UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

Estudio de la aplicación de técnicas de inteligencia artificial en la optimización de procesos en la industria petrolera

TESIS

Que para obtener el título de

Ingeniero Petrolero

P R E S E N T A

Rafael Ramírez de la Cruz

DIRECTOR DE TESIS

M.C. Ulises Neri Flores

Ciudad Universitaria, Cd. Mx., 2024

Agradecimientos

Un millón de gracias a quien siempre estuvo para mí, por todo lo que hizo y por todo lo que pudo ser capaz de hacer si se lo hubiera pedido, mi madre Susana de la Cruz Hernández (q.e.p.d) por ser la mejor mamá del mundo. Siempre agradeceré sus palabras de aliento, su apoyo incondicional que fue el mejor impulso para no rendirme en esos momentos dónde no encontraba una salida.
Ojalá estuvieras aquí para ver todo, me hubiera gustado que me vieras crecer más y vieras todo lo que he logrado y lo que estoy por lograr. No me queda más que agradecerte por lo que me enseñaste que va más allá de lo profesional, el amor que me inculcaste y por hacerme recordar el lado humano de la vida, no sería lo que soy si no hubieras estado conmigo. Vivirás en mi mente y en mi corazón por los años que me queden de vida.

Esto no sería posible sin el apoyo de mi padre, quien ha sido mi mayor fuente de inspiración y apoyo a lo largo de esta travesía académica. Papá, tú has estado ahí en cada paso, alentándome con tu sabiduría y paciencia. Cada logro que obtengo lleva una parte de ti. Te agradezco por ser mi guía, mi amigo y mi padre. Esta tesis es también tuya, porque sin ti, nada de esto sería posible. Gracias, papá, por ser mi faro y mi ejemplo.

A mi hermana, Mariana Ramírez de la Cruz, quiero expresarle mi agradecimiento por su constante aliento y por ser mi compañera en los momentos más difíciles. Gracias por estar siempre a mi lado y por recordarme la importancia de la familia.

Agradezco a mi familia en general por su amor y apoyo constante. Sin ustedes, este logro no habría sido posible. Con cada palabra de aliento y cada gesto de cariño, me han ayudado a llegar a este momento. Gracias por creer en mí y por estar a mi lado en cada paso del camino.

A mi director de tesis, el Maestro Ulises Neri Flores por haberme brindado su conocimiento, su experiencia y su forma de trabajo la cual me permitió conocer mi potencial y explotarlo. Es usted un gran ejemplo a seguir para mí.

Finalmente quiero expresar mi más sincero agradecimiento a la UNAM, que ha sido el lugar donde he crecido académica y personalmente. Gracias a todos los profesores y profesoras que compartieron su conocimiento y pasión por la ingeniería, inspirándome a ir más allá de lo que creía posible. Su dedicación a la enseñanza y su apoyo constante han sido fundamentales para mi desarrollo.

Índice

Resumen **5**

Abstract **5**

Introducción **6**

Justificación **7**

Objetivos **8**

Capítulo 1: Marco Teórico sobre Inteligencia Artificial **9**

Inteligencia Artificial *10*

Capítulo 2: Estado del Arte en la Industria Petrolera **28**

Actualidad de la Inteligencia Artificial en la Industria Petrolera *29*

Factores de crecimiento *33*

Capítulo 3: Resultados y Análisis de Casos de Estudio **43**

Capítulo 4: Implementación de IA en la Industria Petrolera **50**

Tecnologías de Inteligencia Artificial con potencial de aplicarse para la Industria Petrolera *51*

Evaluación de plataformas y técnicas disponibles *52*

Consideraciones en la implementación de plataformas de IA *53*

Propuesta Integral para la implementación exitosa de Inteligencia Artificial en la Industria Petrolera: un enfoque estratégico y práctico *54*

Objetivos estratégicos *54*

Enfoque estratégico y práctico *57*

Capítulo 5: Implementación de la Inteligencia Artificial en la Metodología UNFC de las Naciones Unidas **59**

UNFC: Clasificación Marco de las Naciones Unidas para los Recursos *60*

Beneficios de la integración de la IA en la Metodología UNFC *60*

Aplicaciones Prácticas de la IA en UNFC *61*

Implementación y Sostenibilidad *61*

Metodología UNFC *62*

Aplicación de la Inteligencia Artificial en la Clasificación Marco de las Naciones Unidas para los Recursos (UNFC) *68*

Ejercicio de Clasificación UNFC *122*

Conclusión **128**

Referencias **129**

Anexo 1 ..... **133**

Índice de figuras

Figura 1. Elementos básicos de la IA: Datos, Algoritmo, Entrenamiento. **12**

Figura 2. Análisis de datos desde la fuente hasta la acción. **14**

Figura 3. ¿Cómo trabaja el aprendizaje automático? **15**

Figura 4. Composición de las redes neuronales **18**

Figura 5. Aprendizaje automático vs aprendizaje profundo **21**

Figura 6. Bibliometría de 242 artículos en aprendizaje automático para geociencias por año, 1988-2017 ..... **29**

Figura 7. Rango potencial de impacto del aprendizaje automática sobre el gasto global en E&P convencional. **31**

Figura 8. Proporción de especialistas que reciben capacitación en inteligencia artificial y análisis de datos en la industria petrolera. **32**

Figura 9. Uso esperado de plataformas de IA para flujos de trabajo subterráneos de 2022 a 2024 38

Figura 10. Objetivos de desarrollo sostenible **40**

Figura 12. Impacto cuantificable de Amplified Wells en subconjunto de Oklahoma. **46**

Figura 13. Impacto cuantificable de Amplified Wells ante situación de slugging **47**

Figura 14. UNFC: Clasificación Marco de las Naciones Unidas para los Recursos. **63**

Figura 15. Diagrama de flujo para evaluar el Eje F. **71**

Figura 16. Estructura de un arbol de decisión en la IA. **77**

Figura 17. Matriz de Confusión del modelo de entrenamiento del eje F **85**

Figura 18. Matriz de Confusión del modelo de entrenamiento del eje E **106**

Figura 19. Matriz de Confusión del modelo de entrenamiento del eje G **118**

Figura 20. Resultados del código de predicción para el Eje E **124**

Figura 21. Resultados del código de predicción para el Eje E **125**

Figura 22. Resultados del código de predicción para el Eje G **126**

Figura 23. Clasificación del proyecto mostrado en el Ejercicio Ficticio **127**

Índice de tablas

Tabla 1. Definiciones de inteligencia artificial. **11**

Tabla 2. Casos reales de ejecución de IA en la Industria petrolera. **35**

Tabla 3. Clases y Subclases de UNFC. **65**

Tabla 4. Matriz de Evaluación del eje F **70**

Tabla 5. Matriz de análisis de variables para la evaluación del eje E **89**

Tabla 6. Matriz de análisis y determinación de subcategorías para la evaluación del eje E **90**

Tabla 7. Matriz de análisis para la evaluación del eje G **110**

Tabla 8. Resultados para la evaluación de categorías y clases de UNFC **120**

Tabla 9. Resultados para la evaluación de subclases de UNFC **121**

Tabla 10. Interpretación de variables del Eje E utilizadas en el ejercicio **124**

Resumen

Esta tesis explora el impacto de la inteligencia artificial (IA) en la industria petrolera, específicamente en la optimización de procesos clave. A través de la automatización de estándares como los marcos UNFC y UNRMS de la ONU, se demuestra cómo las técnicas de IA pueden mejorar la gestión de recursos y la toma de decisiones. La metodología incluye un análisis de técnicas avanzadas y su implementación práctica, revelando beneficios significativos y desafíos a superar. Los resultados destacan el potencial de la IA para transformar procesos industriales, estableciendo un modelo replicable para futuros estudios. Este trabajo contribuye a la modernización del sector, ofreciendo un marco integral para la aplicación de tecnologías avanzadas en industrias complejas.

Abstract

This thesis examines the impact of artificial intelligence (AI) on the oil industry, specifically in optimizing key processes. Through the automation of standards such as the UNFC and UNRMS frameworks from the United Nations, it demonstrates how AI techniques can enhance resource management and decision-making. The methodology includes an analysis of advanced techniques and their practical implementation, revealing significant benefits and challenges to overcome. The results highlight AI's potential to transform industrial processes, establishing a replicable model for future studies. This work contributes to the modernization of the sector, providing a comprehensive framework for the application of advanced technologies in complex industries.

Introducción

La industria petrolera ha sido un pilar fundamental de la economía global durante décadas, suministrando energía esencial para diversos sectores. Sin embargo, ha enfrentado desafíos significativos, como fluctuaciones en la demanda, avances tecnológicos y crecientes preocupaciones ambientales. La volatilidad en los precios del petróleo, las presiones para reducir las emisiones de carbono y la necesidad de mejorar la eficiencia han puesto en riesgo su estabilidad y crecimiento.

Simultáneamente, la inteligencia artificial (IA) ha emergido como una tecnología clave capaz de realizar tareas que requieren inteligencia humana, transformando la automatización de procesos complejos. Los algoritmos de IA permiten analizar grandes volúmenes de datos, identificar tendencias y hacer predicciones, facilitando decisiones más informadas y eficientes. Además, automatiza tareas repetitivas o peligrosas, mejorando la seguridad y productividad.

En este contexto, la adopción de IA se presenta como una solución prometedora para transformar la industria petrolera y abordar sus desafíos más urgentes. Este estudio tiene como objetivo analizar las técnicas de IA utilizadas en el sector y proponer un sistema para automatizar la clasificación UNFC (United Nations Framework Classification for Resources) de las Naciones Unidas. La automatización de este sistema estandarizado puede mejorar significativamente la eficiencia y precisión en la categorización de reservas y recursos.

Mediante una revisión exhaustiva de la literatura, análisis de casos prácticos y consultas con expertos, se explora el papel actual y potencial de la IA en la industria petrolera. Tecnologías como el aprendizaje automático y la IA predictiva están optimizando procesos, desde la predicción de yacimientos petroleros hasta el mantenimiento preventivo de equipos, conduciendo a operaciones más seguras y rentables.

No obstante, existen desafíos en su implementación, como la resistencia al cambio, la necesidad de inversiones significativas y preocupaciones sobre la seguridad de los datos. Esta tesis busca identificar estrategias para superar estos obstáculos y facilitar una integración efectiva de la IA en los procesos petroleros, incluyendo la automatización de la clasificación UNFC.

Adicionalmente, se abordan los beneficios ambientales potenciales de aplicar IA en el sector, como la reducción de emisiones y la gestión responsable de recursos naturales, contribuyendo a la sostenibilidad a largo plazo.

La investigación ofrece recomendaciones concretas para la implementación efectiva de la IA en la industria petrolera y destaca oportunidades para futuras investigaciones y desarrollos tecnológicos. La adopción de la IA y la automatización de la clasificación UNFC pueden ser clave para la sostenibilidad y éxito de la industria, mejorando procesos y enfrentando eficazmente los desafíos actuales.

Justificación

La industria petrolera se enfrenta a un entorno en constante evolución, con desafíos significativos relacionados en mayor medida a eficiencia operativa y sobre todo lograr los objetivos en materia de eficiencia energética y sostenibilidad. Se necesita un enfoque innovador e integrado para poder adaptarse a una nueva era, la aplicación de nuevas tecnologías y técnicas de inteligencia artificial puede proporcionar soluciones innovadoras y eficaces para abordar estas problemáticas, optimizando los procesos y mejorando los resultados.

Dada la importancia de estos aspectos y el potencial de la inteligencia artificial para transformar la industria petrolera, este estudio busca explorar y documentar las mejores prácticas, así como proporcionar recomendaciones para una implementación efectiva de estas tecnologías. Los resultados del estudio serán valiosos no solo para las empresas petroleras, sino también para los reguladores y otras partes interesadas que buscan una industria más eficiente y sostenible.

Objetivos

El objetivo principal es identificar y describir las principales técnicas de inteligencia artificial que pueden aplicarse en la industria petrolera, así como sus ventajas y limitaciones. Asimismo, se pretende examinar casos de estudio exitosos en la aplicación de inteligencia artificial en la industria petrolera, con el fin de obtener lecciones aprendidas y buenas prácticas.

Los objetivos particulares son:

- Identificar y analizar áreas clave dentro de la industria petrolera donde la inteligencia artificial pueda tener un impacto significativo, como en exploración, producción, refinamiento y mantenimiento.
- Desarrollar un modelo de inteligencia artificial que permita clasificar correctamente los proyectos según el Marco de Clasificación de las Naciones Unidas para los Recursos (UNFC).
- Evaluar el potencial de la inteligencia artificial para mejorar la sostenibilidad ambiental en la industria petrolera, incluyendo la reducción del impacto ecológico y las emisiones de carbono.
- Determinar los principales obstáculos para la adopción de inteligencia artificial en el sector petrolero.
- Proponer recomendaciones concretas para la implementación efectiva de inteligencia artificial, proporcionando un marco que facilite una transición hacia procesos más eficientes y sostenibles.

Capítulo 1: Marco Teórico sobre Inteligencia Artificial

Inteligencia Artificial

La inteligencia artificial es un campo de la informática que se enfoca en crear sistemas y tecnologías capaces de realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, la toma de decisiones y la percepción visual.

Hoy en día, la inteligencia artificial representa uno de los desafíos más significativos en el campo de la informática. Por esta razón, resulta complicado ofrecer una definición precisa y exhaustiva de este concepto, así como comprender plenamente sus posibilidades y límites.

El concepto de "Inteligencia Artificial" se originó en la conferencia de John McCarthy en 1956, celebrada en el Dartmouth College, en New Hampshire. Durante este encuentro, destacados investigadores como Marvin Minsky, Claude Shannon y N. Rochester, entre otros, exploraron la viabilidad de emular la inteligencia humana a través de máquinas. Según Patrick Henry Winston, profesor del Massachusetts Institute of Technology (MIT), la definición de inteligencia artificial tiene que ser bastante amplia, no solo es acerca del pensamiento, es acerca de percepción y acción. Entonces, se podría decir que la inteligencia artificial se trata de representaciones que respaldan la creación de modelos para facilitar la comprensión del pensamiento, la percepción y la acción.

En su esencia, la IA busca imitar la inteligencia humana utilizando algoritmos y modelos computacionales para procesar datos y tomar acciones de manera autónoma.

Stuart Russell y Peter Norvig, expertos en ciencias de la computación, identifican diversas categorías de inteligencia artificial cómo se muestran en la **Tabla 1.**

Sistemas que piensan como humanos	Sistemas que piensan racionalmente
«El nuevo y excitante esfuerzo de hacer que los computadores piensen... máquinas con mentes, en el más amplio sentido literal». (Haugeland, 1985)	«El estudio de las facultades mentales mediante el uso de modelos computacionales». (Charniak y McDermott, 1985)
«[La automatización de] actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas, aprendizaje». (Bellman,1985)	«El estudio de los cálculos que hacen posible percibir, razonar y actuar». (Winston, 1992)
Sistemas que actuan como humanos	**Sistemas que actuan racionalmente**
«El arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren de inteligencia». (Kurzweil, 1990)	«La Inteligencia Computacional es el estudio del diseño de agentes inteligentes». (Poole et al., 1998)
«El estudio de cómo lograr que los computadores realicen tareas que, por el momento, los humanos hacen mejor». (Rich y Knight, 1991)	«IA... está relacionada con conductas inteligentes en artefactos». (Nilsson, 1998)

Tabla 1. Definiciones de inteligencia artificial.

Elaboración propia a partir de (Russell, S; Norvig, P, 2021, p.2)

Tomando en cuenta las ideas de los siguientes autores podemos llegar a una definición propia de inteligencia artificial:

La Inteligencia Artificial es una disciplina científica que desarrollar sistemas informáticos que pueden ejecutar actividades típicamente asociadas con la inteligencia humana, con la meta de perfeccionarse a medida que ganan experiencia, prescindiendo de una intervención humana continua.

Aspectos Básicos de la Inteligencia Artificial

Para poder hablar de la inteligencia artificial tenemos que remontarnos sus elementos más básicos que son los datos, los algoritmos y el entrenamiento adecuado de estos. Cómo se muestra en la **Figura 1.**

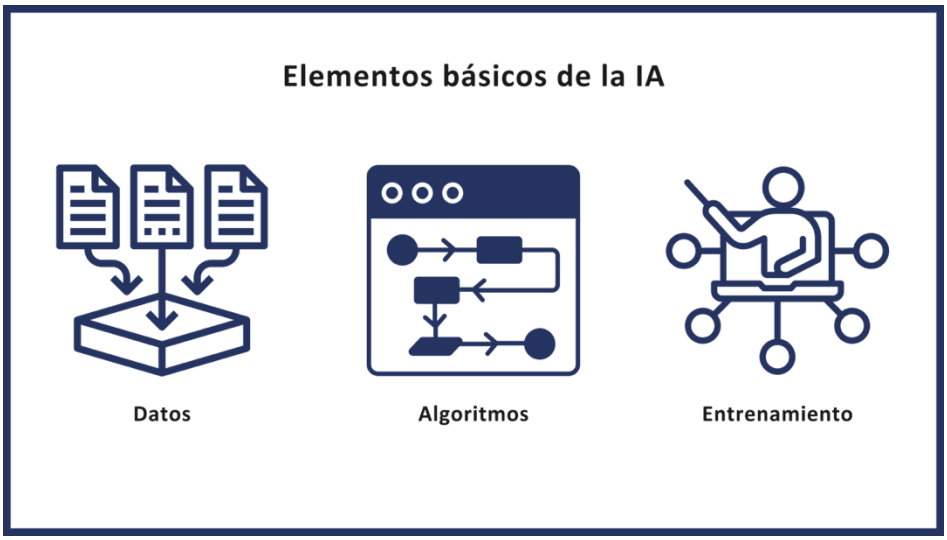

Figura 1. Elementos básicos de la IA: Datos, Algoritmo, Entrenamiento.
Elaboración propia a partir de curso de MIT: Artificial Inteligencie.

Datos

Un dato es una forma en la que la información se guarda, como números, palabras o instrucciones, que luego un sistema de computadora puede usar para entender, conectar ideas o hacer elecciones

Algoritmo

En la ciencia de datos un algoritmo es un conjunto de instrucciones que debe seguir un programa para finalizar una tarea. Comprender los algoritmos no se limita a saber que existen, sino también a saber cuándo usarlos. Esto implica comprender el problema en cuestión, lo cual es crucial en el aprendizaje de algoritmos y estructuras de datos.

Entrenamiento

El entrenamiento es el proceso mediante el cual un sistema de inteligencia artificial aprende a realizar tareas específicas a partir de datos proporcionados. Durante esta etapa, el algoritmo analiza los datos de entrada y ajusta sus parámetros para mejorar la precisión de sus predicciones o decisiones. Este proceso es iterativo y

requiere grandes cantidades de datos para que el sistema pueda identificar patrones y mejorar su desempeño con el tiempo.

Relación entre datos, algoritmos y entrenamiento

Estos tres elementos —datos, algoritmos y entrenamiento— trabajan en conjunto para permitir que la inteligencia artificial funcione de manera efectiva. Los datos alimentan el sistema, los algoritmos procesan y analizan la información, y el entrenamiento permite que la IA aprenda y se perfeccione.

En la industria petrolera, por ejemplo, la IA utiliza datos históricos de perforación, algoritmos de aprendizaje automático y procesos de entrenamiento para predecir la ubicación de yacimientos petrolíferos o para identificar patrones de mantenimiento en los equipos, lo que resulta en operaciones más eficientes y rentables.
Esta combinación de elementos es lo que permite que la inteligencia artificial tome decisiones inteligentes y se adapte a diferentes entornos y desafíos.

Análisis de datos

El análisis de datos consiste en manejar datos para extraer información valiosa que ayude a tomar decisiones informadas. La toma de decisiones basada en datos implica el uso de hechos, datos y métricas para tomar decisiones empresariales, en lugar de confiar en la intuición o las emociones.
Aunque esto parece obvio, muchas organizaciones no aprovechan los datos tanto como podrían. Según el McKinsey Global Institute, las empresas que utilizan datos son mejores para captar y retener clientes, y logran una rentabilidad superior a la media.
En este contexto, el aprendizaje automático (machine learning) juega un papel crucial al proporcionar algoritmos y modelos que pueden automatizar el análisis de datos, permitiendo la identificación de patrones y tendencias en grandes conjuntos de datos. Esto mejora significativamente la precisión y eficiencia del análisis, como se muestra en la **Figura 2**.

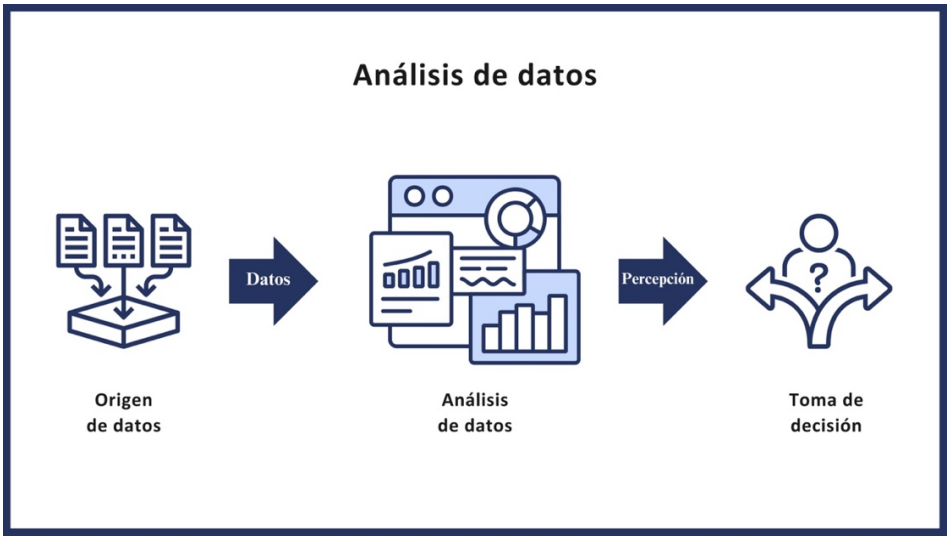

Figura 2. Análisis de datos desde la fuente hasta la acción.
Elaboración propia a partir de curso de The MITPress.

