- Wikipedia. (2025, mayo). *Anaconda (Python distribution)*. [https://en.wikipedia.org/wiki/Anaconda_\(Python_distribution\)](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution))
- Wikipedia. (2025, mayo). *F-score*. <https://en.wikipedia.org/wiki/F-score>
- Wikipedia. (2025, mayo). *Precision and recall*. https://en.wikipedia.org/wiki/Precision_and_recall

ANEXOS

ANEXO 1 Principales componentes y configuración típica de bomba BES

1. Desarmar equipo BES SF2700 de Cía. Halliburton como sigue:
 - SUB-DESCARGA ZENITH, 3-1/2" EUE BOX, SERIE 400.
 - DESCARGA BOLT-ON, 3-1/2" BRD, SERIE 400.
 - BOMBA SUPERIOR SF2700, CW 111S 15XAR1 HSS SS, 111 STG, SERIE 400,S/N:13337993.
 - BOMBA INFERIOR SF2700, CW 111S 15XAR1 HSS SS, 111 STG, SERIE 400, S/N:13337992.
 - SEP. DE FAS 400X VORTEX AR HSS SS, SERIE 400, S/N 13329258.
 - PROTECTOR UT LSBPB AR HSS SS HL, SERIE 400, S/N: 13399072.
 - PROTECTOR LT BPBSL AR HSS SS HL, SERIE 400, S/N: 13399079.
 - MOTOR FMSX1, 168 HP, 2837V, 41.9A, SERIE 456, S/N: 13394296.
 - SENSOR HAL450, OEM, DUAL PRESS, SERIE 450, S/N: TB21486.
 - CENTRALIZAER CR3, CSG 7".
2. Baja BHA de limpieza del Rig para casing de 7" con broca de 6-1/8" en 3-1/2" tubería EUE yTSH BLUE en paradas hasta 7910 Ft y saca realizando registro electromagnético de tubería en boca de pozo hasta superficie y desarma.
3. Técnico Cía. Halliburton arma equipo BES:
 - CENTRALIZADOR CSG 7", LGT: 1.00'.
 - SENSOR HAL OEM DUAL PRESSURE ESP 302 F (TRITON), SERIE: 400, S/N: 2248275 /13978934, LTG:2.35'.
 - MOTOR FMSX1, 168HP, 2837V, 41.9A, 400F SS, SERIE:456, S/N: 13649407, LGT:33.00'.
 - PROTECTOR LT, BPBSL AR HSS SS HL, SERIE: 400, S/N:13892031, LGT:8.20'.
 - PROTECTOR UT, LSBPB AR HSS SS HL, SERIE: 400, S/N:13892043, LGT:8.20'.
 - SEP. DE GAS LT 400X VORTEX, AR HSS SS, SERIE:400, S/N:13945068, LGT: 2.55'.
 - BOMBA INFERIOR SF5900 CW 67S 15 XAR1 HSS SS, SERIE: 400, S/N 14066425, LGT:22.80'.
 - BOMBA SUPERIOR SF900 CW 111S 15 XAR1

- HSS SS, SERIE: 400, S/N: 14066422, LGT:22.80'.
- DESCARGA BTN, 3-1/2, 8RD, MS1-015, SERIE: 400, LGT:0.56'.
 - SUB-DESCARGA ADPT, S/A, 3-1/2 EUE, ECNT, 1/4NPT, SS, SERIE: 400, LTG: 0.86'. LONGITUD TOTAL DE EQUIPO BES: 102.32 Ft
 - Bajan equipo BES SF2700 en 3-1/2 tubería TSH BLUE Y EUE hasta 7935.13 ft (punta de centralizador), midiendo y calibrando (*Intake queda a 7879.83 ft). Realizan empalme de cable a 4256' y 6203'. Peso sarta: subiendo 90 Klbs; bajando=90 Klbs.
 - Retiran conjunto BOP e instala 3-1/2 EUE PIN x 3-1/8 x 3000 PSI cabezal de producción.
 - Realizan prueba de giro del equipo BES con retorno al tanque de lodo del RIG. Datos estáticos: PI: 2811 PSI, PD: 2808 PSI, TI:209 °F; TM:208 °F.
 - GIRO (FWD) CORRECTO: PI:2689 PSI, PD:3491 PSI, TI:208 °F, TM:224 °F, FF:2776/2782 V, FT:1562/2583/2557 V, AMP:30, HZ: 50
 - BOMBA SUPERIOR SF900 CW 111S 15 XAR1 HSS SS, SERIE: 400, S/N: 14066422, LGT:22.80'.
 - DESCARGA BTN, 3-1/2, 8RD, MS1-015, SERIE: 400, LGT:0.56'.
 - SUB-DESCARGA ADPT, S/A, 3-1/2 EUE, ECNT, 1/4NPT, SS, SERIE: 400, LTG: 0.86'. LONGITUD TOTAL DE EQUIPO BES: 102.32 Ft
 - Bajan equipo BES SF2700 en 3-1/2 tubería TSH BLUE Y EUE hasta 7935.13 ft (punta de centralizador), midiendo y calibrando (*Intake queda a 7879.83 ft). Realizan empalme de cable a 4256' y 6203'. Peso sarta: subiendo 90 Klbs; bajando=90 Klbs.
 - Retiran conjunto BOP e instala 3-1/2 EUE PIN x 3-1/8 x 3000 PSI cabezal de producción.
 - Realizan prueba de giro del equipo BES con retorno al tanque de lodo del RIG. Datos estáticos: PI: 2811 PSI, PD: 2808 PSI, TI:209 °F; TM:208 °F.
 - GIRO (FWD) CORRECTO: PI:2689 PSI, PD:3491 PSI, TI:208 °F, TM:224 °F, FF:2776/2782 V, FT:1562/2583/2557 V, AMP:30, HZ: 50