En la industria petrolera, el análisis de datos es fundamental para optimizar operaciones. Los sistemas de IA pueden procesar datos provenientes de sensores en tiempo real, registros históricos de perforación, y análisis geológicos, para predecir la producción de un pozo o detectar fallas potenciales antes de que ocurran. Esto permite a las empresas petroleras tomar decisiones informadas, reducir costos y maximizar la eficiencia operativa.

Aprendizaje automático

El aprendizaje automático (Machine Learning) es un campo de la inteligencia artificial que se enfoca en el diseño y análisis de herramientas informáticas que utilizan experiencias previas para tomar decisiones futuras. Es el estudio de programas que pueden aprender de los datos y mejorar su desempeño con el tiempo, sin necesidad de intervención humana continua.

Según el MIT, el aprendizaje automático proporciona a las máquinas algoritmos que les permiten revisar y analizar grandes cantidades de datos, para luego aprender de ellos y predecir comportamientos futuros. Estos sistemas se perfeccionan de manera autónoma, lo que significa que cuantos más datos procesan, mejor se

vuelven en sus predicciones y decisiones, de ahí el nombre "aprendizaje automático".

Un sistema de Machine Learning no se limita únicamente a aplicar un algoritmo de aprendizaje o una librería; es un proceso completo que generalmente incluye una serie de pasos:

1. Recolectar datos accediendo a una base de datos, usando datos de dominio público o utilizando dispositivos.
2. Procesar los datos asegurando que estos tienen el formato correcto para aportar en el desarrollo del algoritmo.
3. Pre analizar los datos para corregir errores, valores faltantes o encontrar patrones de afluencia en una predicción.
4. Entrenar el algoritmo para que sea capaz de hacer predicciones.
5. Utilizar el modelo y medir su rendimiento.

Para comprender cómo trabaja el aprendizaje automático, la Universidad de Berkeley (2020) dividió el proceso de aprendizaje de un algoritmo de Machine Learning en tres partes principales, como se muestra en la **Figura 3.**

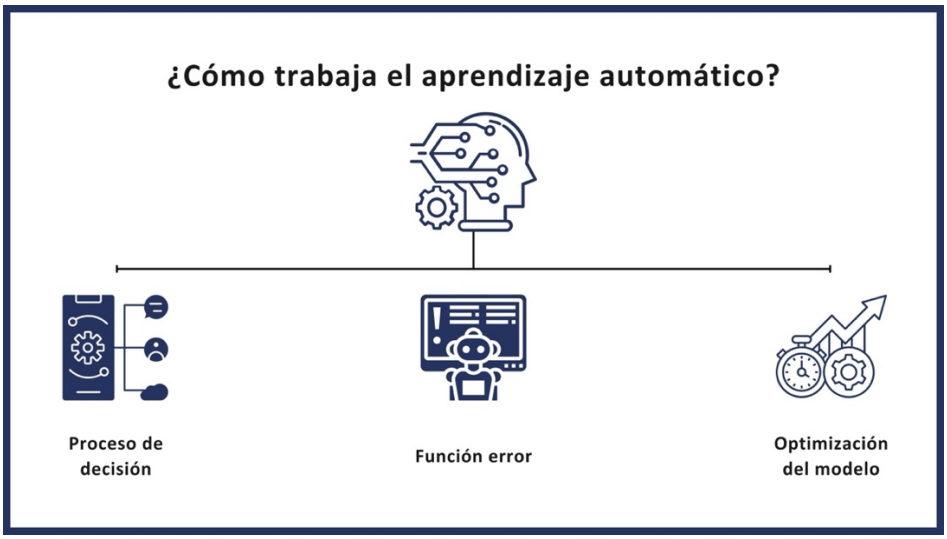

Figura 3. ¿Cómo trabaja el aprendizaje automático?
Elaboración propia a partir de datos de la Universidad de Berkley

Proceso de decisión: Los algoritmos de aprendizaje automático se utilizan para realizar predicciones o clasificaciones basadas en datos de entrada. Estos datos pueden estar etiquetados (con una respuesta conocida) o sin etiquetar, y el algoritmo genera una estimación sobre un patrón en los datos.

Función de error: Esta función evalúa la predicción del modelo comparándola con el resultado real. Si existen ejemplos conocidos, la función de error mide la precisión del modelo y determina cuánto se desvió la predicción del resultado correcto.

Proceso de optimización del modelo: Este proceso busca ajustar los parámetros del modelo para mejorar su precisión, reduciendo la discrepancia entre las predicciones y los resultados reales. El algoritmo repite este ciclo de "evaluar y optimizar" de forma iterativa, actualizando sus pesos hasta que alcanza un umbral de precisión aceptable.

Categorización de modelos de aprendizaje automático

Actualmente, los modelos de aprendizaje automático se dividen en tres categorías principales:

- **Aprendizaje supervisado:** Utiliza conjuntos de datos etiquetados para entrenar al algoritmo, permitiéndole aprender la relación entre la entrada y la salida esperada. Se aplica en técnicas como las redes neuronales, la regresión logística y las máquinas de vectores de soporte. Un ejemplo común es la clasificación de correos electrónicos en "spam" o "no spam".
- **Aprendizaje no supervisado:** Se enfoca en analizar y agrupar datos no etiquetados para descubrir patrones o estructuras subyacentes sin intervención humana. Esto es útil en escenarios como la segmentación de clientes o la detección de patrones en grandes volúmenes de datos. Los métodos incluyen el análisis de componentes principales y técnicas de agrupamiento probabilístico.

- **Aprendizaje semisupervisado:** Combina datos etiquetados y no etiquetados, proporcionando un enfoque intermedio que resulta útil cuando los datos etiquetados son limitados o costosos de obtener. Este enfoque es especialmente relevante en situaciones como el reconocimiento de imágenes, donde solo una pequeña cantidad de datos está etiquetada.

Cada una de estas modalidades de aprendizaje automático ofrece herramientas poderosas para abordar una amplia gama de problemas en el ámbito de la inteligencia artificial y el análisis de datos, permitiendo a las máquinas tomar decisiones precisas y eficientes en diversos contextos.

Redes neuronales artificiales

Las redes neuronales artificiales son modelos de aprendizaje automático inspirados en el funcionamiento del cerebro humano. Al igual que las neuronas biológicas, estas redes están compuestas por nodos (llamados neuronas) que trabajan en conjunto para identificar patrones, sopesar opciones y llegar a conclusiones. Su estructura y funcionamiento permiten a las redes neuronales realizar tareas como la clasificación de datos, la predicción de resultados o el reconocimiento de patrones. (Blais & Mertz, 2018).

Una red neuronal se organiza en capas. Existen tres tipos principales de capas, como se muestra en la **Figura 4:**

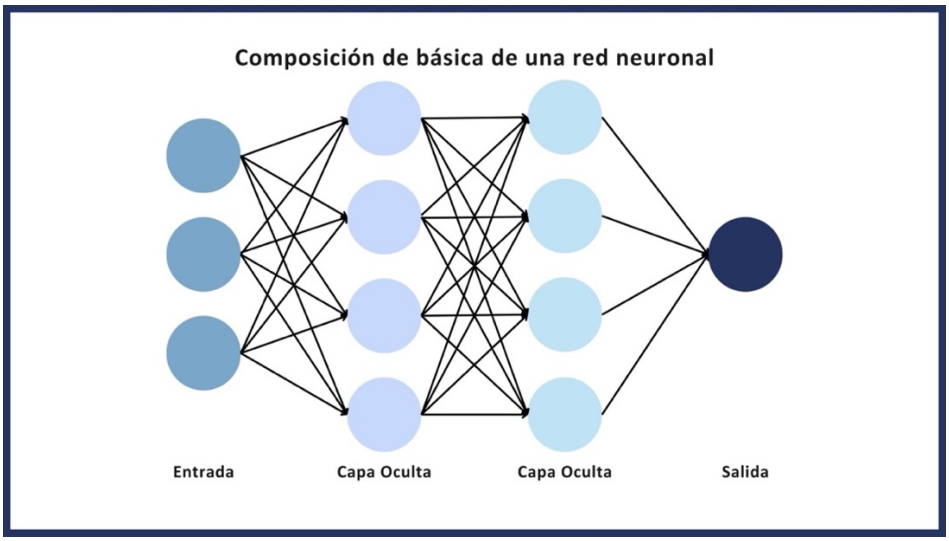

Figura 4. Composición de las redes neuronales
Elaboración propia a partir de datos de IBM

- **Capa de entrada:** Recibe los datos brutos que se van a procesar, como imágenes, texto o números.
- **Capas ocultas:** Estas son capas intermedias que procesan la información mediante operaciones matemáticas complejas. Puede haber una o más capas ocultas en una red neuronal.
- **Capa de salida:** Produce el resultado final, que puede ser una clasificación, un número, una predicción o cualquier otro tipo de salida requerida según la tarea.

Cada nodo o neurona artificial en una red neuronal tiene conexiones con otros nodos, y estas conexiones están ponderadas. Si la salida de un nodo supera un umbral predefinido, el nodo se activa y envía la información a la siguiente capa. De lo contrario, no se transmite información.

Las redes neuronales aprenden a partir de datos de entrenamiento, ajustando sus pesos internos para mejorar su precisión con el tiempo. Este proceso de ajuste es lo que les permite "aprender". Una vez optimizadas, las redes neuronales son herramientas poderosas en el campo de la informática y la inteligencia artificial, ya que permiten clasificar y agrupar datos de manera muy eficiente. Tareas como el reconocimiento de voz o imágenes, que antes requerían horas de procesamiento manual, ahora pueden realizarse en minutos mediante redes neuronales.

Tipos de redes neuronales

Existen distintos tipos de redes neuronales, y cada una está diseñada para tareas específicas. Según IBM (2018), las redes neuronales se clasifican en:

- **Redes de retroalimentación (perceptrones multicapa):** Estas redes tienen una capa de entrada, una o más capas ocultas y una capa de salida. Son fundamentales en aplicaciones como la visión artificial, el procesamiento del lenguaje natural y otras áreas de la inteligencia artificial. Su característica principal es que el flujo de datos es unidireccional, de la entrada a la salida.
- **Redes convolucionales (CNN):** Estas redes están diseñadas principalmente para el reconocimiento de imágenes, la visión artificial y el reconocimiento de patrones. Utilizan operaciones de convolución (matemáticas relacionadas con la multiplicación de matrices) para detectar características visuales en una imagen, como bordes, texturas y formas
- **Redes recurrentes (RNN):** Se distinguen por sus bucles de retroalimentación, lo que les permite recordar información de las entradas anteriores. Son útiles para analizar secuencias temporales de datos y hacer predicciones sobre resultados futuros. Las RNN son ampliamente utilizadas en tareas como la predicción de series temporales, el análisis del mercado de valores o los pronósticos de ventas.

Funcionamiento de una red neuronal

El aprendizaje en una red neuronal se produce a través de un proceso llamado retro propagación. La red neuronal ajusta sus conexiones internas basadas en los datos de entrenamiento para realizar tareas específicas, como la clasificación, predicción o reconocimiento de patrones.

Entrada de datos: La capa de entrada recibe la información inicial (imágenes, texto, números), que se divide en pequeñas unidades llamadas neuronas de entrada. Cada neurona de entrada procesa una pequeña parte de los datos.

Capas ocultas: La información pasa por una o más capas ocultas. Cada neurona en estas capas realiza una operación matemática sobre los datos que recibe, sumando las entradas ponderadas y aplicando una función de activación. El resultado de esta operación se pasa a la siguiente capa.
Capa de salida: Finalmente, los resultados procesados llegan a la capa de salida, donde se genera el resultado final, como una clasificación, un valor numérico o una predicción.

Durante el entrenamiento, la red compara la salida generada con la salida esperada o verdadera. Utiliza una función de pérdida para medir el error entre ambas. Luego, ajusta los pesos de las conexiones entre las neuronas para reducir este error mediante el proceso de retro propagación, que actualiza los pesos en función del error detectado. Este proceso de evaluación y ajuste se repite de manera iterativa hasta que la red alcanza un nivel de precisión aceptable.

Aprendizaje profundo

El aprendizaje profundo es un subconjunto del aprendizaje automático que es esencialmente una red neuronal con tres o más capas. Estas redes neuronales intentan simular el comportamiento del cerebro humano, aunque lejos de igualar su capacidad, lo que les permite "aprender" de grandes cantidades de datos. Si bien una red neuronal de una sola capa puede aun así realizar predicciones aproximadas, las capas ocultas adicionales ayudan a optimizar y refinar la precisión. (*¿Qué Es el Deep Learning?* | *IBM*, s. f.)
El deep learning apoya en muchas aplicaciones y servicios de la inteligencia artificial que mejoran la automatización al encargarse de tareas analíticas y físicas sin necesidad de intervención humana.
El beneficio que puedes tener con el aprendizaje profundo recae en su automatización, fiabilidad, escalabilidad y flexibilidad.

Los algoritmos de deep learning tienen aplicaciones en:

- Visión Artificial
- Procesamiento de lenguaje natural
- Mantenimiento predictivo
- Robótica

Aprendizaje profundo vs aprendizaje automático

La evolución de la inteligencia artificial ha dado lugar a dos conceptos clave: el aprendizaje profundo (deep learning) y el aprendizaje automático (machine learning). Ambos son subcampos de la IA, pero tienen diferencias significativas en sus métodos y aplicaciones, como podemos ver en la **Figura 5**.

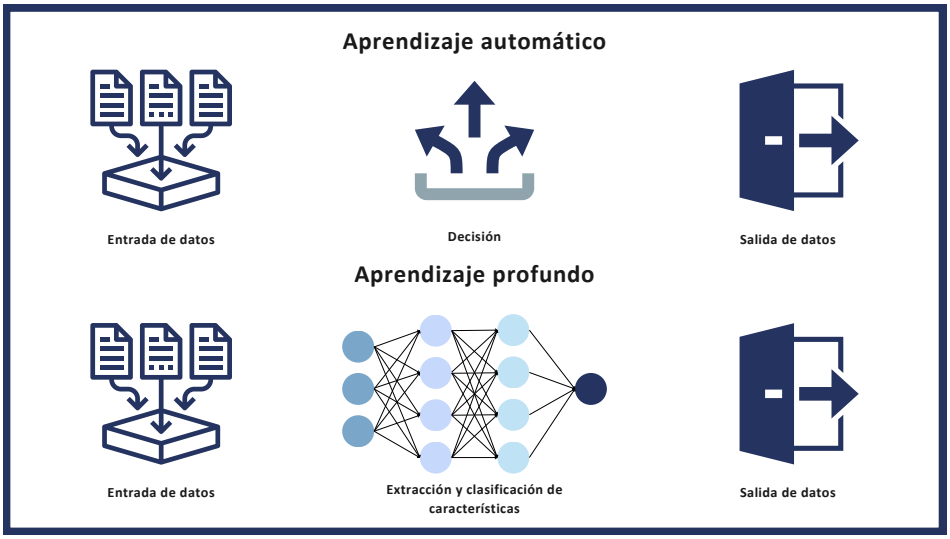

Figura 5. Aprendizaje automático vs aprendizaje profundo
Elaboración propia a partir de datos de IBM y MIT

El aprendizaje automático se basa en la creación de algoritmos que permiten a las máquinas identificar patrones y tomar decisiones basadas en datos. Estos algoritmos requieren que las características más relevantes de los datos sean seleccionadas y extraídas manualmente, lo que implica un conocimiento profundo del dominio del problema. Técnicas como los árboles de decisión, las máquinas de soporte vectorial (SVM) y los modelos de regresión son ampliamente utilizados en machine learning para resolver problemas que involucran datos estructurados y

semiestructurados, tales como la predicción de resultados, la clasificación de información y la detección de anomalías.

En contraste, el aprendizaje profundo es una evolución del aprendizaje automático que emplea redes neuronales artificiales con múltiples capas (de ahí el término “profundo”). Este enfoque es capaz de aprender automáticamente representaciones jerárquicas de los datos, eliminando la necesidad de extracción manual de características. En lugar de depender de la intervención humana para identificar qué características son importantes, el deep learning construye estas representaciones directamente a partir de los datos crudos, lo que le permite sobresalir en problemas donde los datos son no estructurados o complejos, como imágenes, voz o texto. Las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN) son algunos ejemplos de las arquitecturas más utilizadas en deep learning, que han sido cruciales para el avance de campos como la visión por computadora y el procesamiento del lenguaje natural.

Mientras que el machine learning tradicional puede ser eficaz para problemas donde las características relevantes de los datos están bien comprendidas y pueden ser extraídas manualmente, el deep learning sobresale en problemas donde la extracción de características es compleja y no lineal.

Las aplicaciones del machine learning incluyen desde análisis predictivo en finanzas y sistemas de recomendación en comercio electrónico, hasta la detección de fraudes y el diagnóstico médico. Por otro lado, el deep learning ha revolucionado industrias como la automoción, donde es fundamental para el desarrollo de vehículos autónomos, y la tecnología de consumo, donde se utiliza en asistentes virtuales, reconocimiento facial y traducción automática.

En términos de rendimiento, el aprendizaje profundo ofrece una mayor precisión y capacidad de generalización en problemas donde los datos son masivos y no estructurados, pero a costa de una mayor demanda computacional y de datos. El aprendizaje automático tradicional, aunque más limitado en cuanto a capacidad para procesar grandes volúmenes de datos no estructurados, sigue siendo

altamente eficiente y menos costoso en términos de recursos para tareas bien definidas con datos estructurados.

Mientras que el aprendizaje automático es más adecuado para problemas donde las características de los datos son conocidas y relativamente simples de extraer, el aprendizaje profundo es la mejor opción para problemas complejos que requieren la interpretación de datos no estructurados o altamente dimensionales.

Visión artificial

La visión artificial o visión por computadora es una parte de la IA que ayuda a las computadoras y sistemas a entender y sacar información valiosa de fotos, videos y otras imágenes, permitiéndoles tomar decisiones o dar recomendaciones basadas en lo que ven.

La IA es el cerebro, mientras que la visión artificial les permite observar y comprender entrenando a las máquinas para poder realizar estas tareas basados en datos y algoritmos.

Para hacer funcionar una visión artificial se necesita del aprendizaje automático y una red neuronal convolucional. El primero usa modelos y algoritmos que permiten a la computadora aprender por sí sola sobre los datos visuales y la red neuronal realiza convoluciones y verifica la exactitud de sus predicciones repetidamente hasta que estas se vuelven precisas. Luego, comenzará a reconocer o ver imágenes de manera similar a los humanos.

Procesamiento del lenguaje natural

El procesamiento del lenguaje natural, o PLN, combina la lingüística computacional, que se basa en el modelado del lenguaje humano mediante reglas, con modelos estadísticos y de aprendizaje automático. Esta combinación permite a las computadoras y dispositivos digitales reconocer, comprender y generar tanto texto como voz humana de manera eficiente (Kavlakoglu, 2020).

El PLN es fundamental para analizar grandes volúmenes de datos no estructurados, como textos, lo que facilita la obtención de información valiosa y una comprensión

más profunda de los datos en cuestión. Esto lo convierte en una herramienta clave para transformar el texto en bruto en conocimiento accionable, especialmente en campos como el análisis de sentimientos, la minería de opiniones y la automatización de tareas repetitivas relacionadas con el lenguaje.
Sin embargo, la complejidad inherente del lenguaje humano plantea desafíos significativos en el desarrollo de software capaz de interpretar con precisión el significado de datos textuales o de voz. Elementos como los homónimos, homófonos, sarcasmo, modismos y las excepciones gramaticales complican la tarea de enseñar a las aplicaciones de PLN a comprender correctamente el lenguaje humano. Los desarrolladores deben diseñar modelos robustos que puedan lidiar con estas peculiaridades del lenguaje para garantizar la efectividad de las soluciones basadas en PLN.

El PLN se encuentra en el núcleo de una amplia gama de aplicaciones y dispositivos, habilitando capacidades como:

- Resumir grandes volúmenes de texto.
- Traducir textos de un idioma a otro.
- Responder a comandos mediante asistentes virtuales.
- Autenticar usuarios a través del reconocimiento de voz.
- Generar automáticamente texto, gráficos u otro tipo de contenido bajo demanda.

Gracias a su capacidad para procesar y entender el lenguaje humano, el PLN sigue impulsando la evolución de tecnologías como los chatbots, los sistemas de recomendación y los asistentes inteligentes, facilitando la interacción entre humanos y máquinas de manera más natural y fluida.

Mantenimiento predictivo

El mantenimiento preventivo se refiere a la realización de actividades de mantenimiento programadas regularmente para prevenir posibles fallos en el futuro. En términos simples, se trata de arreglar las cosas antes de que se rompan (Mora, 2020). Sin embargo, con la incorporación de tecnologías avanzadas como el machine learning y el análisis de datos, el concepto ha evolucionado hacia el mantenimiento predictivo.

Mediante el uso de algoritmos de machine learning y el análisis de datos operativos en tiempo real, los ingenieros pueden predecir cuándo es necesario realizar mantenimiento en equipos y maquinaria, evitando así problemas antes de que ocurran. Este enfoque predictivo no solo permite optimizar la vida útil de los equipos, sino que también ayuda a reducir tiempos de inactividad no planificados, a mantener la continuidad operativa y a asegurar el cumplimiento de las garantías y normas de seguridad. En definitiva, el mantenimiento predictivo es clave para resolver posibles fallos antes de que estos afecten la producción.

Existen cuatro tipos principales de mantenimiento que se alinean con los principios del mantenimiento predictivo:

- **Mantenimiento basado en el uso:** Se lleva a cabo teniendo en cuenta las condiciones de operación y el uso promedio diario de los equipos. Es ideal para entornos donde la maquinaria tiene ciclos de uso intensivos o variables.
- **Mantenimiento basado en el tiempo:** Programado en función de un calendario fijo o cuando se aproxima la fecha de vencimiento de los componentes, este tipo de mantenimiento se basa en un criterio temporal más que en las condiciones reales del equipo.
- **Mantenimiento predictivo:** Utiliza datos en tiempo real y modelos predictivos para programar acciones de mantenimiento correctivo antes de que ocurra una falla. La predicción se basa en patrones de desgaste o anomalías detectadas en los datos operativos.

- **Mantenimiento prescriptivo:** Va un paso más allá del predictivo, ya que no solo indica cuándo se producirá una falla, sino también por qué ocurre y cuál es la mejor manera de prevenirla o corregirla. Ofrece recomendaciones precisas para la toma de decisiones, lo que permite a las empresas actuar con mayor rapidez y efectividad.

El mantenimiento predictivo y prescriptivo están transformando la forma en que las industrias gestionan el ciclo de vida de sus activos, permitiendo un enfoque más proactivo que minimiza el riesgo de interrupciones inesperadas y maximiza la eficiencia operativa.

Robótica

La manipulación robótica se refiere al movimiento de piezas y herramientas en el espacio mediante algún tipo de mecanismo, lo que implica la necesidad de representar con precisión las posiciones y orientaciones tanto de los objetos como del mecanismo en sí (J. Craig, 2006). La robótica es un campo multidisciplinario que combina la ingeniería, la ciencia y la tecnología para diseñar, crear y operar robots, dispositivos programables capaces de realizar tareas que típicamente requieren intervención humana.

Este campo abarca múltiples disciplinas, como la mecánica, la electrónica, la informática y la inteligencia artificial (IA). Cada una de estas áreas contribuye al desarrollo de robots más sofisticados y autónomos, capaces de realizar tareas complejas con un alto grado de precisión y eficiencia. Los robots han demostrado ser fundamentales en entornos donde las tareas son repetitivas, peligrosas o donde se requiere una alta precisión, como en líneas de ensamblaje, operaciones quirúrgicas o exploración espacial.

El futuro de la robótica está siendo impulsado por los avances en inteligencia artificial y aprendizaje automático. Estos avances permiten a los robots no solo ejecutar tareas preprogramadas, sino también adaptarse a nuevas situaciones y aprender de su entorno. Los robots dotados de IA pueden interactuar de manera

más natural con los humanos, resolver problemas no predefinidos y operar en entornos dinámicos y complejos. Esta capacidad de adaptación es clave para aplicaciones en sectores como la logística, la atención médica y la manufactura avanzada.

Se prevé que la robótica continuará transformando áreas esenciales de la vida cotidiana y la industria, donde los robots y los humanos trabajarán cada vez más en colaboración. Esta interacción entre humanos y robots, conocida como robótica colaborativa o "cobots", está mejorando la seguridad y eficiencia en el lugar de trabajo al compartir responsabilidades y combinar las fortalezas humanas con la precisión y resistencia de las máquinas.

La robótica no solo está cambiando la manera en que operan las industrias, sino que también está impactando nuestras vidas diarias, desde robots que asisten en el hogar hasta vehículos autónomos. La continua evolución de esta tecnología promete revolucionar aún más el trabajo, la medicina y la vida cotidiana, mejorando tanto la calidad como la seguridad en diversas actividades.

Capítulo 2: Estado del Arte en la Industria Petrolera

Actualidad de la Inteligencia Artificial en la Industria Petrolera

La inteligencia artificial, aunque relativamente nueva en la industria petrolera, ha estado ganando terreno rápidamente en los últimos años. Si bien no es completamente desconocida, su adopción y aplicación se están intensificando significativamente. Actualmente, un número creciente de empresas en el sector están empleando y mejorando técnicas de IA para que esta tecnología represente una transformación fundamental en sus operaciones.

Podemos notar que desde el año 2010, la publicación de trabajos y artículos sobre el uso del machine learning en geociencias ha experimentado un crecimiento exponencial, como podemos verlo en la **Figura 6.** Este aumento se ha visto impulsado por varios factores clave que han confluido para transformar esta área de investigación.

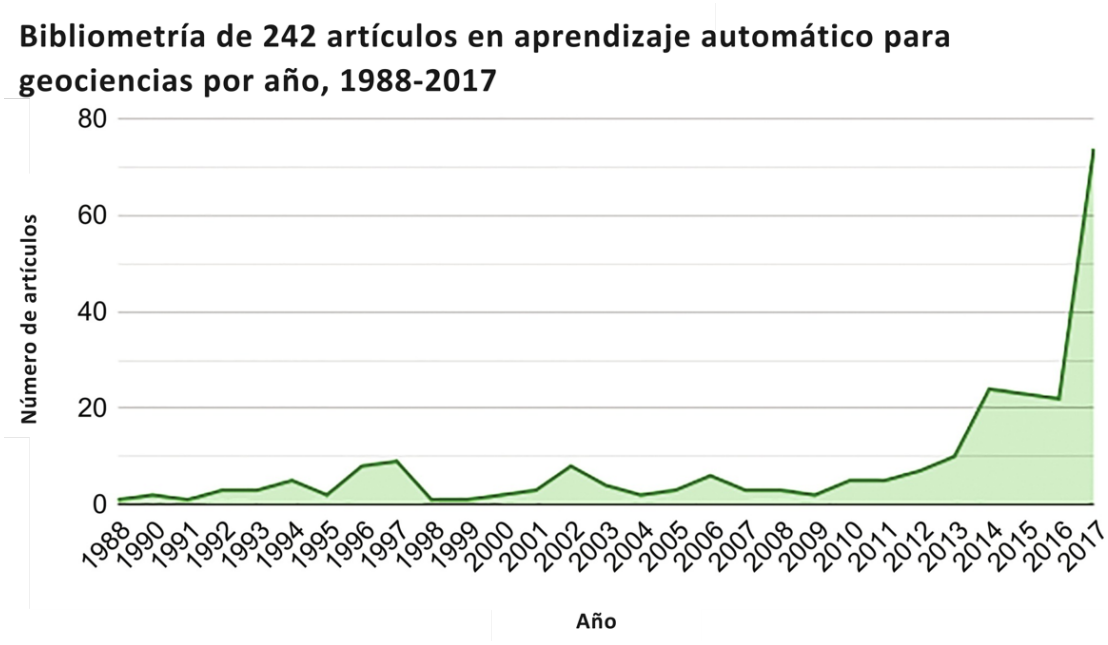

Figura 6. Bibliometría de 242 artículos en aprendizaje automático para geociencias por año, 1988-2017. Elaboración propia con datos de (Sören Dramsch, 2020,p7)

El desarrollo de nuevas tecnologías y algoritmos de aprendizaje automático ha mejorado considerablemente la capacidad de procesar y analizar grandes volúmenes de datos geocientíficos. Herramientas como las redes neuronales profundas, los bosques aleatorios y los métodos de aprendizaje por refuerzo han demostrado ser especialmente eficaces en la interpretación de datos complejos y en la predicción de fenómenos geológicos. Además, la creciente disponibilidad de grandes conjuntos de datos, obtenidos mediante mejoras en la recopilación de información geoespacial, satelital y sensores en tiempo real, ha proporcionado una base sólida para el entrenamiento de modelos de machine learning, lo que ha permitido abordar problemas geocientíficos con un nivel de precisión sin precedentes.

El interés y la inversión en investigación interdisciplinaria han sido también factores cruciales en este crecimiento. Las colaboraciones entre expertos en geociencias, informática y estadística han generado nuevas metodologías y aplicaciones prácticas, que han sido ampliamente reconocidas en la literatura académica. Este enfoque multidisciplinario ha abierto nuevas oportunidades para la innovación y la resolución de problemas complejos en el ámbito de las geociencias. En este contexto, algunas de las principales compañías petroleras, como ExxonMobil, Total y Royal Dutch Shell, han estado a la vanguardia en la adopción de IA. Estas empresas han invertido considerablemente en la investigación y el desarrollo de tecnologías basadas en IA, así como en la capacitación de su personal para aprovechar al máximo estas herramientas. Gracias a ello, han mejorado su competitividad en un mercado cada vez más globalizado y exigente.

Según el informe de desempeño de 2024 de la consultora Kimberlite International Oilfield Research, el machine learning y la IA lideran la lista de tecnologías emergentes con mayor potencial para aumentar la productividad, reducir riesgos y generar ahorros de costos. Alrededor del 79% de los operadores a nivel mundial han identificado las soluciones de machine learning como una prioridad estratégica, como se detalla en la **Figura 7.**

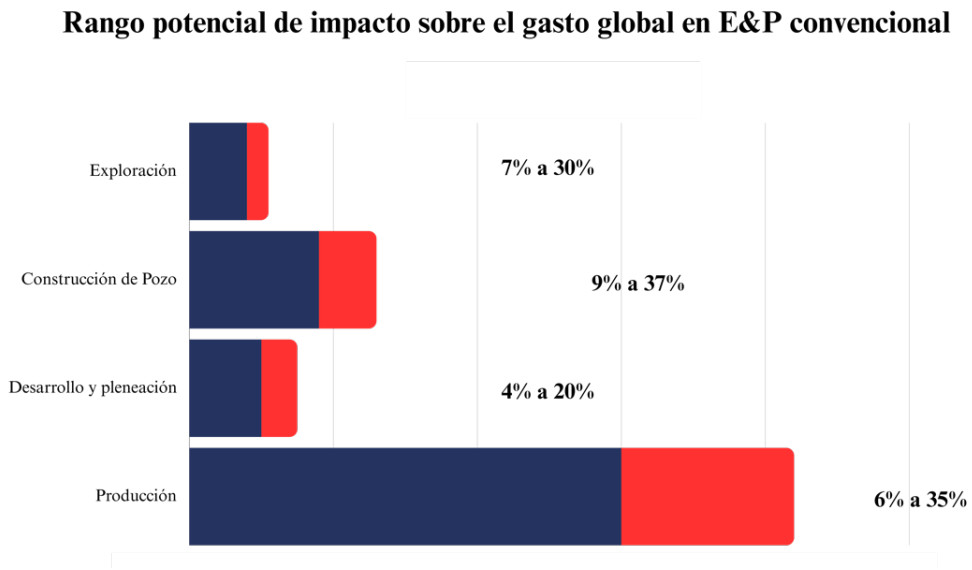

Figura 7. Rango potencial de impacto del aprendizaje automática sobre el gasto global en E&P convencional. Elaboración propia con datos de Kimberlite,LLC-2024 G&G Software Supplier Performance Report.

A pesar de que las grandes compañías petroleras han liderado la adopción de IA, las empresas más pequeñas también están comenzando a beneficiarse de esta tecnología. A medida que los costos de implementación disminuyen y las barreras de entrada se reducen, las compañías más pequeñas pueden aplicar IA para mejorar sus propias operaciones y mantenerse competitivas en un mercado en constante evolución.

La IA está transformando la industria petrolera de manera profunda, desde la optimización de procesos hasta la predicción de fallas en equipos. Está desempeñando un papel crucial en la mejora de la eficiencia operativa y en la toma de decisiones estratégicas. Sin embargo, a pesar de los avances tecnológicos, aún existe un desafío importante: la falta de especialistas capacitados en IA y análisis de datos en la industria petrolera, como se evidencia en la **Figura 8.**

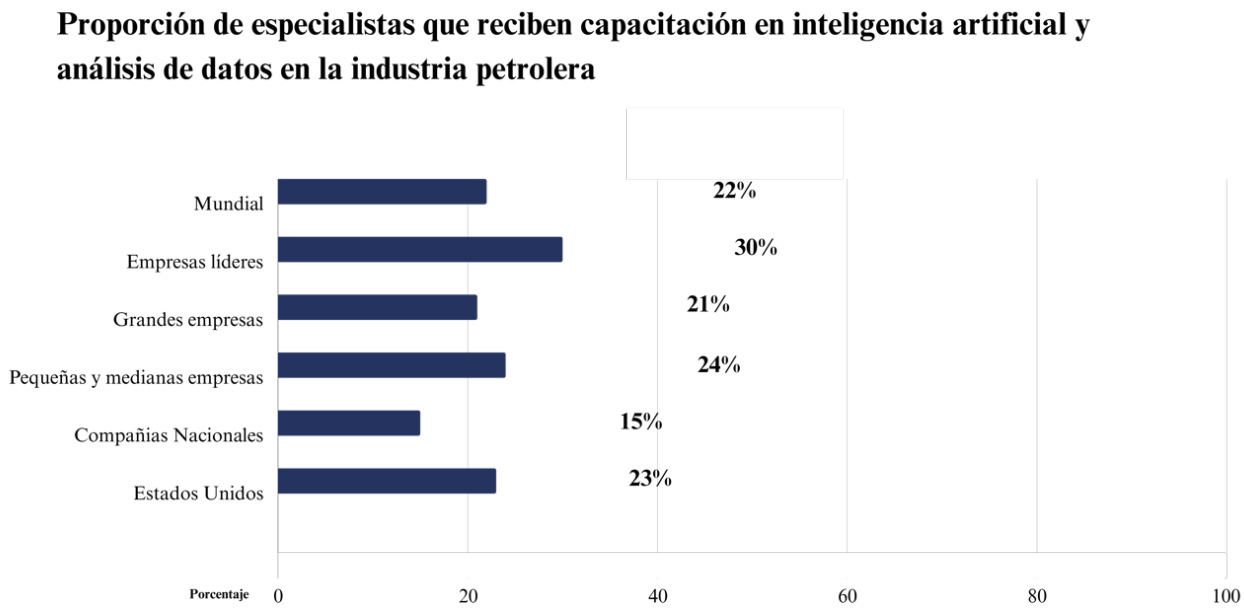

Figura 8. Proporción de especialistas que reciben capacitación en inteligencia artificial y análisis de datos en la industria petrolera.
Elaboración propia con datos de Kimberlite,LLC-2024 G&G Software Supplier Performance Report.

Para lograr una adopción plena de estas tecnologías innovadoras, es esencial contar con una fuerza laboral preparada y actualizada. Es imperativo aumentar la capacitación en IA y análisis de datos entre los profesionales del sector petrolero, no solo para garantizar una implementación efectiva de las soluciones de IA existentes, sino también para fomentar el desarrollo de nuevas herramientas y algoritmos específicos para los desafíos únicos de esta industria.

La inversión en programas de formación y desarrollo profesional en IA y análisis de datos debe ser una prioridad tanto para las empresas petroleras como para las instituciones educativas. Al fomentar una cultura de aprendizaje continuo y proporcionar los recursos adecuados, se podrá maximizar el potencial transformador de la IA en la industria petrolera y garantizar su crecimiento sostenible en el futuro.

Factores de crecimiento

El desarrollo de tecnologías de IA ha permitido avances significativos en la capacidad de procesar y analizar grandes volúmenes de datos geocientíficos. Según un informe de MarketsandMarkets (2023), se espera que el mercado global de IA en el sector petrolero crezca de 2.85 mil millones de dólares en 2023 a 4.81 mil millones de dólares en 2027, con una tasa compuesta anual (CAGR) del 11.2%.

Este crecimiento se atribuye a varios factores, entre los que destacan:

- **Reducción de costos operativos:** La IA ha permitido a las empresas petroleras optimizar la producción en campos maduros, lo que ha reducido los costos operativos en un 15-25% en algunos casos. Esto incluye la optimización de la producción de pozos y la mejora en la recuperación de reservas mediante el uso de modelos predictivos.

- **Automatización avanzada:** El uso de técnicas de machine learning y deep learning para controlar equipos y sistemas en tiempo real ha permitido reducir el tiempo de inactividad no planificado en las plataformas petroleras en un 30-50%, según un estudio de McKinsey & Company (2022).

La inteligencia artificial ha comenzado a transformar la industria petrolera de manera significativa, optimizando operaciones, reduciendo costos y mejorando la sostenibilidad. Aunque existen desafíos, como la falta de personal capacitado, el avance continuo de esta tecnología promete revolucionar aún más el sector en los próximos años. Las principales compañías petroleras ya están experimentando los beneficios de la IA, y a medida que las barreras de entrada disminuyen, esta tendencia se extenderá a empresas de todos los tamaños.

Análisis de casos reales de la implementación de la inteligencia artificial en la industria petrolera.

Con los antecedentes propuestos, es fundamental analizar casos reales de implementación de la inteligencia artificial en la industria petrolera para entender con mayor profundidad cómo estas tecnologías están siendo aplicadas de manera concreta y los resultados tangibles que han generado.

A continuación, en la **Tabla 2** se mencionarán los casos específicos más relevantes de empresas petroleras que trabajan con inteligencia artificial, el tipo de operación que realizan, una descripción y un breve análisis de estos.

Esta refleja cómo la inteligencia artificial (IA) está siendo integrada estratégicamente en diversas áreas de la industria petrolera por grandes compañías que buscan aumentar la eficiencia, reducir costos y mejorar la seguridad operativa. Los ejemplos presentados muestran una tendencia clara: la inteligencia artificial se está utilizando no solo para optimizar procesos específicos, sino también para transformar la manera en que se manejan los datos y se toman decisiones clave.

Empresas	Técnica de IA	Área de la industria	Descripción
ExxonMobil y MIT	Robótica	Exploración	El equipo del MIT colabora con ingenieros de ExxonMobil para implementar un software de inteligencia artificial similar en robots de exploración submarina. Estos robots se emplearán para automatizar la detección y caracterización de filtraciones naturales.
Total y Google Cloud	Visión por computadora Aprendizaje profundo	Exploración y producción	Total Oil y Google Cloud colaboran para implementar un sistema de inteligencia artificial que mejore los procesos de exploración y producción de la empresa. Antes, los geólogos de Total analizaban manualmente datos sísmicos en imágenes, pero ahora buscan digitalizar este proceso utilizando IA para detectar patrones en el subsuelo que puedan indicar riesgos económicos en nuevos yacimientos petroleros. El objetivo es agilizar la recopilación y análisis de datos sísmicos.
Baker Hughes y NVIDIA	Aprendizaje profundo	Análisis integral de datos	Esta compañía de Colombia está empleando tecnologías de visión por computadora y procesamiento de lenguaje natural para automatizar la recuperación de información de documentos no estructurados, como informes de mantenimiento y operaciones.
Ecopetrol	Procesamiento de Lenguaje Natural Visión por computadora	Análisis de datos	Esta compañía de Colombia está empleando tecnologías de visión por computadora y procesamiento de lenguaje natural para automatizar la recuperación de información de documentos no estructurados, como informes de mantenimiento y operaciones.
Repsol	Aprendizaje automático	Ciberseguridad	Repsol está adoptando herramientas avanzadas de IA y aprendizaje automático para proteger sus activos energéticos y redes informáticas, mejorando la visibilidad y monitorización de sus entornos operativos
SLB	Creación de un ambiente cognitivo	Exploración y producción	La empresa SLB se encuentra inmersa en el desarrollo de un ambiente cognitivo innovador, que está redefiniendo la forma en que se abordan los desafíos en la industria petrolera. Este entorno revolucionario está compuesto por una suite integral de programas técnicos petroleros, diseñados para optimizar cada fase de la exploración, extracción y procesamiento de recursos energéticos. Además, este sistema pionero se apoya en una infraestructura cloud native de vanguardia, permitiendo una escalabilidad, flexibilidad y eficiencia sin precedentes. Esta arquitectura basada en la nube proporciona la capacidad de adaptarse dinámicamente a las demandas del mercado y a las necesidades específicas de cada proyecto, garantizando así una operación fluida y ágil. Asimismo, la empresa SLB ha consolidado un robusto ecosistema de datos, fundamentado en la recopilación, análisis y gestión inteligente de información relevante para la toma de decisiones estratégicas. Este conjunto de datos integra múltiples fuentes de información, desde datos geológicos y geofísicos hasta datos operativos y de rendimiento, proporcionando una visión holística y precisa del panorama petrolero.
Shell	Aprendizaje profundo Redes neuronales Procesamiento de Lenguaje Natural	Exploración y producción	Shell ha lanzado un asistente de inteligencia artificial, el primero en una empresa de lubricantes, para ayudar a sus clientes en la búsqueda de productos. Tradicionalmente, los clientes debían buscar manualmente entre una gran base de datos. Utilizando avatares como Emma y Ethan, Shell simplifica esta búsqueda a través de lenguaje natural. El asistente virtual opera en el sitio web de la empresa, ofreciendo detalles sobre la disponibilidad, tamaños de paquetes y especificaciones técnicas de los lubricantes.