ANEXO 2 PARTE DEL DATASET CON SUS PARAMETROS ELECTRICOS

FECHA	ARE	Pc	Cs	P	P	BFP	BOI	BWP	BSW	Ap	HZ	AMP M	IOL M	PIP	PDE	T	TM	GA	GO
03/09/2018	M1	80	0	30	22	1124	685,64	438,36	39%	19,8	56,6	44	2686	651	3038	197	338	69	100,64
07/09/2018	M1	80	0	30	22	1125	686,25	438,75	39%	19,8	56,6	44	2686	649	3036	197	340	69	100,55
09/09/2018	M1	80	0	30	22	1135	692,35	442,65	39%	19,8	57	44	2766	647	3041	197	341	69	99,66
12/09/2018	M1	85	0	30	22	1136	692,96	443,04	39%	19,8	57	44	2766	647	3039	197	341	69	99,57
17/09/2018	M1	85	0	30	22	1138	694,18	443,82	39%	19,8	57	44	2766	645	3048	197	341	69	99,40
29/09/2018	M1	80	0	30	22	1138	694,18	443,82	39%	19,8	57	44	2766	648	3047	197	333	69	99,40
08/10/2018	M1	80	0	30	22	1139	694,79	444,21	39%	19,8	57	44	2766	642	3046	197	333	70	100,75
13/10/2018	M1	85	0	30	22	1137	693,57	443,43	39%	19,8	57	44	2766	642	3045	197	333	69	99,49
15/10/2018	M1	85	0	30	22	1139	694,79	444,21	39%	19,8	57	44	2766	641	3055	197	333	70	100,75
22/10/2018	M1	80	0	28	22	1141	696,01	444,99	39%	19,8	57	44	2766	647	3046	197	332	70	100,57
28/10/2018	M1	80	0	28	22	1143	697,23	445,77	39%	19,8	57	44	2766	642	3047	197	332	70	100,40
01/11/2018	M1	80	0	28	22	1144	697,84	446,16	39%	19,8	57	44	2766	640	3045	197	333	70	100,31
06/11/2018	M1	85	0	30	22	1145	698,45	446,55	39%	19,8	57	44	2766	641	3048	197	332	69,845	100,00
13/11/2018	M1	80	0	30	22	1144	697,84	446,16	39%	19,8	57	44	2766	638	3045	197	331	69,784	100,00
17/11/2018	M1	85	0	30	22	1026	625,86	400,14	39%	19,8	57	43	2766	642	3042	197	331	62,586	100,00
20/11/2018	M1	85	0	28	22	1145	698,45	446,55	39%	19,8	57	44	2766	638	3057	197	330	70,06	100,31
28/11/2018	M1	85	0	28	22	1144	697,84	446,16	39%	19,8	57	44	2766	637	3054	197	330	70	100,31
30/11/2018	M1	90	0	29	22	1145	698,45	446,55	39%	19,8	57	44	2766	636	3053	197	330	70	100,22
05/12/2018	M1	90	0	30	22	1143	697,23	445,77	39%	19,8	57	44	2766	637	3062	197	330	70	100,40
12/12/2018	M1	100	0	30	22	1101	671,61	429,39	39%	19,8	57	44	2766	637	3078	197	330	67	99,76
14/12/2018	M1	100	0	30	22	1095	667,95	427,05	39%	19,8	57	44	2766	636	3075	197	330	67,0	100,31
16/12/2018	M1	100	40	29	22	1099	670,39	428,61	39%	19,8	57	44	2766	637	3069	197	330	67,0	99,94
19/12/2018	M1	90	30	30	22	1140	695,40	444,60	39%	19,8	57	44	2766	635	3061	197	330	69,5	100,00
22/12/2018	M1	90	28	30	22	1138	694,18	443,82	39%	19,8	57	43	2766	636	3052	197	330	69,0	99,40
27/12/2018	M1	90	15	30	22	1139	694,79	444,21	39%	19,8	57	43	2766	634	3047	197	330	69,0	99,31
04/01/2019	M1	85	10	30	22	1137	693,57	443,43	39%	19,8	57	43	2766	636	3066	197	330	69,0	99,49
10/01/2019	M1	90	10	28	22	1136	692,96	443,04	39%	19,8	57	44	2766	639	3067	197	331	69,0	99,57
13/01/2019	M1	90	10	30	22	1133	691,13	441,87	39%	19,8	57	43	2766	637	3069	197	330	69,1	100,00
18/01/2019	M1	90	10	30	22	1084	661,24	422,76	39%	19,8	57	43	2766	636	3070	197	330	66,0	99,81
20/01/2019	M1	90	10	30	22	1084	661,24	422,76	39%	19,8	57	43	2766	636	3070	197	330	66,0	99,81

Anexo3 .evaluacion con metricas en jupyter notebook

```
# 13. Evaluación con métricas
# =====

from sklearn.metrics import (classification_report, confusion_matrix,
                             roc_auc_score, precision_score, recall_score,
                             f1_score, accuracy_score,
                             mean_squared_error, mean_absolute_error, r2_score)

import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

# Clasificación
print("\n📊 Clasificación (anticipación a 7 días):")
print(classification_report(y7_test, y7_pred))

sns.heatmap(confusion_matrix(y7_test, y7_pred), annot=True, fmt='d', cmap='Blues')
plt.title("Matriz de Confusión - Falla en 7 días")
plt.xlabel("Predicción"); plt.ylabel("Real")
plt.show()

# Métricas extendidas
print("\n📈 Métricas extendidas:")
print(f"✅ Accuracy: {accuracy_score(y7_test, y7_pred):.4f}")
print(f"🎯 Precision: {precision_score(y7_test, y7_pred):.4f}")
print(f"📄 Recall: {recall_score(y7_test, y7_pred):.4f}")
print(f"👉 F1-score: {f1_score(y7_test, y7_pred):.4f}")
print(f"📊 AUC-ROC: {roc_auc_score(y7_test, y7_proba):.4f}")

# Regresión sobre probabilidades
print("\n📉 Métricas tipo regresión (sobre probabilidad de falla):")
print(f"📏 RMSE: {np.sqrt(mean_squared_error(y7_test, y7_proba):.4f}")
print(f"📏 MAE: {mean_absolute_error(y7_test, y7_proba):.4f}")
print(f"📏 R²: {r2_score(y7_test, y7_proba):.4f}")
```

ANEXO 4 Instalacion de programas para modelados

```
Collecting shap
  Downloading shap-0.47.2-cp312-cp312-win_amd64.whl.metadata (25 kB)
Requirement already satisfied: numpy in c:\users\artur\anaconda3\lib\site-packages (from shap) (1.26.4)
Requirement already satisfied: scipy in c:\users\artur\anaconda3\lib\site-packages (from shap) (1.13.1)
Requirement already satisfied: scikit-learn in c:\users\artur\anaconda3\lib\site-packages (from shap) (1.5.1)
Requirement already satisfied: pandas in c:\users\artur\anaconda3\lib\site-packages (from shap) (2.2.2)
Requirement already satisfied: tqdm>=4.27.0 in c:\users\artur\anaconda3\lib\site-packages (from shap) (4.66.5)
Requirement already satisfied: packaging>20.9 in c:\users\artur\anaconda3\lib\site-packages (from shap) (24.1)
Collecting slicer==0.0.8 (from shap)
  Downloading slicer-0.0.8-py3-none-any.whl.metadata (4.0 kB)
Requirement already satisfied: numba>=0.54 in c:\users\artur\anaconda3\lib\site-packages (from shap) (0.60.0)
Requirement already satisfied: cloudpickle in c:\users\artur\anaconda3\lib\site-packages (from shap) (3.0.0)
Requirement already satisfied: typing-extensions in c:\users\artur\anaconda3\lib\site-packages (from shap) (4.11.0)
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in c:\users\artur\anaconda3\lib\site-packages (from numba>=0.54->shap) (0.43.0)
Requirement already satisfied: colorama in c:\users\artur\anaconda3\lib\site-packages (from tqdm>=4.27.0->shap) (0.4.6)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\artur\anaconda3\lib\site-packages (from pandas->shap) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\artur\anaconda3\lib\site-packages (from pandas->shap) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\artur\anaconda3\lib\site-packages (from pandas->shap) (2023.3)
Requirement already satisfied: joblib>=1.2.0 in c:\users\artur\anaconda3\lib\site-packages (from scikit-learn->shap) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in c:\users\artur\anaconda3\lib\site-packages (from scikit-learn->shap) (3.5.0)
Requirement already satisfied: six>=1.5 in c:\users\artur\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas->shap) (1.16.0)
Downloading shap-0.47.2-cp312-cp312-win_amd64.whl (545 kB)
----- 0.0/545.2 kB ? eta -:-:--
----- 262.1/545.2 kB ? eta -:-:--
----- 545.2/545.2 kB 2.6 MB/s eta 0:00:00
Downloading slicer-0.0.8-py3-none-any.whl (15 kB)
Installing collected packages: slicer, shap
Successfully installed shap-0.47.2 slicer-0.0.8
```