Tabla 2. Casos reales de ejecución de IA en la Industria petrolera.

Elaboración propia a partir de datos de MIT, Intell y Emer

ExxonMobil, en colaboración con el MIT, se enfoca en el uso de robótica para mejorar la exploración submarina, una tarea que tradicionalmente es costosa y conlleva riesgos. Al automatizar la detección de filtraciones naturales, no solo se reducen los tiempos, sino también los costos y los riesgos asociados a estas operaciones.

Total y Google Cloud se centran en la exploración y producción utilizando aprendizaje profundo y visión por computadora. Esto refleja la capacidad de la IA para transformar procesos que antes requerían intervención humana intensiva, como el análisis de datos sísmicos, ahora realizados de manera más rápida y precisa, lo que acelera el descubrimiento de yacimientos.

Por otro lado, empresas como Repsol y Ecopetrol están aplicando la IA en áreas relacionadas con la ciberseguridad y el procesamiento de datos no estructurados, respectivamente. La implementación de estas tecnologías no solo optimiza la operación diaria, sino que mejora la seguridad de sus sistemas y la gestión de información crítica.

Finalmente, Baker Hughes y SLB muestran cómo la IA puede integrarse para abordar el análisis masivo de datos sísmicos y crear ambientes cognitivos que optimicen la exploración y producción en su totalidad. Esto resalta la capacidad de la IA para aportar valor en todas las fases del ciclo de vida del petróleo, desde la exploración hasta la producción y más allá.

En conjunto, estos ejemplos revelan que la IA no es una simple herramienta auxiliar, sino un factor disruptivo que está redefiniendo cómo se llevan a cabo operaciones clave en la industria petrolera. Las compañías están invirtiendo no solo en tecnología, sino también en colaboraciones estratégicas que les permiten estar a la vanguardia de esta transformación tecnológica.

Tendencias emergentes en la integración de la inteligencia artificial en los procesos subterráneos de la industria petrolera

Además de los datos tangibles sobre el uso real de la inteligencia artificial en las empresas petroleras que se han presentado, es importante destacar el estudio llevado a cabo por la empresa Kimberlite. Este estudio se centró en proyectar el uso esperado de plataformas habilitadas por IA para los flujos de trabajo subterráneos en el período comprendido entre 2022 y 2024.

Los resultados de esta encuesta, presentados en la **Figura 9**, ofrecen una visión integral sobre las tendencias emergentes en la integración de la inteligencia artificial (IA) en los procesos subterráneos de la industria petrolera. Desde la simulación avanzada de yacimientos hasta la optimización de la producción, las plataformas impulsadas por IA están desempeñando un papel cada vez más crucial para mejorar la eficiencia operativa, la precisión en la toma de decisiones y la rentabilidad en todas las fases del ciclo de vida del petróleo.

Estos datos proporcionan una perspectiva estratégica para las empresas del sector, permitiéndoles identificar áreas clave donde la IA puede generar el mayor impacto. Además, les facilita planificar sus inversiones y desarrollos tecnológicos de manera proactiva para maximizar las oportunidades emergentes. En un entorno competitivo y en constante cambio, donde la innovación y la eficiencia son esenciales, entender cómo la IA puede transformar los flujos de trabajo subterráneos es fundamental para asegurar la sostenibilidad y el éxito a largo plazo de las operaciones petroleras.

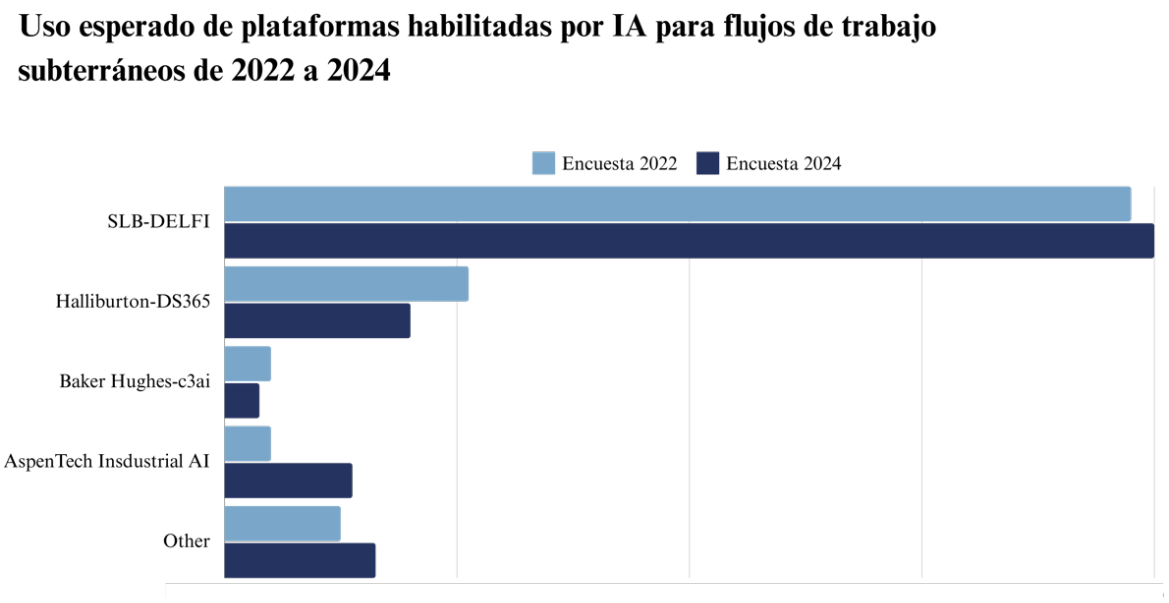

Figura 9. Uso esperado de plataformas de IA para flujos de trabajo subterráneos de 2022 a 2024. Elaboración propia con datos de Kimberlite,LLC-2024 G&G Software Supplier Performance Report.

La gráfica destaca el claro enfoque de las compañías petroleras en el uso de la inteligencia artificial para optimizar procesos clave, como la optimización de la producción y la simulación de yacimientos, áreas que lideran en términos de prioridad. Estas categorías subrayan cómo la IA está impulsando mejoras en eficiencia y precisión, factores esenciales en un entorno cada vez más competitivo y orientado a maximizar la rentabilidad operativa.

Otras áreas, como el mantenimiento predictivo y el análisis de datos geológicos, también tienen un impacto significativo, lo que indica que las empresas están utilizando IA para gestionar mejor los activos y reducir el tiempo de inactividad no planificado. La menor inversión en otras aplicaciones sugiere que estas tecnologías podrían estar en fases iniciales de adopción o desarrollo.

Los datos de la **Figura 9** revelan que las empresas están priorizando aquellas aplicaciones de IA que aportan mejoras inmediatas en la optimización de sus operaciones, mientras siguen explorando otras áreas con un potencial prometedor a largo plazo.

La inteligencia artificial en la sostenibilidad de la industria petrolera

La inteligencia artificial (IA), además de sus beneficios en la optimización de procesos y reducción de costos, está emergiendo como una herramienta esencial para promover prácticas más sostenibles y mitigar los impactos ambientales y sociales negativos. En la industria petrolera, la IA está siendo utilizada para abordar algunos de los desafíos más críticos en términos de sostenibilidad, desde la reducción de emisiones hasta la optimización del uso de recursos naturales y la mejora en la eficiencia energética.

En diversos sectores, la IA está jugando un papel crucial para avanzar hacia la sostenibilidad. Su capacidad para analizar grandes volúmenes de datos y tomar decisiones rápidas la convierte en un recurso invaluable para gestionar recursos de manera eficiente, optimizar la transición hacia energías renovables y minimizar el impacto ambiental de las operaciones industriales. No obstante, es vital que el desarrollo y la aplicación de estas tecnologías estén guiados por principios éticos y un enfoque claro en el bienestar a largo plazo de las personas y el planeta.

La IA tiene el potencial de facilitar el cumplimiento de los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas (**Figura 10**). Esto se debe a su capacidad para mejorar la gestión de recursos, prevenir crisis ambientales, optimizar la educación y la salud, y promover la equidad de género y el desarrollo económico inclusivo. En el contexto de la industria petrolera, la IA puede ser clave para mitigar el impacto ambiental y contribuir a la transición hacia un futuro más sostenible.

Figura 10. Objetivos de desarrollo sostenible
(Naciones Unidas, 2017)

La industria petrolera desempeña un rol crucial en la economía mundial, así como en muchas economías nacionales, incluidas aquellas en desarrollo y emergentes, siendo un pilar clave del sistema energético global que impulsa tanto el crecimiento económico como social. En concordancia con los Objetivos de Desarrollo Sostenible (ODS), que promueven la equidad social y la sostenibilidad ambiental, el PNUD, la Corporación Financiera Internacional y la IPIECA crearon en 2017 un Atlas que aborda los aspectos ambientales y sociales vinculados a esta industria.

Este Atlas no solo resalta la importancia de la industria del petróleo y gas en la economía, sino también su papel esencial en la lucha contra el cambio climático. Ofrece una serie de acciones que pueden contribuir a estabilizar las emisiones y las temperaturas, alineándose con el Acuerdo de París. Además, mapea las contribuciones actuales del sector e impulsa a las empresas a identificar nuevas oportunidades para avanzar hacia los 17 ODS, como se presenta en la **Figura 11** (IPIECA, 2017).

Figura 11. Mapeo de la industria petrolera y de gas hacia los Objetivos de Desarrollo Sostenible. IPIECA, 2017

El compromiso de las empresas de petróleo y gas con la sostenibilidad se evidencia en su enfoque integral hacia la integridad, la responsabilidad y la transparencia en todas sus actividades. Estas compañías no solo persiguen la rentabilidad, sino que también reconocen su obligación de generar un impacto positivo en la sociedad y en el entorno en el que operan. Adoptar una estrategia sostenible no solo responde a consideraciones éticas, sino que también puede ser ventajoso a largo plazo, dado que la conciencia y la demanda de sostenibilidad continúan aumentando globalmente.

Es importante destacar que, aunque los principales proyectos de IA en la industria petrolera se centran en mejorar la extracción, producción y distribución de hidrocarburos, muchos de ellos también están alineados con prácticas más sostenibles. Estas incluyen la adopción de tecnologías más limpias, la reducción de emisiones de gases de efecto invernadero, la implementación de programas de conservación ambiental y la participación en iniciativas de responsabilidad social corporativa. A medida que la conciencia sobre el cambio climático y la sostenibilidad sigue creciendo, se espera que estas prácticas se fortalezcan y se expandan.

Un caso destacado en este ámbito es el de Amplified Industries, una empresa que ha desarrollado sensores de bajo costo y larga duración para monitorear campos petrolíferos en tiempo real. Estos sensores, combinados con algoritmos de IA, permiten a los operadores detectar fugas de metano y derrames de salmuera, problemas que anteriormente requerían inspecciones manuales exhaustivas. Gracias a esta tecnología, los operadores pueden actuar antes de que estos problemas se conviertan en desastres ambientales.

Amplified Industries trabaja con varias empresas petroleras en Estados Unidos, operando miles de pozos, muchos de los cuales son antiguos y propensos a fallos. Sus soluciones no solo mejoran la eficiencia operativa, sino que también minimizan los riesgos ambientales al reducir las emisiones y prevenir fugas. Mannai afirma que la modernización de la infraestructura de datos y la integración de sensores avanzados será clave para asegurar una industria petrolera más eficiente y sostenible en el futuro.

La IA, con su capacidad para optimizar procesos y reducir el impacto ambiental, se posiciona como una fuerza transformadora en la industria de los hidrocarburos, no solo ayudando a las empresas a mejorar su rendimiento, sino también permitiendo una transición hacia operaciones más limpias y responsables.

Capítulo 3: Resultados y Análisis de Casos de Estudio

Caso de Estudio

En este apartado se examinará un estudio de caso realizado y publicado por la empresa Amplified Industries, centrado en los beneficios que ofrece la inteligencia artificial para la detección de problemas y el control inteligente en unidades de bombeo.

Beneficios de la detección de problemas con IA y el control inteligente de la Unidad de Bombeo

La industria petrolera ha confiado históricamente en sistemas manuales o parcialmente automatizados, los cuales presentan limitaciones para gestionar de manera eficaz desafíos dinámicos, como el comportamiento de los pozos horizontales. El caso de Amplified Industries ilustra de manera destacada cómo la Inteligencia Artificial (IA) está revolucionando este sector, particularmente en la mejora de la producción y el manejo de activos.

Se implementó un sistema completo, que incluye el paquete de monitoreo de pozos y el sistema de monitoreo de tanques. Cada sistema se instala en menos de 30 minutos e incluye todo lo necesario para instrumentar un activo. Además de ofrecer automatización y monitoreo de campo completo, el sistema proporciona a los usuarios un nivel de profundidad de datos extraordinario, incluyendo la capacidad de detectar, diagnosticar y optimizar condiciones del subsuelo directamente en el cabezal del pozo.

A continuación, se analizan los elementos principales del caso de estudio en cuestión:

Contexto y Problemas

Los pozos analizados pertenecen a operadores independientes en Texas y Oklahoma, están en declive y se enfrentan a desafíos operativos que dificultan su producción eficiente.

Retos operativos:

- Los pozos enfrentan el desafío de mantener la eficiencia en activos en declive y técnicamente complejos.
- Problemas específicos incluyen el "slugging" (acumulación de gas y líquidos), ineficiencia en bombas, y diagnósticos poco claros en tiempo real.

Solución Propuesta

La empresa Amplified Industries implementó su sistema completo llamado Amplified Wells, que incluye:

- Controladores inteligentes conectados.
- Monitoreo de tanques y pozos.
- Herramientas avanzadas de diagnóstico y optimización, con una instalación rápida (menos de 30 minutos).

El sistema proporciona datos detallados y en tiempo real, como gráficas de bombeo por cada ciclo, accesibles a través de plataformas web y móviles. Permite:

- Detectar problemas en la superficie y en el subsuelo, como fugas de tuberías, bloqueo por gas, piezas desgastadas o ineficiencia general de la bomba.
- Optimizar automáticamente el rendimiento del pozo y generar alertas de alto impacto.

Solución basada en IA:

La implementación del sistema Amplified Wells representa una evolución hacia la adopción de IA en la industria. Este sistema incluye:

- Monitoreo y Control Automatizado: La IA procesa datos en tiempo real para optimizar parámetros de bombeo y detectar problemas.
- Detección y Diagnóstico: Algoritmos inteligentes detectan anomalías como fugas, bloqueo por gas o piezas desgastadas con mayor precisión y rapidez que las inspecciones tradicionales.

- Optimización Continua: El sistema ajusta automáticamente las operaciones para minimizar tiempo de funcionamiento, consumo energético y desgaste, manteniendo niveles óptimos de producción.

Impacto cuantificable:

Para un subconjunto de pozos en el centro de Oklahoma, la optimización de control logró reducir el tiempo total de operación en más del 50% sin afectar la producción total. No solo se redujo significativamente el consumo de electricidad, sino que las estimaciones indicaron que la vida útil del equipo se duplicará efectivamente, ya que el llenado promedio de las bombas ha aumentado cerca del 100% y los golpes de fluido han sido casi completamente eliminados (**Figura 12**).

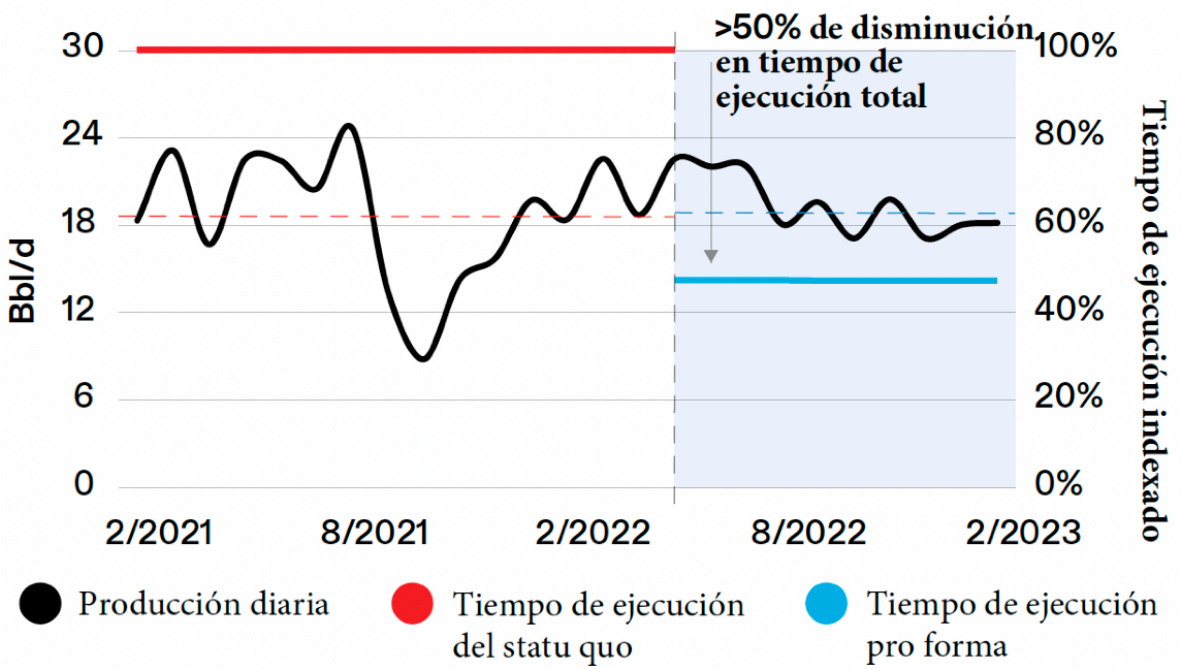

Figura 12. Impacto cuantificable de Amplified Wells en subconjunto de Oklahoma. Amplified Industries (2023)

En una serie de pozos horizontales ubicados entre Texas y el centro de Oklahoma, se han observado comportamientos significativos de slugging de gas. Al implementar el sistema en estos pozos, los operadores lograron reducir el tiempo de operación en un 30% o más sin afectar la producción. (**Figura 13**) Lo más importante es que el sistema de control activo estabilizó el slugging de gas, haciendo que el comportamiento de los pozos sea mucho más predecible y fácil de manejar, lo que prolonga considerablemente la vida útil del equipo y reduce los costos de levantamiento

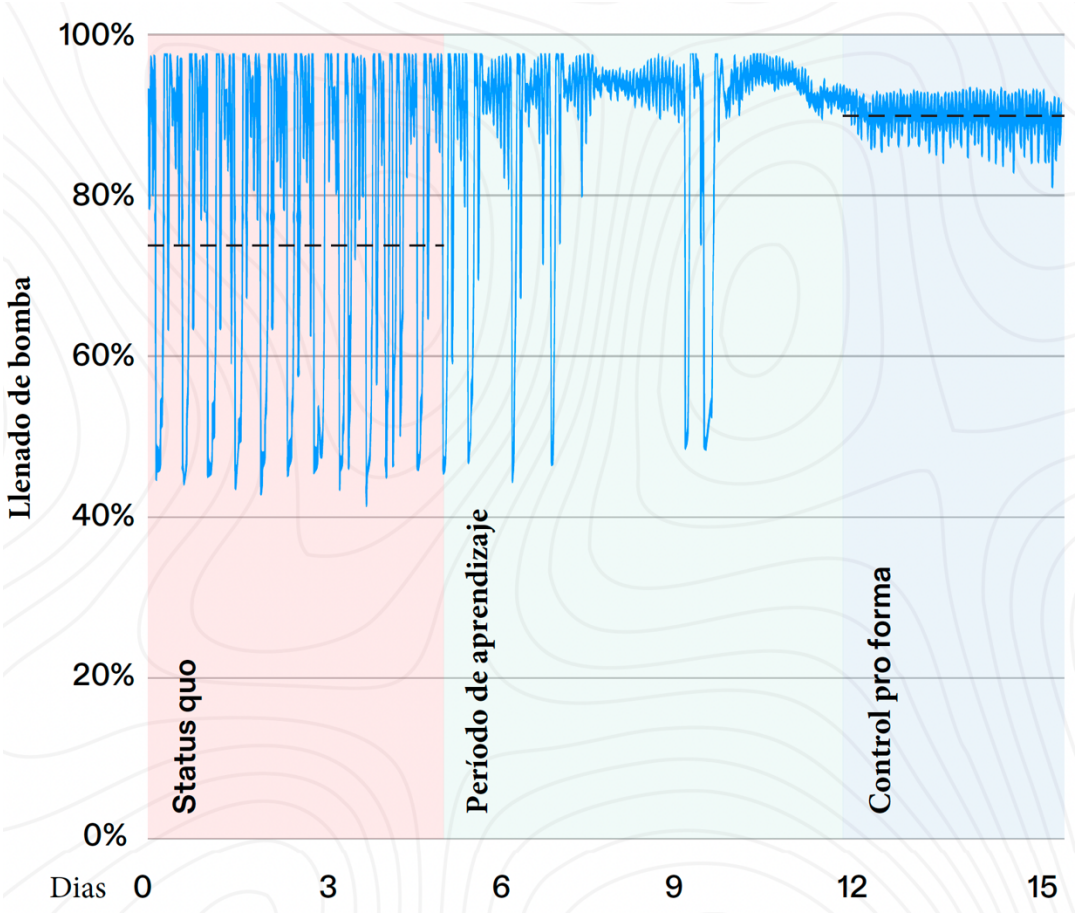

Figura 13. Impacto cuantificable de Amplified Wells ante situación de slugging
Amplified Industries (2023)

Impacto Económico y Ambiental

- **Reducción del Consumo Energético y Costos Operativos:** La optimización del funcionamiento de las unidades de bombeo permite una operación más eficiente al reducir significativamente el tiempo de actividad innecesaria. Esto genera ahorros en el consumo eléctrico, ya que las bombas operan únicamente cuando es necesario, minimizando los picos de demanda energética. Además, la reducción del uso excesivo de las bombas prolonga su vida útil aproximadamente al doble. Como resultado, se disminuyen los costos asociados con la compra e instalación de nuevos equipos, así como los gastos de mantenimiento no planificado.

- **Disminución de la Huella de Carbono:** La menor dependencia de energía, combinada con una optimización en los recursos utilizados, tiene un impacto directo en la sostenibilidad ambiental:

    - Reducción de emisiones de gases de efecto invernadero: Un menor consumo energético implica una menor generación de emisiones, especialmente en regiones donde la electricidad proviene de fuentes fósiles.
    - Producción más sostenible: La estabilización de los procesos operativos reduce las fugas o desperdicios de gas y líquidos, disminuyendo la contaminación ambiental.

- **Mantenimiento Preventivo y Mitigación de Problemas:** La IA permite prever fallas en el sistema antes de que estas se conviertan en problemas mayores. Esto impacta en la reducción de tiempos de inactividad, disminución del mantenimiento correctivo y se traduce en una mejor planificación de recursos, ya que, con una detección precisa de problemas,

las empresas pueden optimizar la logística del del personal y los materiales requeridos para reparaciones, reduciendo costos adicionales.

Conclusión

El caso de Amplified Industries demuestra cómo la implementación de soluciones avanzadas basadas en Inteligencia Artificial puede revolucionar la gestión y optimización de activos en la industria petrolera. La solución Amplified Wells no solo aborda problemas operativos comunes, como el "slugging" y la ineficiencia del bombeo, sino que también permite a los operadores mantener una producción estable, reducir costos y prolongar la vida útil de los equipos.

Entre los principales logros destacan la reducción del 50% en el tiempo de operación, el aumento del 20% en el llenado de las bombas y el duplicar la vida útil de los equipos. Además, la capacidad de diagnosticar y resolver problemas en tiempo real minimiza tiempos de inactividad y costos imprevistos, mientras que la disminución del consumo energético y la huella de carbono subraya un compromiso con la sostenibilidad.

Este caso es un ejemplo claro de cómo la tecnología inteligente puede transformar la eficiencia, rentabilidad y sostenibilidad de las operaciones en la industria del petróleo y gas. Soluciones como Amplified Wells representan el futuro de la industria, al combinar innovación tecnológica con resultados prácticos y medibles.

Capítulo 4: Implementación de IA en la Industria Petrolera

Tecnologías de Inteligencia Artificial con potencial de aplicarse para la Industria Petrolera

La integración de la inteligencia artificial (IA) en la industria petrolera está generando beneficios sustanciales en términos de eficiencia, seguridad y sostenibilidad. Las compañías petroleras están utilizando IA para optimizar procesos que abarcan desde la exploración de yacimientos y la producción de hidrocarburos hasta la distribución y comercialización de productos. Esto ha permitido a las empresas aumentar la productividad, reducir costos y mejorar significativamente los tiempos de ejecución.

Uno de los principales beneficios de la IA en este sector es su capacidad para mejorar la eficiencia operativa. Los algoritmos avanzados de aprendizaje automático permiten a las empresas optimizar la producción de pozos, identificar yacimientos con mayor precisión y predecir el rendimiento de activos. Además, la IA facilita la seguridad en la industria petrolera al emplear robots y sistemas automatizados para realizar tareas peligrosas y monitorear las condiciones de trabajo en tiempo real, mitigando riesgos para los empleados y el medio ambiente.
Otro aspecto clave es la capacidad de la IA para impulsar la innovación.

Las tecnologías de procesamiento de lenguaje natural (PLN), por ejemplo, están permitiendo a las empresas petroleras desarrollar nuevos servicios y mejorar la experiencia del cliente mediante asistentes virtuales que proporcionan soporte técnico, agilizan la resolución de consultas y optimizan la selección de productos.
A continuación, se presenta un análisis y evaluación de algunas de las plataformas y técnicas de IA más relevantes y con mayor potencial de aplicación en la industria petrolera.

Evaluación de plataformas y técnicas disponibles

IBM Watson

IBM Watson es una plataforma avanzada de inteligencia artificial y análisis de datos que ha demostrado ser eficaz en el manejo de grandes volúmenes de información. Su capacidad para ejecutar procesos de aprendizaje automático y análisis predictivo lo convierte en una solución clave para la industria petrolera.

Google Cloud AI

Google Cloud AI es una plataforma poderosa que ofrece herramientas avanzadas de aprendizaje automático, análisis de big data y gestión de infraestructuras en la nube.
Google Cloud AI sobresale por su escalabilidad y su capacidad para procesar datos en tiempo real, ofreciendo una ventaja competitiva en la gestión de grandes volúmenes de información. La infraestructura de Google permite una integración fluida en la nube, lo que es esencial para las operaciones distribuidas de la industria petrolera.

Microsoft Azure AI

Microsoft Azure AI ofrece una gama completa de servicios de inteligencia artificial, todos ellos integrados en su plataforma en la nube. Microsoft Azure AI es ideal para empresas petroleras que buscan integrar plataformas de IA con sistemas preexistentes. Su capacidad para ejecutar simulaciones y análisis predictivos proporciona una solución versátil que mejora la toma de decisiones y la planificación operativa.

C3.ai

C3.ai es una plataforma sólida para las empresas energéticas debido a su enfoque especializado en mejorar la eficiencia operativa y reducir el impacto ambiental. Su capacidad para integrar datos operativos en tiempo real proporciona una visión integral para la toma de decisiones estratégicas.

Petroleum Experts (PETEX)

Petroleum Experts (Petex) es una plataforma avanzada de software de simulación y modelado diseñada específicamente para la industria del petróleo y gas. Su capacidad para modelar yacimientos y realizar simulaciones de flujo lo convierte en una herramienta esencial para las empresas que buscan optimizar la producción y mejorar la gestión de sus recursos.

Consideraciones en la implementación de plataformas de IA

Las plataformas de inteligencia artificial ofrecen ventajas competitivas significativas para la industria petrolera, pero su implementación exitosa depende de varios factores clave:

- Integración: La compatibilidad de las plataformas de IA con los sistemas preexistentes es fundamental para garantizar una adopción eficiente y sin interrupciones.
- Escalabilidad: Es crucial que las plataformas sean capaces de manejar el creciente volumen de datos que genera la industria en todas sus fases.
- Especialización: Plataformas como C3.ai y Petex, diseñadas específicamente para las necesidades del sector energético, ofrecen ventajas claras en el mantenimiento predictivo y modelado de yacimientos.

Propuesta Integral para la implementación exitosa de Inteligencia Artificial en la Industria Petrolera: un enfoque estratégico y práctico

La industria petrolera se enfrenta a un punto crítico donde la eficiencia operativa, la reducción de costos y la sostenibilidad deben equilibrarse en un entorno de crecientes presiones regulatorias y demandas medioambientales. En este contexto, la inteligencia artificial (IA) se posiciona como una tecnología transformadora que tiene el potencial de optimizar procesos, mejorar la toma de decisiones y proporcionar un enfoque predictivo para la gestión de activos. No obstante, una implementación exitosa de IA en este sector requiere una estrategia integral que aborde no solo los aspectos tecnológicos, sino también las consideraciones éticas, operativas y regulatorias.

Para que las empresas petroleras aprovechen al máximo el potencial de la IA y transformen sus operaciones, es fundamental un enfoque estratégico que contemple tanto las necesidades tecnológicas como la creación de un entorno propicio para la integración eficiente de estas soluciones. Esta propuesta describe los objetivos principales y los pasos prácticos para lograr una implementación eficaz y sostenible de la IA en la industria petrolera.

Objetivos estratégicos

Optimización de operaciones

El primer objetivo es mejorar la eficiencia operativa mediante la automatización de procesos clave en las áreas de exploración, perforación y producción. La IA puede integrarse con los sistemas de control existentes para:

- **Automatizar tareas repetitivas y complejas**, lo que minimiza la intervención humana en actividades riesgosas o de alto consumo de tiempo.

- **Optimizar el uso de recursos energéticos y materiales**, reduciendo el desperdicio y maximizando la eficiencia.
- **Utilizar modelos predictivos** que permitan una planificación más precisa de las operaciones, minimizando los riesgos de interrupciones inesperadas.

Impacto esperado: Reducción de costos operativos, mayor eficiencia y una mejora en la seguridad operativa al reducir la exposición a tareas peligrosas.

Mantenimiento predictivo

Uno de los beneficios más inmediatos de la IA es su capacidad para anticipar fallos en los equipos y prever necesidades de mantenimiento mediante el uso de datos en tiempo real. La implementación de sistemas de mantenimiento predictivo permite:

- **Analizar en tiempo real los datos operativos de equipos y maquinaria**, identificando patrones que indiquen deterioro o problemas inminentes.
- **Aplicar modelos de aprendizaje automático** basados en datos históricos para prever fallos, evitando paradas no programadas.
- **Planificar el mantenimiento proactivo**, lo que reduce costos y maximiza la vida útil de los activos.
- **Impacto esperado:** Aumento de la eficiencia operativa, reducción de costos asociados a mantenimientos de emergencia y prolongación del ciclo de vida de los equipos.

Toma de decisiones basada en datos

La capacidad de la IA para procesar grandes volúmenes de datos en tiempo real permite a las empresas petroleras tomar decisiones más precisas y fundamentadas. Esto incluye:

- **Análisis de datos provenientes de múltiples fuentes**, como sensores, equipos de perforación y datos geológicos, que mejoran la visibilidad de las operaciones.
- **Modelos predictivos y simulaciones**, que permiten evaluar el impacto potencial de diferentes decisiones antes de implementarlas.
- **Mejora en la transparencia y trazabilidad**, lo que facilita la colaboración entre equipos y mejora la alineación con los objetivos estratégicos.

Impacto esperado: Mayor precisión en las decisiones operativas, optimización de los recursos y capacidad de respuesta ante imprevistos, mejorando la competitividad.

Sostenibilidad y reducción de impacto ambiental

Un enfoque clave de la implementación de IA debe centrarse en la sostenibilidad, ayudando a las empresas a minimizar su impacto ambiental. La IA puede:

- **Optimizar el consumo energético** en las operaciones de perforación, producción y transporte, reduciendo las emisiones y mejorando la eficiencia.
- **Monitorear en tiempo real** las emisiones y las condiciones ambientales en los sitios de operación, garantizando el cumplimiento normativo.
- **Facilitar el cumplimiento de las normativas ambientales** mediante reportes automáticos y alertas para una mejor gestión del impacto ambiental.

Impacto esperado: Reducción del impacto ambiental, mejora en la eficiencia energética y cumplimiento con las regulaciones internacionales de sostenibilidad.

Enfoque estratégico y práctico

Para implementar la IA de manera efectiva en la industria petrolera, es necesario seguir una estrategia estructurada que contemple los siguientes pasos:

Evaluación inicial de infraestructura y capacidades

- Realizar una auditoría detallada de la infraestructura tecnológica existente para identificar las brechas tecnológicas y determinar cómo la IA puede generar el mayor impacto.
- Definir los objetivos y el retorno de inversión esperado en función de los resultados específicos que se desean alcanzar con la IA.

Desarrollo de capacidades internas

- Capacitar al personal en el uso y gestión de herramientas de IA, asegurando que los equipos puedan aprovechar al máximo las nuevas tecnologías.
- Formar equipos multidisciplinarios que combinen el conocimiento técnico de la IA con la experiencia operativa y gerencial.

Selección de tecnologías y proveedores

- Identificar las plataformas y soluciones de IA más adecuadas, priorizando aquellas que sean escalables y personalizables para las necesidades de la industria petrolera.

- Considerar el uso de soluciones basadas en la nube que permitan una gestión eficiente de grandes volúmenes de datos y ofrezcan flexibilidad para futuras expansiones.

Implementación piloto

- Iniciar con proyectos piloto en áreas específicas de la operación para probar las capacidades de la IA y ajustar los modelos antes de la implementación generalizada.
- Usar los resultados de los pilotos como base para la expansión de la IA en otras áreas de la empresa.

Monitoreo y mejora continua

- Establecer sistemas de monitoreo en tiempo real para medir el rendimiento de la IA y realizar ajustes continuos a los modelos.
- Auditar regularmente los sistemas de IA para garantizar que sigan alineados con los objetivos estratégicos y mantener su relevancia en un entorno cambiante.

La implementación exitosa de la IA en la industria petrolera no es simplemente una cuestión de adopción tecnológica, sino de alineación estratégica con los objetivos operativos y de sostenibilidad a largo plazo.

Capítulo 5: Implementación de la Inteligencia Artificial en la Metodología UNFC de las Naciones Unidas

UNFC: Clasificación Marco de las Naciones Unidas para los Recursos

La Clasificación Marco de Naciones Unidas para Recursos (UNFC) es una metodología reconocida a nivel internacional que permite evaluar y gestionar recursos mediante un marco coherente y estandarizado. Este sistema clasifica los proyectos según su viabilidad técnica, económica y ambiental. La incorporación de la inteligencia artificial (IA) a esta metodología representa una oportunidad innovadora para optimizar el proceso, incrementando la precisión y la eficiencia en la gestión de los recursos naturales.

La IA, con su capacidad para procesar grandes volúmenes de datos, identificar patrones y generar análisis predictivos, proporciona una base robusta para automatizar y perfeccionar la aplicación del UNFC. Mediante algoritmos de aprendizaje automático, es posible evaluar de manera más eficiente parámetros clave que determinan la viabilidad de los proyectos, como los niveles de reservas, la factibilidad técnica y el impacto ambiental, entre otros.

Beneficios de la integración de la IA en la Metodología UNFC

Automatización del Proceso de Clasificación: La inteligencia artificial (IA) permite automatizar el análisis y la clasificación de recursos según los criterios del UNFC, agilizando la toma de decisiones y minimizando la necesidad de intervención manual.

Procesamiento de Grandes Volúmenes de Datos: Con la integración de la IA, el sistema es capaz de procesar y analizar grandes conjuntos de datos geológicos, económicos y técnicos de manera eficiente, mejorando significativamente la precisión en la evaluación de los proyectos.

Actualización y Monitoreo Continuo: Los algoritmos de IA pueden adaptarse automáticamente a medida que se incorporan nuevos datos, asegurando que las evaluaciones sean precisas, actualizadas y basadas en la información más reciente.

Modelado Predictivo: Mediante técnicas avanzadas de IA, como el aprendizaje supervisado y no supervisado, es posible prever el desempeño a largo plazo de los proyectos, identificando con antelación riesgos y oportunidades potenciales.

Aplicaciones Prácticas de la IA en UNFC

La incorporación de inteligencia artificial (IA) en el marco UNFC permite automatizar procesos clave en la clasificación de recursos, mejorando la precisión y eficiencia en la gestión. Entre sus aplicaciones destacan:

- **Evaluación de la Viabilidad Económica:** Los modelos predictivos basados en IA analizan datos económicos históricos y actuales para evaluar la viabilidad a largo plazo de un proyecto, lo que facilita decisiones más informadas y estratégicas.
- **Optimización de Parámetros Técnicos:** La IA procesa datos técnicos complejos, como la capacidad de extracción de recursos y el estado de la infraestructura, permitiendo ajustar los planes de desarrollo de manera más precisa y eficiente.
- **Análisis de Impacto Ambiental:** Los algoritmos de IA contribuyen a evaluar y mitigar los impactos ambientales de los proyectos, garantizando el cumplimiento de normativas internacionales y fomentando prácticas sostenibles.

Implementación y Sostenibilidad

Este enfoque basado en inteligencia artificial (IA) tiene la capacidad de evolucionar de forma continua gracias a la integración de herramientas de monitoreo y actualización automática. A medida que los proyectos avanzan y se recopilan nuevos datos, el sistema puede ajustar sus predicciones y recomendaciones, asegurando que los proyectos de recursos sean cada vez más eficientes, sostenibles y rentables.

El principal objetivo de esta integración es demostrar la viabilidad y los beneficios de incorporar la IA para complementar y optimizar la metodología UNFC. Además de automatizar y mejorar los procesos actuales, esta tecnología proporciona una base más robusta para la toma de decisiones estratégicas en la gestión de recursos naturales a escala global.

Metodología UNFC

La Clasificación Marco de las Naciones Unidas para los Recursos (UNFC, por sus siglas en inglés) es un sistema desarrollado por la Comisión Económica de las Naciones Unidas para Europa (UNECE). Su objetivo principal es ofrecer un marco unificado y estandarizado para la evaluación y gestión de recursos naturales. Este sistema permite analizar la viabilidad ambiental, social, económica y técnica de proyectos relacionados con la explotación de recursos, proporcionando una clasificación clara en función de su madurez y sostenibilidad (UNFC, 2019).

Objetivo Principal

El UNFC tiene como objetivo principal proporcionar un marco coherente para evaluar los recursos naturales considerando su viabilidad en áreas clave: ambiental, social, económica y técnica. Su estructura facilita la toma de decisiones estratégicas sobre la factibilidad de los proyectos, asegurando su alineación con los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas. Esto lo convierte en una herramienta esencial para fomentar prácticas responsables y sostenibles en la gestión de recursos como hidrocarburos, minerales, energías renovables y aguas subterráneas.

La actualización del UNFC en 2019 amplió significativamente su alcance, permitiendo su aplicación en una mayor variedad de recursos y sectores. Además, reforzó su relevancia para las industrias que buscan adoptar prácticas sostenibles y alinearse con objetivos de desarrollo a largo plazo.

¿Cómo funciona el UNFC?

El sistema UNFC utiliza un enfoque tridimensional para la clasificación de recursos (**Figura 14**), basado en tres ejes principales:

1. **Eje E**
   Evalúa la viabilidad ambiental, social y económica de los proyectos (E1, E2, E3).

2. **Eje F**
   Considera la factibilidad técnica y madurez del proyecto (F1, F2, F3, F4).

3. **Eje G:**
   Mide el grado de certeza en la estimación de los recursos disponibles (G1, G2, G3, G4).

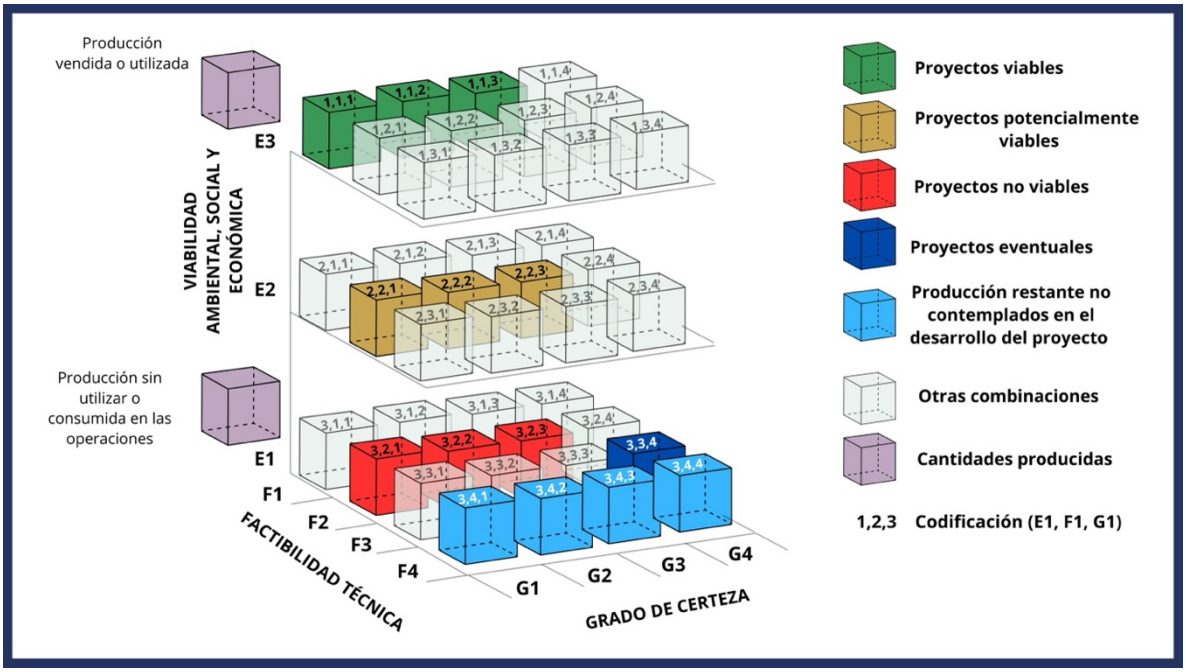

Figura 14. UNFC: Clasificación Marco de las Naciones Unidas para los Recursos.
Elaboración propia con datos de Naciones Unidas

La combinación de estos ejes genera una clasificación tridimensional que permite evaluar la madurez de un proyecto, la certeza sobre los recursos disponibles y su viabilidad desde el punto de vista de la sostenibilidad. Esta estructura multidimensional es fundamental para garantizar una evaluación exhaustiva y precisa de los proyectos en diversos sectores, como el petrolero.