ANEXO 5 INSTALACION DE MODELO GXBOOST

```
!pip install xgboost catboost
```

Collecting xgboost

Downloading xgboost-3.0.2-py3-none-win_amd64.whl.metadata (2.1 kB)

Collecting catboost

Downloading catboost-1.2.8-cp312-cp312-win_amd64.whl.metadata (1.5 kB)

Requirement already satisfied: numpy in c:\users\artur\anaconda3\lib\site-packages (from xgboost) (1.26.4)

Requirement already satisfied: scipy in c:\users\artur\anaconda3\lib\site-packages (from xgboost) (1.13.1)

Collecting graphviz (from catboost)

Downloading graphviz-0.20.3-py3-none-any.whl.metadata (12 kB)

Requirement already satisfied: matplotlib in c:\users\artur\anaconda3\lib\site-packages (from catboost) (3.9.2)

Requirement already satisfied: pandas>=0.24 in c:\users\artur\anaconda3\lib\site-packages (from catboost) (2.2.2)

Requirement already satisfied: plotly in c:\users\artur\anaconda3\lib\site-packages (from catboost) (5.24.1)

Requirement already satisfied: six in c:\users\artur\anaconda3\lib\site-packages (from catboost) (1.16.0)

Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\artur\anaconda3\lib\site-packages (from pandas>=0.24->catboost) (2024.1)

Requirement already satisfied: pytz>=2020.1 in c:\users\artur\anaconda3\lib\site-packages (from pandas>=0.24->catboost) (2024.1)

Requirement already satisfied: tzdata>=2022.7 in c:\users\artur\anaconda3\lib\site-packages (from pandas>=0.24->catboost) (2023.3)

Requirement already satisfied: contourpy>=1.0.1 in c:\users\artur\anaconda3\lib\site-packages (from matplotlib->catboost) (1.2.0)

Requirement already satisfied: cycler>=0.10 in c:\users\artur\anaconda3\lib\site-packages (from matplotlib->catboost) (0.11.0)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\artur\anaconda3\lib\site-packages (from matplotlib->catboost) (4.51.0)

Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\artur\anaconda3\lib\site-packages (from matplotlib->catboost) (1.4.4)

Requirement already satisfied: packaging>=20.0 in c:\users\artur\anaconda3\lib\site-packages (from matplotlib->catboost) (24.1)

Requirement already satisfied: pillow>=8 in c:\users\artur\anaconda3\lib\site-packages (from matplotlib->catboost) (10.4.0)

Requirement already satisfied: pyparsing>=2.3.1 in c:\users\artur\anaconda3\lib\site-packages (from matplotlib->catboost) (3.1.2)

Requirement already satisfied: tenacity>=6.2.0 in c:\users\artur\anaconda3\lib\site-packages (from plotly->catboost) (8.2.3)

Downloading xgboost-3.0.2-py3-none-win_amd64.whl (150.0 MB)

ANEXO 6 VALIDACIONES EN SERIE RED NEURONAL

Epoch 1/50

7/7 ----- 1s 41ms/step - accuracy: 0.5267 - loss: 0.7163 - val_accuracy: 0.7143 - val_loss: 0.5954

Epoch 2/50

7/7 ----- 0s 17ms/step - accuracy: 0.7276 - loss: 0.5994 - val_accuracy: 0.7857 - val_loss: 0.5119