Clases y Subclases

El UNFC clasifica los proyectos en diversas clases y subclases mediante la combinación de tres ejes principales: **E** (viabilidad económica y social), **F** (madurez técnica) y **G** (grado de certeza de los recursos). Estas clases reflejan la viabilidad del proyecto, su estado de desarrollo técnico y la certeza de los recursos disponibles, ofreciendo una visión integral del estado actual de los proyectos y su potencial futuro. La **Tabla 3** presenta una descripción detallada de las clases y subclases más relevantes.

<table>
<tr><td rowspan="13">Cantidades Totales</td><td rowspan="2">Extraídas</td><td colspan="4">Producción destinada a la venta</td></tr>
<tr><td colspan="4">Producción no destinada a la venta</td></tr>
<tr><td colspan="2" rowspan="2">Clase</td><td rowspan="2">Sub-Clase</td><td colspan="3">Categorías Mínimas</td></tr>
<tr><td>E</td><td>F</td><td>G</td></tr>
<tr><td rowspan="7">Recursos conocidos</td><td rowspan="3">Proyectos comerciales</td><td>En producción</td><td>1</td><td>1.1</td><td>1,2,(3)</td></tr>
<tr><td>Aprobado para Desarrollo</td><td>1</td><td>1.2</td><td>1,2,(3)</td></tr>
<tr><td>Justificado para Desarrollo</td><td>1</td><td>1.3</td><td>1,2,(3)</td></tr>
<tr><td rowspan="2">Proyectos con potencial comercial</td><td>Desarrollo pendiente</td><td>2</td><td>2.1</td><td>1,2,3</td></tr>
<tr><td>Desarrollo en Pausa</td><td>2</td><td>2.2</td><td>1,2,3</td></tr>
<tr><td rowspan="2">Proyectos no comerciales</td><td>Desarrollos no Clarificado</td><td>3.2</td><td>2.2</td><td>1,2,3</td></tr>
<tr><td>Desarrollo no Viable</td><td>3.3</td><td>2.3</td><td>1,2,3</td></tr>
<tr><td colspan="2">Cantidades remanentes no desarrollados de proyectos identificados</td><td>3.3</td><td>4</td><td>1,2,3</td></tr>
<tr><td rowspan="2">Recursos Potenciales</td><td>Proyectos de exploración</td><td>[No cuenta con Sub-Clases definidas]</td><td>3.2</td><td>3</td><td>4</td></tr>
<tr><td colspan="2">Cantidades remanentes no desarrolladas de proyectos prospectivos</td><td>3.3</td><td>4</td><td>4</td></tr>
</table>

Tabla 3. Clases y Subclases de UNFC.

Elaboración propia con datos de Naciones Unidas

Este enfoque estructurado no solo es relevante para la clasificación de recursos, sino también para guiar decisiones financieras, estratégicas y operativas en el manejo de proyectos de recursos naturales.

Ventajas y Desventajas del UNFC

Ventajas

- **Transparencia y uniformidad:** Ofrece un marco estandarizado que permite realizar evaluaciones consistentes en distintas jurisdicciones y sectores.
- **Orientación hacia la sostenibilidad:** Integra criterios ambientales y sociales, promoviendo la alineación de la gestión de recursos con los Objetivos de Desarrollo Sostenible (ODS).
- **Versatilidad:** Es aplicable a una amplia variedad de recursos, desde energías renovables hasta recursos fósiles, lo que lo convierte en una herramienta flexible y adaptable.

Desventajas

- **Complejidad:** La clasificación basada en múltiples ejes y subcategorías puede resultar complicada para organizaciones que no disponen de personal capacitado en el sistema.
- **Dependencia de datos de alta calidad:** Su eficacia está directamente relacionada con la disponibilidad de datos precisos y completos, lo que puede ser un desafío en regiones remotas o con infraestructura limitada.

Uso del UNFC en la Industria Petrolera

En el sector petrolero, el UNFC se emplea principalmente para clasificar y gestionar los recursos de hidrocarburos, proporcionando a las empresas un marco integral para evaluar proyectos de exploración y producción. Este enfoque multidimensional considera no solo la viabilidad económica, sino también los impactos sociales y ambientales. Algunas de las aplicaciones más relevantes incluyen:

- **Evaluación de reservas y recursos:** Clasificar los hidrocarburos en función de su certeza y viabilidad económica, lo que permite priorizar proyectos más seguros, rentables y sostenibles.
- **Planificación estratégica:** Facilitar el desarrollo de planes a largo plazo, identificando los recursos más valiosos y rentables según las condiciones del mercado y las regulaciones ambientales, optimizando así la toma de decisiones empresariales.

Cómo la IA puede Automatizar el Proceso de Evaluación del UNFC

La inteligencia artificial (IA) tiene el potencial de transformar la evaluación y clasificación de proyectos dentro del marco del UNFC, gracias a su capacidad para automatizar procesos y mejorar la precisión en el análisis. Integrando algoritmos avanzados de aprendizaje automático, procesamiento de datos geológicos y modelos predictivos, las aplicaciones más destacadas de la IA en este contexto incluyen:

1. **Automatización de la recopilación y procesamiento de datos:** Sensores geológicos en tiempo real, datos de producción y métricas económicas pueden ser procesados por sistemas de IA, generando clasificaciones precisas y actualizadas de los proyectos dentro del UNFC.
2. **Modelos predictivos para evaluaciones continuas:** La IA permite realizar evaluaciones predictivas constantes, ajustando automáticamente las clasificaciones en función de los nuevos datos recibidos, garantizando así una gestión dinámica y precisa.
3. **Monitoreo en tiempo real para la optimización de recursos:** Combinada con tecnologías del internet de las cosas, la IA facilita el monitoreo continuo de operaciones, detectando anomalías y ajustando las clasificaciones del UNFC con base en datos actualizados, optimizando el uso de recursos y reduciendo riesgos.

4. **Automatización del cumplimiento normativo:** La IA puede supervisar el cumplimiento de regulaciones ambientales y elaborar reportes automáticos, asegurando que los proyectos clasificados bajo el UNFC operen dentro de los límites legales establecidos.

El UNFC es una herramienta poderosa para la evaluación y gestión de recursos en la industria petrolera, proporcionando un marco estandarizado y multidimensional que aborda los aspectos económicos, técnicos y ambientales de los proyectos. La integración de inteligencia artificial tiene el potencial de automatizar y mejorar significativamente este proceso, permitiendo evaluaciones más rápidas, precisas y sostenibles, y ayudando a las empresas petroleras a tomar decisiones informadas que alineen sus operaciones con los objetivos de desarrollo sostenible.

Aplicación de la Inteligencia Artificial en la Clasificación Marco de las Naciones Unidas para los Recursos (UNFC)

El proyecto ha sido concebido como la fase inicial de un sistema más amplio que abarcará todos los ejes del UNFC (**E**, **F** y **G**). Para su desarrollo, el programa se ha entrenado utilizando datos ficticios creados en colaboración con el Centro de Excelencia para la Administración de Recursos Sostenibles (ICE-SRM) de Latinoamérica y el Caribe de la ONU-UNECE. Este sistema servirá como fundamento para un proyecto más ambicioso que permitirá clasificar proyectos a mayor escala en el futuro.

El enfoque modular del proyecto facilita una implementación gradual, adaptándose a las necesidades específicas de empresas y gobiernos. Abarca aspectos clave como la viabilidad económica, técnica, social, ambiental y geológica de los proyectos, garantizando una integración escalable y alineada con los objetivos de sostenibilidad.

Metodología de desarrollo por ejes

La metodología del UNFC es un marco sólido y estandarizado que clasifica proyectos de recursos naturales en función de su viabilidad técnica, económica y el grado de certeza de los recursos. Su estructura lógica y jerárquica lo hace ideal para ser automatizado mediante inteligencia artificial (IA). Utilizando técnicas como el aprendizaje automático (*machine learning*) y las redes neuronales, la IA puede aprender de datos históricos de proyectos y realizar predicciones o clasificaciones de nuevos proyectos con alta precisión.

La integración de la IA con el UNFC permite automatizar la clasificación de proyectos en los tres ejes clave del marco:

Eje E (viabilidad económica y social): Se pueden emplear redes neuronales para analizar factores económicos y sociales, detectando patrones y tendencias relevantes para clasificar proyectos.

Eje F (viabilidad técnica): Algoritmos como árboles de decisión pueden ser utilizados para evaluar y clasificar proyectos según su madurez técnica y desarrollo.

Eje G (certeza geológica): Modelos de aprendizaje automático pueden procesar datos geológicos para determinar el nivel de certeza de los recursos asociados a cada proyecto.

El proceso de automatización se basa en el entrenamiento de modelos de *machine learning* con conjuntos de datos históricos de proyectos previamente clasificados bajo el UNFC. Estos modelos identifican patrones complejos en los datos que vinculan las características de los proyectos con sus clasificaciones respectivas. Esto no solo aumenta la eficiencia del sistema, sino que también mejora la precisión y uniformidad en la evaluación y clasificación de proyectos.

Eje F (Factibilidad técnica)

El **Eje F** del UNFC evalúa la viabilidad técnica de la extracción en un proyecto en desarrollo.

La **Tabla 4** presenta un mapeo detallado que facilita la evaluación y clasificación de la Factibilidad Técnica de un proyecto bajo el Eje F del marco UNFC. Este mapeo, elaborado a partir de datos proporcionados por la ONU, tiene como objetivo categorizar los proyectos según su grado de madurez técnica y su viabilidad de ejecución.

Categoría UNFC	Definición
F1	La factibilidad técnica del proyecto de desarrollo se ha confirmado
F2	La factibilidad técnica del proyecto de desarrollo este sujeto a una evaluación más profunda
F3	La factibilidad técnica del proyecto de desarrollo no puede evaluarse debido a la escasez de los datos
F4	No se ha identificado ningún proyecto de desarrollo

Tabla 4. Matriz de Evaluación del eje F

Elaboración propia con datos de Naciones Unidas

Esquema del Eje F

El esquema del Eje F (**Figura 15**) representa un diagrama de flujo que guía la evaluación de un proyecto en diferentes etapas, desde la producción hasta la aprobación de planes de desarrollo y migración.

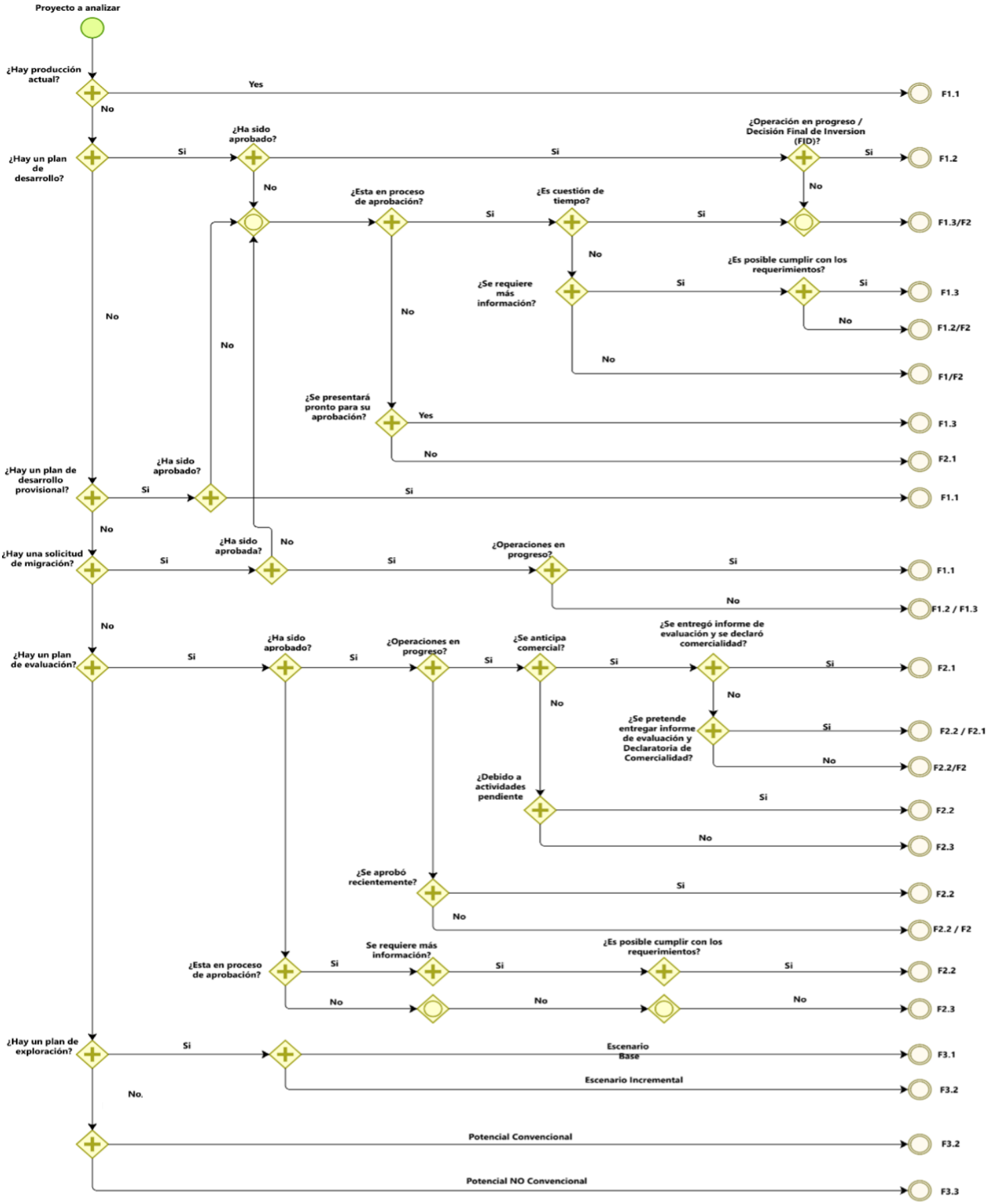

Figura 15. Diagrama de flujo para evaluar el Eje F.
Elaboración propia con datos de ICE-SRM y CNH

En este diagrama, se toman decisiones clave sobre la viabilidad del proyecto, como la existencia de un plan de desarrollo, la aprobación de operaciones, y la comercialidad esperada. Cada decisión lleva a una clasificación en una de las categorías del Eje F:

- F1 (viabilidad técnica confirmada)
- F2 (viabilidad técnica sujeta a evaluación adicional)
- F3 (viabilidad técnica incierta)

Es esencial porque establece un proceso estructurado para evaluar los proyectos en función de su madurez técnica y estado de aprobación, facilitando la identificación de posibles obstáculos técnicos y áreas que requieren mayor investigación. Así, las categorías y subcategorías proporcionadas (como F1.1, F1.2, F2.1, etc.) detallan la precisión de la evaluación técnica, permitiendo a las empresas priorizar proyectos con mayores probabilidades de éxito técnico.

Estructura del conjunto de datos del Eje F

En la clasificación de proyectos del Eje F (viabilidad técnica), la estructura de los datos se basa en diversas características que describen el estado de avance del proyecto desde un punto de vista técnico. Estas características están relacionadas con la fase en la que se encuentra el proyecto (si ha iniciado producción, si tiene planes de desarrollo aprobados, si existen obstáculos pendientes, etc.) y con su capacidad para cumplir con los requerimientos técnicos necesarios para proceder.

Los datos se estructuran de manera que cada proyecto se describe mediante una combinación de factores binarios o categóricos que indican si ciertos hitos técnicos se han alcanzado o no. Por ejemplo, se evalúan si las licencias y permisos necesarios están en regla, si hay actividades pendientes, si existe un plan aprobado, si se han presentado informes técnicos, entre otros.

Los árboles de decisión son la herramienta ideal para clasificar estos proyectos debido a su capacidad para manejar datos categóricos, su claridad en la toma de decisiones secuencial, y su interpretabilidad. Permiten evaluar la factibilidad técnica de un proyecto de manera lógica y transparente, asegurando que cada decisión esté respaldada por reglas claras. Además, su flexibilidad permite que el modelo se ajuste a medida que cambian las condiciones del proyecto, lo que es esencial para la gestión continua de proyectos de recursos.

Beneficios de la IA en la clasificación del Eje F

El modelo de árbol de decisión es particularmente adecuado para clasificar proyectos en el Eje F por varias razones importantes las cuales las dividiremos en generales y específicas.

Beneficios generales

1. Precisión y Consistencia Mejoradas: Al utilizar un modelo entrenado con datos históricos y criterios definidos, se reduce la variabilidad causada por interpretaciones humanas.
2. Ahorro de Tiempo y Recursos: La automatización permite procesar múltiples proyectos simultáneamente, liberando recursos humanos para tareas más estratégicas.
3. Análisis en Tiempo Real: La capacidad de obtener clasificaciones instantáneas facilita la toma de decisiones oportunas.
4. Reducción de Errores Humanos: Minimiza la posibilidad de errores derivados de fatiga o malinterpretación de los criterios.
5. Capacidad de Aprendizaje Continuo: El modelo puede ser actualizado y mejorado con nuevos datos, adaptándose a cambios en los criterios o en el entorno operativo.

6. Transparencia y Trazabilidad: Los algoritmos de decisión proporcionan una ruta clara de cómo se llegó a una clasificación, lo que es esencial para auditorías y cumplimiento normativo.

Beneficios específicos

Toma de decisiones secuencial

Un árbol de decisión es una representación visual y lógica de una secuencia de decisiones que se basan en las características del proyecto. Cada "nodo" del árbol hace una pregunta sobre una característica clave del proyecto, como, por ejemplo, "¿el plan de desarrollo ha sido aprobado?". Dependiendo de la respuesta ("sí" o "no"), el modelo sigue un camino específico hacia la siguiente decisión.
Esta estructura imita cómo un experto en la materia podría evaluar manualmente la viabilidad técnica de un proyecto, tomando en cuenta múltiples factores uno por uno. Así, el árbol puede combinar información dispersa sobre diferentes hitos del proyecto y llegar a una conclusión sobre su viabilidad técnica de forma lógica y estructurada.

Interpretabilidad

Una de las grandes ventajas de los árboles de decisión es su interpretabilidad. A diferencia de otros modelos de machine learning más complejos (como las redes neuronales), los árboles de decisión ofrecen reglas claras y comprensibles que explican por qué un proyecto fue clasificado de una manera particular.
Por ejemplo, si un proyecto es clasificado como viable técnicamente (F1), el árbol de decisión permite ver las decisiones que llevaron a esa conclusión. Esto es esencial cuando se necesita justificar la clasificación ante stakeholders o reguladores, ya que cada paso de la evaluación es completamente transparente y trazable.

Manejo de datos categóricos

La mayoría de las características que describen el estado técnico de un proyecto suelen ser categóricas o binarias, lo que significa que pueden tomar valores como "sí" o "no", o indicar diferentes estados como "aprobado", "pendiente", o "no aplicable". Los árboles de decisión son excepcionalmente buenos para manejar este tipo de datos discretos, ya que dividen los datos en subconjuntos lógicos según las respuestas a preguntas simples.

Por ejemplo, si una de las características del proyecto es "¿existe un plan de evaluación técnica aprobado?", el árbol puede bifurcarse dependiendo de si la respuesta es afirmativa o negativa. Con cada pregunta, el árbol va reduciendo las posibilidades hasta llegar a una clasificación final, que será F1 (viable), F2 (viable con condiciones), o F3 (no viable).

Simplicidad en la interpretación de múltiples factores

Un proyecto técnico involucra muchas variables: permisos, infraestructura, planes de desarrollo, migración de fases, entre otros. Un árbol de decisión puede procesar estas variables de forma conjunta, pero de manera organizada. El árbol toma todas las características relevantes, las evalúa una por una y proporciona un camino claro que lleva a la clasificación.

Por ejemplo, supongamos que hay un proyecto en el que faltan permisos, pero que ya tiene un plan aprobado y la infraestructura está en progreso. El árbol de decisión evaluará estos factores de forma independiente, pero luego los combinará para dar una clasificación completa del proyecto, señalando las áreas donde el proyecto está fuerte o donde existen debilidades que podrían retrasar su avance técnico.

Capacidad para manejar proyectos en diferentes fases

Los proyectos de recursos naturales suelen pasar por diferentes fases, desde la exploración inicial hasta la producción plena. A lo largo de estas fases, las

características del proyecto cambian: algunos obstáculos se superan y otros pueden surgir. Un árbol de decisión es flexible en el sentido de que puede actualizarse para incluir nuevos datos a medida que el proyecto avanza, y tomar en cuenta esta evolución.

Por ejemplo, si un proyecto inicialmente se clasificó como F2 (viable con condiciones pendientes) debido a la falta de ciertos permisos, y más tarde esos permisos son aprobados, el árbol de decisión puede actualizar la clasificación del proyecto a F1 (viable). Este tipo de flexibilidad es crucial en el contexto de proyectos de largo plazo, donde las condiciones cambian con el tiempo.

Rendimiento eficiente y simplicidad computacional

Los árboles de decisión son eficientes computacionalmente hablando, lo que los hace ideales para manejar múltiples proyectos simultáneamente. Pueden procesar un conjunto de datos relativamente grande sin requerir grandes recursos computacionales, lo que es útil cuando se está evaluando una cartera de proyectos en paralelo.

Funcionamiento del modelo de árboles de decisión

Los árboles de decisión son uno de los métodos más intuitivos y poderosos dentro del ámbito del aprendizaje automático. Es un algoritmo de aprendizaje supervisado no paramétrico, que se utiliza tanto para tareas de clasificación como de regresión (IBM, s.f.).

Estos modelos permiten realizar predicciones y clasificaciones a partir de un conjunto de datos mediante la descomposición recursiva de las características en subconjuntos, representando así una serie de decisiones jerárquicas que facilitan la comprensión y el análisis de los datos **(Figura 16**).

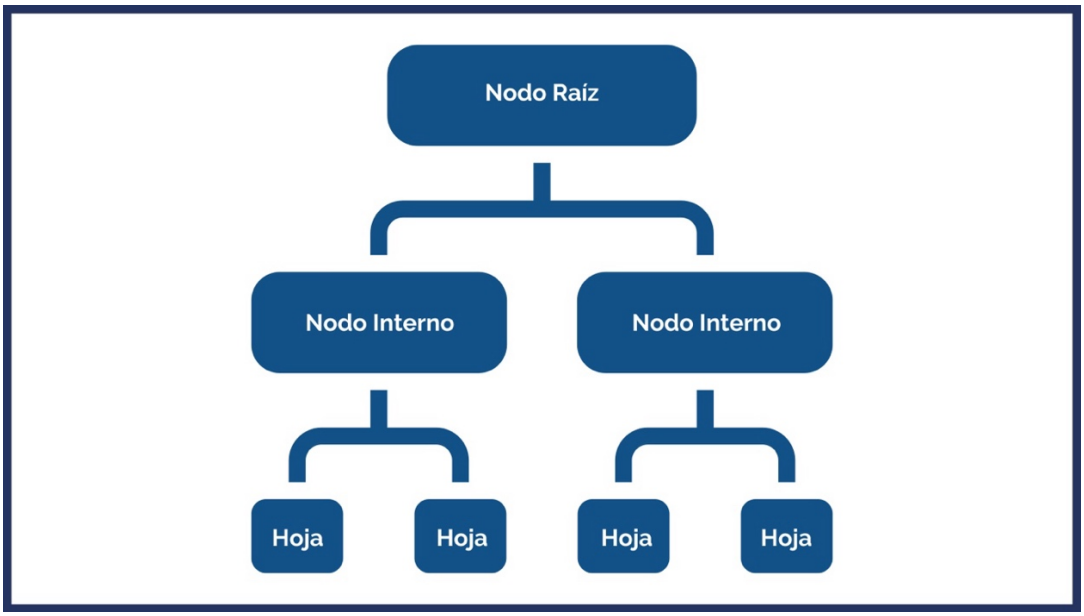

Figura 16. Estructura de un arbol de decisión en la IA.
Elaboración propia con datos de IBM

Cada nodo del árbol de decisión representa una característica o atributo del conjunto de datos, y las ramas representan posibles valores que esa característica puede tomar.

En términos matemáticos, los árboles de decisión se pueden describir como gráficos dirigidos y acíclicos (DAGs, por sus siglas en inglés) que representan un flujo lógico de decisiones. Un gráfico dirigido tiene nodos conectados por aristas que indican una dirección, lo que significa que se puede navegar de un nodo a otro siguiendo las conexiones. En un árbol de decisión, cada nodo representa un punto de decisión basado en una característica de los datos, y cada rama indica la elección resultante de esa decisión. Este tipo de gráfico es acíclico porque no existen ciclos, lo que garantiza que no se puede regresar a un nodo ya visitado.
La estructura jerárquica de un árbol de decisión permite descomponer un problema grande en subproblemas más pequeños y manejables.

Cada nodo divide el conjunto de datos en subconjuntos según una característica, y este proceso continúa hasta llegar a una hoja, que proporciona la predicción o clasificación final. Esta propiedad es clave para resolver problemas de aprendizaje supervisado, ya que el modelo se vuelve modular y más fácil de interpretar. (IBM, s.f.).

Justificación matemática de un árbol de decisión

Matemáticamente, los árboles de decisión son modelos eficaces que optimizan el proceso de toma de decisiones a través de la reducción de la incertidumbre (entropía o Gini). Este enfoque recursivo permite segmentar los datos de manera lógica y estructurada, lo que resulta en predicciones precisas basadas en las características de entrada. Además, las técnicas de poda ayudan a evitar el sobreajuste, mejorando la capacidad del modelo para generalizar en datos nuevos. En su forma más básica, un árbol de decisión estructura el proceso de toma de decisiones en una serie de preguntas basadas en características, que dividen repetidamente los datos hasta llegar a una clasificación o predicción.

Selección de características para la división

En cada nodo del árbol, es necesario elegir la característica que mejor divide el conjunto de datos. Este proceso de selección se basa en la reducción de la incertidumbre, y las dos métricas más utilizadas para medir dicha reducción son la entropía (utilizada en la ganancia de información) y el índice de Gini.

Entropía y ganancia de información

La entropía es una medida de la incertidumbre en un conjunto de datos y se define de la siguiente manera:

$$H(S) = -\sum_{i=1}^{c} pi\ log_2\ pi\ \ldots$$ **Ecuación 1**

Donde:

- S: es el conjunto de datos.
- pi:es la proporción de ejemplos que pertenecen a la clase iii.
- c : es el número de clases.

La ganancia de información evalúa cuánto reduce la entropía al dividir el conjunto de datos sobre una característica A. Se calcula como:

$$IG(S,A) = H(S) - \sum_{v \in valores(A)} \frac{|S_v|}{|S|} H(S_v) \ldots \textbf{Ecuación 2}$$

Donde:

- S_v: es el subconjunto de S donde la característica A toma el valor v.
- $\frac{|S_v|}{|S|}$: es la proporción de ejemplos en S_v con respecto al conjunto S.
- $\mathrm{H}(S_v)$: es la entropía del subconjunto S_v .

La característica que maximice la ganancia de información se selecciona como la mejor división en ese nodo del árbol.

Índice de Gini

El índice de Gini mide la pureza de un conjunto de datos y se define como:

$$\boldsymbol{Gini(S) = 1 - \sum_{i=1}^{c} p_1^2 \ldots \text{ Ecuación 3}}$$

Donde:

- p_i es la proporción de elementos de la clase i en el conjunto S.

La reducción del índice de Gini al dividir un conjunto de datos por una característica A se define como:

$$\Delta Gini(S,A) = Gini(S) - \sum_{v \in valores(A)} \frac{|S_v|}{|S|} Gini(Sv) \ldots \textbf{Ecuación 4}$$

El atributo que minimice el índice de Gini se selecciona para dividir el conjunto de datos en ese nodo del árbol.

Construcción recursiva del árbol

El proceso de división continúa de forma recursiva. Una vez seleccionada la característica que mejor divide el conjunto de datos, se generan subconjuntos, y el mismo proceso se aplica a cada uno de ellos. El árbol crece hasta que se cumple una de las siguientes condiciones:

- Todos los ejemplos en un subconjunto pertenecen a una misma clase (pureza).
- Se alcanza un número mínimo de ejemplos en los subconjuntos.
- Se alcanza una profundidad máxima del árbol.

Predicción con árboles de decisión

Una vez construido el árbol, se puede realizar la predicción para nuevas instancias. El proceso de predicción consiste en seguir las ramas del árbol en función de las características de la nueva instancia hasta llegar a una hoja, que corresponde a la clase o valor de salida.

Evitar el sobreajuste

Uno de los principales desafíos de los árboles de decisión es su tendencia al sobreajuste cuando se les permite crecer demasiado, ajustándose de manera excesiva a los detalles y ruidos específicos del conjunto de entrenamiento. Para evitar este problema y mejorar la capacidad de generalización, se emplea la técnica de poda del árbol, que consiste en eliminar nodos que no contribuyen significativamente a la precisión del modelo en nuevos datos.
Un método ampliamente utilizado es la poda por reducción de error, que evalúa si la eliminación de una rama mejora o mantiene la precisión del modelo en un conjunto de datos de validación. Si la poda no aumenta el error de clasificación y reduce la complejidad del árbol, la rama se elimina.

Proceso de Poda Recursiva

El procedimiento de poda se realiza de forma recursiva, comenzando desde las hojas del árbol y avanzando hacia la raíz. Las ramas se evalúan de la siguiente manera:

- Evaluación de cada nodo interno: Para cada nodo interno, se analiza la posibilidad de eliminar las sub-ramas bajo ese nodo.
- Criterio de poda: Si la eliminación de una sub-rama no aumenta el error en el conjunto de validación, se realiza la poda.
- Transformación del nodo: Cuando una sub-rama es eliminada, el nodo interno se convierte en una hoja que predice la clase más frecuente entre las instancias que llegan a ese nodo.

Comparación del Error Antes y Después de la Poda

El error de clasificación del árbol se compara antes y después de la poda. Si el error en el conjunto de validación disminuye o se mantiene constante tras la poda, se considera que la poda es beneficiosa, ya que ha simplificado la estructura del árbol sin comprometer su precisión. Esto ayuda a reducir la complejidad del modelo y mejorar su capacidad de generalización.

Iteración

Este proceso de poda se repite para cada nodo interno del árbol, de manera recursiva, hasta que no se observen más mejoras en la precisión del modelo o hasta que se alcance una versión simplificada del árbol que generalice mejor en los datos de validación.

Automatización de la Clasificación del Eje F del UNFC mediante árboles de decisión

La gestión eficiente de proyectos de recursos naturales es esencial para el desarrollo sostenible y la toma de decisiones estratégicas en el sector energético y minero. El Marco de Clasificación de las Naciones Unidas para Recursos (UNFC) proporciona una estructura integral para evaluar proyectos en función de su viabilidad económica, social y técnica (UNECE, 2013). Dentro de este marco, el Eje F se centra en la viabilidad del proyecto, considerando factores técnicos y de madurez del proyecto.

Objetivos de la automatización

- Estandarización: Aplicar criterios uniformes en la clasificación de proyectos.
- Eficiencia Operativa: Reducir el tiempo de evaluación y acelerar la toma de decisiones.
- Escalabilidad: Manejar eficientemente un mayor número de proyectos sin comprometer la calidad de la evaluación.
- Integración Tecnológica: Incorporar herramientas de inteligencia artificial para mejorar los procesos existentes.
- Transparencia: Proporcionar un modelo interpretable que permita entender y justificar las decisiones tomadas.

Implementación del modelo en el Eje F

Preprocesamiento y preparación de datos

Se generó un archivo Excel con información de 200 proyectos (**Anexo 1**), estructurado en 28 categorías que describen los aspectos técnicos clave de cada proyecto. Estas categorías fueron codificadas en formato binario: '1' indica que la característica está presente, y '0' que está ausente. Este archivo sirvió como base

para entrenar un modelo de árbol de decisión, en el cual cada una de las 28 categorías se consideró como un nodo de decisión. El modelo evaluó estas categorías de forma secuencial para clasificar los proyectos según su viabilidad técnica, de acuerdo con el Eje F del UNFC. Los proyectos fueron categorizados como:

- F1: Técnicamente viables.
- F2: Viables con condiciones pendientes.
- F3: No viables.

Antes de entrenar el modelo, se realizó un preprocesamiento de los datos para garantizar su calidad y consistencia. Este proceso incluyó:

- Eliminación de datos faltantes: Se eliminaron o imputaron registros con valores ausentes.
- Codificación de variables categóricas: Las variables categóricas se transformaron en formato binario.
- Normalización: No fue necesaria, ya que los árboles de decisión no dependen de la escala de las variables.

Selección de hiperparámetros

Se ajustaron los hiperparámetros del modelo para optimizar su rendimiento:

- Profundidad Máxima (max_depth): Limitada para prevenir el sobreajuste.
- Criterio de Impureza (criterion): Se utilizó entropy para basar las divisiones en la ganancia de información.
- Mínimo de Muestras por Hoja (min_samples_leaf): Establecido para asegurar que las hojas tengan un número suficiente de muestras.

Código de Entrenamiento

El modelo de clasificación implementado es un Random Forest Classifier, un enfoque que combina múltiples árboles de decisión para generar predicciones robustas y minimizar el riesgo de sobreajuste. Para garantizar la reproducibilidad de los resultados, se configura con un estado aleatorio fijo (random_state=42). El modelo es entrenado utilizando los datos del conjunto de entrenamiento (X_train, y_train).

Tras el entrenamiento, el modelo es evaluado utilizando el conjunto de prueba (X_test) para generar predicciones (y_pred). Estas predicciones permiten analizar el rendimiento del modelo en datos no vistos previamente.

Se calcula la precisión (accuracy) empleando la función accuracy_score de la biblioteca Scikit-learn. La precisión mide la proporción de instancias correctamente clasificadas con respecto al total, proporcionando una visión clara y cuantitativa de la efectividad del modelo.
Este proceso permite evaluar de manera sistemática el rendimiento del modelo, asegurando una interpretación clara de sus capacidades predictivas y áreas de mejora potencial.
El código completo de este proceso se incluye en el **Anexo 1** para su consulta y referencia.

Matriz de confusión

La matriz de confusión constituye una herramienta fundamental para evaluar y comprender el desempeño de un modelo de clasificación. Complementada con la métrica de precisión, proporciona una visión holística de la efectividad del modelo y destaca áreas potenciales para su mejora.

En la **Figura 17** se ilustra la matriz de confusión utilizada para analizar el rendimiento del modelo en una tarea de clasificación multiclase. En esta representación, el eje vertical corresponde a las etiquetas reales, reflejando los valores verdaderos del conjunto de datos de prueba, mientras que el eje horizontal representa las etiquetas predichas por el modelo para las mismas instancias.

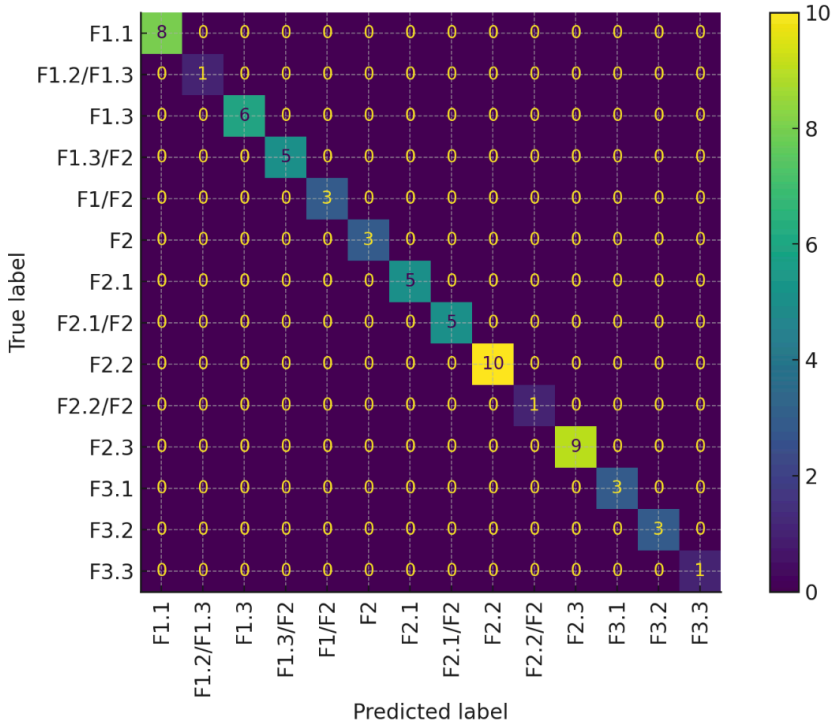

Figura 17. Matriz de Confusión del modelo de entrenamiento del eje F
Elaboración propia

Este tipo de análisis permite discernir de manera clara los aciertos y errores del modelo en cada clase específica. Al identificar patrones de confusión, como clases que son frecuentemente equivocadas entre sí, se obtiene información valiosa que puede orientar ajustes y optimizaciones en el modelo. Además, facilita la identificación de posibles sesgos o desequilibrios en los datos, lo que es esencial para mejorar la precisión y la generalización del modelo en aplicaciones futuras.

Observaciones

Diagonal principal:

- La diagonal principal de la matriz contiene los valores correspondientes a las instancias correctamente clasificadas para cada clase.
- En este caso, todas las instancias del conjunto de prueba se encuentran en esta diagonal, indicando que el modelo clasificó correctamente cada instancia en su respectiva categoría.

Celdas fuera de la diagonal:

- Estas celdas representan errores de clasificación, donde el modelo predijo incorrectamente la clase de una instancia. En esta matriz, estas celdas están vacías, lo cual demuestra la ausencia de errores de clasificación.

Rendimiento del Modelo

El modelo alcanzó una precisión del 100%, lo que indica que todas las predicciones realizadas coincidieron perfectamente con las etiquetas reales. Este resultado refleja un desempeño impecable en la tarea de clasificación evaluada.
Además, las celdas situadas fuera de la diagonal principal, que representan los errores de clasificación, se encuentran completamente vacías. Esto confirma que no se registraron instancias mal clasificadas, lo que refuerza la exactitud y fiabilidad del modelo. Este nivel de rendimiento demuestra que el modelo es altamente eficaz para distinguir entre las diferentes clases presentes en el conjunto de prueba, evidenciando su capacidad para generalizar correctamente a partir de los datos proporcionados.

Código de Predicción

El código (disponible en el **Anexo 1**) implementa un flujo estructurado para realizar predicciones de clasificación utilizando un modelo preentrenado guardado en modelo_arbol_decision.pkl. A continuación, se resumen sus principales componentes:

- Asegura la compatibilidad y reutilización del modelo.
- Incluye la función para validar entradas binarias (0 o 1), mientras que solicita y organiza los datos del usuario en un formato procesable por el modelo.
- Clasifica las entradas y emplea un diccionario de explicaciones que traduce las predicciones (F1, F2, F3) a descripciones claras y accionables para el usuario.
- Previene fallos críticos por entradas no válidas o problemas de compatibilidad, mejorando la robustez del código.

Ventajas:

- Recolecta datos directamente del usuario.
- Valida entradas y maneja errores eficazmente.
- Simplifica el proceso de predicción.
- Ofrece explicaciones claras y accesibles.

Limitaciones:

- Aún no es escalable para grandes volúmenes de datos.
- El código es práctico y funcional para realizar predicciones con un modelo preentrenado, pero puede mejorarse al incorporar soporte para entradas masivas, validaciones avanzadas y explicaciones más detalladas.

Eje E (Viabilidad Ambiental, Social y Económica)

Para la evaluación del eje E, se elaboró una matriz que permite identificar los principales riesgos sociales, ambientales y legales, junto con las consideraciones económicas clave de cada proyecto. El eje E abarca la evaluación económica del proyecto, integrando además los factores ambientales, sociales y regulatorios a lo largo de todo su ciclo de vida, buscando lograr un equilibrio. Los subproductos de la extracción de petróleo y gas juegan un papel importante en el cumplimiento de varios Objetivos de Desarrollo Sostenible (ODS).
Las categorías del eje E abordan factores no técnicos que afectan directamente la viabilidad comercial del proyecto, como los precios de los productos, costos de producción, condiciones del mercado, y los marcos legales y fiscales, así como las regulaciones ambientales y sociales.

La Comisión Nacional de Hidrocarburos (CNH) desarrolló una matriz para evaluar la viabilidad socioeconómica del eje E como parte de la prueba piloto de 2019. Esta matriz tiene como propósito describir con mayor precisión los riesgos específicos de los proyectos y evaluar su viabilidad basándose en los siguientes aspectos:

- Los factores socio-organizativos incluyen aspectos relacionados con la estructura social y económica del área, tales como la presencia de comunidades indígenas, uso del suelo, niveles de marginación, desarrollo humano, actividades económicas y uso del agua.
- Los factores ambientales abarcan el entorno físico, incluyendo zonas de protección, áreas naturales protegidas, humedales, especies protegidas y áreas de uso crítico del suelo.
- También se consideran las condiciones legales y regulatorias.

Matriz del Eje E

La viabilidad de los proyectos en el marco del eje E se clasifica en tres niveles: alta, media o baja, dependiendo de las variables ambientales, sociales, legales y económicas (**Tabla 5**). Los expertos con experiencia en la zona del proyecto pueden utilizar la siguiente matriz como una herramienta cualitativa para identificar los principales riesgos en estas áreas y evaluar la factibilidad de llevar a cabo el proyecto (CNH, 2019).

Variables Ambientales	Alto (Más probable)	Medio (Probable)	Bajo (No probable)
¿El proyecto está localizado en un área restringida?	No	Quizás	Sí
¿Flora y fauna enlistadas en la NOM-059-SEMARNAT-2010?	No	Quizás	Sí
¿Hay Programa de Ordenamiento Ecológico Territorial?	No	Quizás	Sí
¿Hay uso crítico de suelo?	No	Quizás	Sí
Variables Socio - Organizacionales	**Alto (Más probable)**	**Medio (Probable)**	**Bajo (No probable)**
¿Presencia de comunidades indígenas? (Comunidades > 50 personas)	No	Quizás	Sí
¿Hay alguna región indígena?	No	Quizás	Sí
¿Hay pertenencia social y de la tierra?	No	Quizás	Sí
¿Hay marginalización? Según el índice de marginalización**.**	No	Quizás	Sí
¿El proyecto interfiere con alguna actividad económica?	No	Quizás	Sí
¿Hay afectación al proyecto por el agua?	No	Quizás	Sí
Variables legales	**Alto (Más probable)**	**Medio (Probable)**	**Bajo (No probable)**
¿Hay alguna afectación por las variables legales en el proyecto?	No	Quizás	Sí
¿Hay permisos y aprobaciones ambientales?	No	Quizás	Sí
¿Hay evaluaciones sociales?	No	Quizás	Sí

Tabla 5. Matriz de análisis de variables para la evaluación del eje E

Elaboración propia con datos de CNH

Las subcategorías de este eje (**Tabla 6**) son fundamentales para entender si un proyecto puede ser implementado de manera rentable y sostenible ya que, reflejan el nivel de certeza sobre la posibilidad de que un proyecto sea económicamente

viable y sustentable, teniendo en cuenta factores como el mercado, la legislación, los impactos sociales y ambientales.