Epoch 3/50

7/7 ----- 0s 14ms/step - accuracy: 0.8580 - loss: 0.5072 - val_accuracy: 0.8214 - val_loss: 0.4545

Epoch 4/50

7/7 ----- 0s 13ms/step - accuracy: 0.7863 - loss: 0.4956 - val_accuracy: 0.8393 - val_loss: 0.4152

Epoch 5/50

7/7 ----- 0s 14ms/step - accuracy: 0.8302 - loss: 0.4560 - val_accuracy: 0.8393 - val_loss: 0.3858

Epoch 6/50

7/7 ----- 0s 12ms/step - accuracy: 0.8346 - loss: 0.4170 - val_accuracy: 0.8750 - val_loss: 0.3641

Epoch 7/50

7/7 ----- 0s 12ms/step - accuracy: 0.9018 - loss: 0.3355 - val_accuracy: 0.8929 - val_loss: 0.3499

Epoch 8/50

7/7 ----- 0s 12ms/step - accuracy: 0.8690 - loss: 0.3466 - val_accuracy: 0.8929 - val_loss: 0.3375

Epoch 9/50

7/7 ----- 0s 14ms/step - accuracy: 0.8595 - loss: 0.3426 - val_accuracy: 0.8750 - val_loss: 0.3276

Epoch 10/50

7/7 ----- 0s 12ms/step - accuracy: 0.8647 - loss: 0.3321 - val_accuracy: 0.8750 - val_loss: 0.3195

Epoch 11/50

7/7 ----- 0s 13ms/step - accuracy: 0.8967 - loss: 0.3138 - val_accuracy: 0.8750 - val_loss: 0.3143

Epoch 12/50

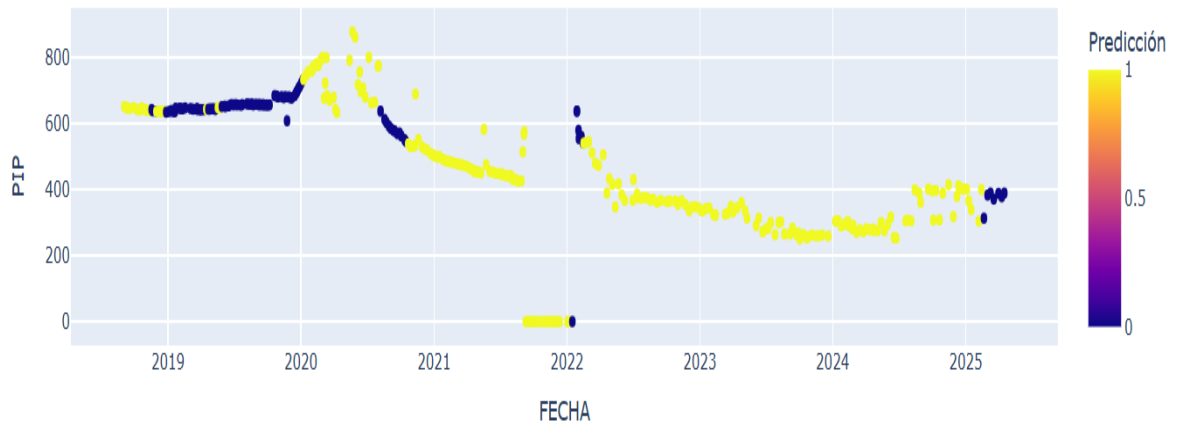
7/7 ----- 0s 14ms/step - accuracy: 0.8925 - loss: 0.3174 - val_accuracy: 0.8750 - val_loss: 0.3096

Epoch 13/50

7/7 ----- 0s 12ms/step - accuracy: 0.9232 - loss: 0.2754 - val_accuracy: 0.8929 - val_loss: 0.3058

ANEXO 7 PREDICCIÓN EN FALLA SOBRE PRESIÓN

Predicción de fallas sobre presión



ANEXO 8 MODELO CALIBRADO CATBOOST

```
from sklearn.calibration import CalibratedClassifierCV

modelo_calibrado = CalibratedClassifierCV(modelo_cat, method='sigmoid')
modelo_calibrado.fit(X7_train, y7_train)
```

```
▸ CalibratedClassifierCV ⓘ ?
▸ estimator: CatBoostClassifier
  ▸ CatBoostClassifier
```

ANEXO 9 F1-SCORE CON 5 MODELOS

```
import matplotlib.pyplot as plt

# F1-scores de los 5 modelos
f1_scores = {
    'Random Forest': 0.9505,
    'XGBoost': 0.9600,
    'CatBoost': 0.9515,
    'Red Neuronal': 0.9114,
    'SVM': 0.9048
}
```