En estas se pueden encontrar adiciones que permiten mayor detalle en la evaluación del proyecto, por ejemplo:

Categoría	Subcategoría	Definición
E1	E1.1	El desarrollo es viable en base a las condiciones actuales y suposiciones realistas de las condiciones futuras.
	E1.2	El desarrollo no es viable basado en las condiciones actuales del proyecto, pero se hace viable a través de subsidios gubernamentales y otras consideraciones.
E2	E2.1	No todas las contingencias económicas, sociales y ambientales han sido resueltas, pero existe una alta probabilidad de que se resuelvan en un futuro previsible. Los problemas aún no se han resuelto, pero existe una alta probabilidad de su resolución, evidenciado por un intento activo de resolver todos los impedimentos (contingencias) con una alta probabilidad de éxito, y por un historial de proyectos similares en el área.
	E2.2	No todas las cuestiones económicas, sociales y ambientales han sido resueltas, y no hay actividad para resolver los impedimentos. Sin embargo, existe una probabilidad media de su resolución en un futuro previsible.
	E2.3	No todas las cuestiones económicas, sociales y ambientales han sido resueltas y los ámbitos no resueltos no pueden ser influenciados por las partes interesadas. Se espera que se resuelvan en un futuro previsible.
E3	E3.1	Existe estimación de proyectos que se prevé que se desarrollen a futuro, pero que no se utilizarán o consumara operaciones.
	E3.2	La viabilidad no se puede determinar debido a la información insuficiente.
	E3.3	Sobre la base de suposiciones realistas sobre las condiciones futuras, actualmente no se considera que no haya perspectivas razonables en el futuro previsible.

Tabla 6. Matriz de análisis y determinación de subcategorías para la evaluación del eje E
Elaboración propia con datos del ICE-SRM de la ONU

Beneficios de la IA en la clasificación del eje e (viabilidad ambiental, social y económica)

El uso de técnicas de redes neuronales permitirá una clasificación precisa y adaptable en el eje E, con beneficios como:

- Evaluación más rápida y precisa: Al utilizar redes neuronales, el sistema de IA puede procesar grandes volúmenes de datos ambientales, sociales y económicos de manera rápida, proporcionando evaluaciones detalladas de la viabilidad de los proyectos bajo diferentes condiciones.

- Monitoreo continuo: Las redes neuronales pueden adaptarse y aprender de los cambios en los datos en tiempo real, lo que permite un monitoreo constante y actualizaciones automáticas en las evaluaciones de viabilidad a medida que cambian las condiciones del mercado, sociales o regulatorias.
- Reducción del sesgo: Las redes neuronales, al basarse en datos objetivos y análisis de patrones complejos, ayudan a reducir el sesgo humano en la evaluación de proyectos, asegurando decisiones más justas y fundamentadas.
- Proyección futura: Las redes neuronales pueden simular diferentes escenarios futuros basándose en variables ambientales, sociales y económicas, lo que ayuda a prever el impacto a largo plazo de cada proyecto, facilitando una planificación más sólida

Funcionamiento del modelo de redes neuronales

Las redes neuronales son una de las técnicas más avanzadas y versátiles dentro del aprendizaje automático. Estas están inspiradas en el funcionamiento del cerebro humano y consisten en capas de nodos (neuronas) interconectadas que permiten procesar datos complejos. Se utilizan principalmente para resolver problemas de clasificación, regresión y predicción (IBM, s.f.).

En el contexto de la clasificación, las redes neuronales permiten realizar predicciones a partir de un conjunto de datos mediante el ajuste iterativo de pesos asignados a las conexiones entre neuronas, lo que facilita la identificación de patrones ocultos y complejos en los datos.

Estructura de una red neuronal

Una red neuronal típica está compuesta por tres tipos de capas:

- Capa de entrada: Recibe los datos de entrada (características del conjunto de datos). Cada nodo en esta capa representa una característica del conjunto de datos.
- Capas ocultas: Estas son capas intermedias que transforman los datos mediante funciones de activación no lineales. Aquí es donde se realiza el aprendizaje profundo, ya que cada nodo en una capa oculta toma una combinación ponderada de los nodos de la capa anterior y aplica una función de activación para generar su salida.
- Capa de salida: Proporciona el resultado final de la clasificación o predicción. En tareas de clasificación, cada nodo en la capa de salida representa una clase, y el nodo con el valor más alto será la clase predicha por la red.

Justificación matemática de una red neuronal

Matemáticamente, una red neuronal ajusta los pesos $\boldsymbol{w}$ y los sesgos $\boldsymbol{b}$ de cada conexión entre neuronas utilizando una función de pérdida y un algoritmo de optimización (como el descenso de gradiente). La función de pérdida mide qué tan lejos está la predicción de la red de los valores reales, mientras que el descenso de gradiente ajusta los pesos para minimizar esta pérdida. El objetivo es encontrar los pesos óptimos que minimicen la diferencia entre la salida predicha y la real.

Una sola neurona en una red neuronal toma una combinación lineal de las entradas, les aplica una función de activación, y genera una salida:

$$z = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b = \sum_{i=1}^{n} w_ix_i + b\ldots$$ **Ecuación 5**

Donde:

w_i son los pesos de las conexiones

x_i son las entradas (características del conjunto de datos)

b es el sesgo

z es el valor de la combinación lineal de las entradas

La salida de la neurona es $a = f(z)$ donde f es una función de activación

Función de activación

La función de activación introduce no linealidad al modelo, lo que permite que la red neuronal aprenda patrones complejos. Algunas de las funciones de activación más comunes son:

ReLU (Rectified Linear Unit):

$$f(z) = \max(0, z)\ldots$$ **Ecuación 6**

Sigmoide:

$$f(z) = \frac{1}{1+e^{-z}}\ldots$$ **Ecuación 7**

Tangente hiperbólica (tanh):

$$f(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} = \tanh(z)\ldots$$ **Ecuación 8**

Softmax (usada en la capa de salida para la clasificación multiclase):

$$f(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \ldots \textbf{Ecuación 9}$$

Donde C es el número de clases.

Cálculo del error (función de pérdida)

El objetivo de la red neuronal es minimizar el error entre la salida predicha y el valor real de las etiquetas. Esto se mide a través de una función de pérdida, que evalúa qué tan lejos están las predicciones del resultado esperado. Algunas funciones de pérdida comunes son:

Error cuadrático medio (MSE) para problemas de regresión:

$$L(, y, \hat{y}) = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \ldots \textbf{Ecuación 10}$$

Donde y_i es el valor real y $\hat{y}_i$ es la predicción.

Entropía cruzada para clasificación binaria:

$$L(, y, \hat{y}) = -(ylog(\hat{y}) + (1 - y)\log(1 - \hat{y})) \ldots \textbf{Ecuación 11}$$

Entropía cruzada categórica para clasificación multiclase:

$$L(, y, \hat{y}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \ \ldots \textbf{Ecuación 12}$$

Retro propagación (cálculo de gradientes)

El entrenamiento de una red neuronal implica la optimización de los pesos y sesgos para minimizar la función de pérdida. Esto se logra utilizando el descenso de gradiente y su variante, retro propagación, que ajusta los pesos en función del gradiente de la pérdida con respecto a cada peso.

Para un peso w, el ajuste se realiza en la dirección opuesta al gradiente:

$$w \leftarrow w - n\frac{\partial L}{\partial w} \ldots \textbf{Ecuación 13}$$

Donde:

n es la tasa de aprendizaje

$\frac{\partial L}{\partial w}$ es el gradiente de la función de pérdida con respecto al peso w.

Regla de la cadena para retro propagación

Para calcular los gradientes en cada capa, se aplica la regla de la cadena:

$$\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial a} \cdot \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_i} \ldots \textbf{Ecuación 14}$$

Donde:

$\frac{\partial L}{\partial a}$ es el gradiente de la pérdida con respecto a la salida de la neurona

$\frac{\partial a}{\partial z}$ es el gradiente de la salida de la neurona con respecto a su entrada (derivada de la función de activación)

$\frac{\partial z}{\partial w_i} = x_i$ que es el valor de la entrada en ese peso.

El proceso de retropropagación ajusta los pesos de cada neurona en función del error que propaga hacia atrás desde la capa de salida hasta las capas ocultas.

Optimización del modelo (descenso de gradiente estocástico)

El descenso de gradiente estocástico (SGD) es el algoritmo más común para optimizar los pesos. En lugar de calcular el gradiente usando todo el conjunto de datos (descenso de gradiente completo), el SGD actualiza los pesos después de ver un solo ejemplo o un pequeño subconjunto.

La actualización de los pesos se realiza como sigue:

$$w \leftarrow w - n\frac{1}{m}\sum_{i=1}^{m}\frac{\partial L(y_i,\hat{y}_i)}{\partial w} \ldots \textbf{Ecuación 15}$$

Donde:
m es el número de ejemplos de subconjunto
n es la taza de aprendizaje

Regularización para evitar sobreajuste

Para evitar el sobreajuste, se utilizan varias técnicas de regularización que penalizan pesos grandes en la función de pérdida:

Regularización L2 (Ridge)

$$L_{total} = L(y,\hat{y}) + \lambda\sum_{i=1}^{n} w_i^2 \ldots \textbf{Ecuación 16}$$

Donde λ es el factor de regularización que controla la penalización.

Regularización L1 (Lasso):

$$L_{total} = L(y,\hat{y}) + \lambda\sum_{i=1}^{n} |w_i| \ldots \textbf{Ecuación 17}$$

Dropout

El dropout es una técnica de regularización que desconecta aleatoriamente un conjunto de neuronas durante cada paso de entrenamiento para evitar que el modelo dependa demasiado de ciertas neuronas. Matemáticamente, esto se puede describir como:

$$a_{drop}^{[l]} = a^{[l]} \cdot D^{[l]} \ldots$$ **Ecuación 18**

Donde:

$D^{[l]}$ es una máscara binaria que indica si una neurona en la capa lll está activa o no

$a_{drop}^{[l]}$ es la activación de la capa lll después de aplicar el dropout.

El desarrollo matemático abarca desde el funcionamiento esencial de una neurona, las funciones de activación, hasta los procesos clave de retro propagación, optimización y regularización. Las redes neuronales, aunque complejas, son capaces de modelar relaciones no lineales en los datos, basándose en principios sólidos de cálculo y optimización para su entrenamiento.

Proceso de Entrenamiento de Redes Neuronales

El entrenamiento de una red neuronal consiste en exponer la red a datos de entrenamiento, calcular el error entre las predicciones y los valores reales, y ajustar los pesos de sus conexiones para reducir dicho error. Este ajuste se realiza mediante el algoritmo de retropropagación, que distribuye el error desde la capa de salida hacia las capas anteriores, modificando los pesos de acuerdo con el gradiente de la función de pérdida respecto a cada peso.

Retropropagación y Optimización

La retropropagación, en combinación con el descenso de gradiente estocástico (SGD), es clave para minimizar la función de pérdida. Este método ajusta iterativamente los pesos y los sesgos de la red, permitiendo que esta aprenda patrones en los datos y mejore sus predicciones con cada iteración.
El proceso de entrenamiento sigue las siguientes fases principales:

- **Cálculo de la Pérdida**
  Se determina la diferencia entre la salida predicha por la red y el valor real utilizando una función de pérdida específica, como el error cuadrático medio para problemas de regresión o la entropía cruzada para problemas de clasificación.
- **Retropropagación del Error**
  El error calculado en la capa de salida se propaga hacia atrás a través de la red. Durante este proceso, se obtienen los gradientes de la función de pérdida con respecto a cada peso y sesgo, los cuales son utilizados para ajustar las conexiones.
- **Actualización de Pesos**
  Los pesos y sesgos se actualizan utilizando el descenso de gradiente. Este método aplica pequeños ajustes en la dirección que reduce el valor de la función de pérdida, mejorando así la precisión del modelo.

Cada uno de estos pasos se repite iterativamente a lo largo de varias épocas, lo que permite a la red neuronal refinar su capacidad de modelar los datos y generalizar a nuevos conjuntos de datos.

Funciones de Activación

Las funciones de activación introducen no linealidad en las redes neuronales, permitiéndoles modelar patrones complejos en los datos. Algunas de las funciones de activación más utilizadas son:

- **ReLU (Rectified Linear Unit):**
  Activa las neuronas solo cuando la entrada es mayor que 0. Es ampliamente empleada en capas ocultas debido a su eficiencia computacional y su capacidad para procesar grandes volúmenes de datos sin problemas de saturación.
- **Sigmoide:**
  Convierte los valores de entrada en un rango entre 0 y 1, siendo útil para problemas de clasificación binaria. Sin embargo, puede sufrir problemas de gradientes pequeños en redes profundas.
- **Softmax:**
  Una generalización de la función sigmoide para problemas de clasificación multiclase. Asigna probabilidades a cada clase, de forma que la suma de las probabilidades es igual a 1.

Técnicas para Evitar el Sobreajuste

El sobreajuste es un desafío común en redes neuronales. Ocurre cuando el modelo se adapta excesivamente al conjunto de entrenamiento, reduciendo su capacidad para generalizar en datos nuevos. Algunas estrategias para prevenirlo son:

- **Regularización:**
  Métodos como la regularización L2 (ridge) penalizan los pesos grandes en el modelo, promoviendo soluciones más simples y menos propensas al sobreajuste.
- **Dropout:**
  Durante el entrenamiento, se desactivan aleatoriamente un porcentaje de neuronas en cada iteración. Esto fuerza a la red a aprender representaciones más generales y menos dependientes de neuronas específicas.
- **Detección temprana (Early Stopping):**

Monitorea el rendimiento del modelo en un conjunto de validación y detiene el entrenamiento cuando este rendimiento comienza a empeorar, evitando que el modelo se ajuste en exceso a los datos de entrenamiento.

Predicción con Redes Neuronales

Después de entrenar la red neuronal, esta puede utilizarse para realizar predicciones con nuevos datos. El proceso es el siguiente:

- Las nuevas entradas se introducen a la red a través de la capa de entrada.
- Estas entradas pasan a través de las capas ocultas, donde se aplican los pesos, sesgos y funciones de activación aprendidos durante el entrenamiento.
- Finalmente, la capa de salida genera las predicciones. En problemas de clasificación, el nodo con el valor más alto en la capa de salida indica la clase predicha.

Este flujo garantiza que la red procese los datos de manera eficiente y proporcione resultados basados en el conocimiento adquirido durante su entrenamiento.

Automatización de la Clasificación del Eje E del UNFC Mediante Redes Neuronales

La automatización de la clasificación de proyectos en el contexto del Marco de Clasificación de las Naciones Unidas para Recursos (UNFC) es un paso crucial para mejorar la eficiencia y precisión en la toma de decisiones dentro de la industria de recursos naturales. El UNFC, establecido por la Comisión Económica para Europa de las Naciones Unidas (UNECE), es una herramienta integral para evaluar la viabilidad económica, social y ambiental de los proyectos. El Eje E del UNFC se enfoca en evaluar la viabilidad económica y social, tomando en cuenta factores como la aceptación social, las restricciones ambientales y los aspectos legales de los proyectos (UNECE, 2013).

Objetivos de la automatización

- **Estandarización**: Aplicar criterios uniformes en la evaluación de la viabilidad económica, social y legal de los proyectos. Al estandarizar el proceso mediante el uso de redes neuronales, se elimina la variabilidad inherente a las evaluaciones manuales, proporcionando resultados más consistentes y alineados con las directrices del UNFC.
- **Eficiencia Operativa**: Reducir el tiempo necesario para evaluar proyectos en el Eje E, permitiendo tomar decisiones informadas de manera más rápida. Al automatizar el proceso, se agiliza la evaluación de múltiples proyectos simultáneamente.
- **Escalabilidad**: La red neuronal permite manejar grandes volúmenes de datos sin comprometer la precisión de los resultados. Esto facilita la evaluación de un mayor número de proyectos a medida que se incorporan más recursos al sistema.
- **Integración Tecnológica**: La incorporación de redes neuronales no solo permite mejorar el proceso de evaluación existente, sino también integrar los resultados de manera fluida en plataformas digitales de gestión de proyectos, proporcionando una visión global de los riesgos y oportunidades de cada proyecto.
- **Transparencia**: La red neuronal puede ser interpretada a través de visualizaciones y métricas que expliquen cómo se toman las decisiones. Esto es crucial para garantizar que las evaluaciones sean comprensibles tanto para expertos técnicos como para partes interesadas que requieren justificaciones claras de las decisiones.

Implementación del Modelo en el Eje E

El proceso de automatización de la clasificación en el Eje E se basa en la recopilación de datos sociales, ambientales, legales y económicos relevantes para cada proyecto. Estos datos son transformados en un conjunto de características

que alimentan una red neuronal entrenada para clasificar proyectos en las subcategorías del Eje E (por ejemplo, E1.1, E1.2, E2.1, E3.1, etc.).

El entrenamiento de la red neuronal se llevó a cabo utilizando un set de datos de proyectos con características definidas como:

- **Variables ambientales**: Incluyendo la ubicación en áreas restringidas o la afectación por flora y fauna.
- **Variables sociales**: Como la presencia de comunidades indígenas y la interferencia con actividades económicas locales.
- **Variables legales**: Relacionadas con la obtención de permisos y evaluaciones sociales.

Los datos fueron preprocesados y escalados antes de ser utilizados para entrenar la red neuronal, que fue diseñada con varias capas ocultas para capturar la complejidad de las interacciones entre las variables de entrada. Una vez entrenada, la red neuronal es capaz de predecir de manera eficiente la clasificación del Eje E de un proyecto nuevo.

Preprocesamiento y Preparación de Datos

El archivo Proyectos_EjeE.xlsx (mostrado en el **Anexo 1**) contiene información estructurada de proyectos ficticios para entrenar la red neuronal. Está compuesto por varias columnas, donde cada una representa una característica o variable relevante para la evaluación de la viabilidad económica, social, ambiental y legal del proyecto, de acuerdo con las exigencias del Eje E del UNFC.
Antes de entrenar el modelo de redes neuronales para la clasificación del Eje E, se realizaron varios pasos de preprocesamiento de datos para garantizar que el modelo pudiera aprender de manera eficiente y producir predicciones precisas.

- **Eliminación de Datos Faltantes**: Se revisaron los datos para identificar registros incompletos. En los casos en que había valores faltantes, se aplicaron técnicas de imputación, como rellenar los valores con la mediana de las características o eliminar las filas con datos faltantes, dependiendo de la importancia de la variable en cuestión.

- **Codificación de Variables Categóricas**: Dado que las redes neuronales requieren datos numéricos, se transformaron las variables categóricas (como las opciones "No", "Quizás", "Sí") en valores numéricos. Para esto, se utilizó una codificación ordinal, donde:
    - "No" se convirtió en 0
    - "Quizás" se convirtió en 1
    - "Sí" se convirtió en 2

- **Normalización de los Datos**: A diferencia de los árboles de decisión, las redes neuronales pueden verse afectadas por diferentes escalas en los valores de entrada. Por lo tanto, se aplicó un proceso de **normalización** para escalar todas las variables de entrada dentro de un rango estándar, típicamente utilizando un **escalado estándar** (media 0, desviación estándar Esto garantizó que las características con valores más grandes no dominen las más pequeñas, lo que mejoró el rendimiento del modelo.

Cada fila del archivo Excel corresponde a un proyecto, con las características relevantes (variables) en las columnas, y la clasificación del Eje E en la columna final.

Selección de Hiperparámetros

El modelo de redes neuronales utilizado para clasificar el Eje E fue ajustado con diversos hiperparámetros para asegurar un equilibrio adecuado entre complejidad del modelo y rendimiento:

- Número de Neuronas y Capas Ocultas: Se experimentó con diferentes números de neuronas y capas ocultas. La arquitectura final incluyó dos capas ocultas: la primera con 64 neuronas y la segunda con 32 neuronas. Estas capas permitieron capturar las interacciones complejas entre las variables económicas, sociales, ambientales y legales.
- Función de Activación: Se utilizó la función de activación ReLU (Rectified Linear Unit) en las capas ocultas, ya que es una de las funciones más efectivas para modelos de redes neuronales profundas, permitiendo aprender relaciones no lineales complejas. En la capa de salida, se utilizó la función softmax, apropiada para tareas de clasificación multiclase como la del Eje E.
- Optimizador: Se utilizó el optimizador Adam (Adaptative Moment Estimation) debido a su efectividad en tareas de aprendizaje profundo, ajustando dinámicamente el learning rate durante el entrenamiento.
- Número de Épocas: El modelo fue entrenado durante 50 épocas, con un tamaño de lote de 8, lo que permitió actualizar los pesos del modelo de manera eficiente mientras se prevenía el sobreajuste.

Código de Entrenamiento

El código que se utilizó para entrenar la red neuronal (mostrado en el **Anexo 1**) está basado en Keras (TensorFlow), un framework popular para el desarrollo de modelos de redes neuronales.

El modelo fue entrenado utilizando el conjunto de datos preprocesado y escalado, compuesto por 13 características de entrada, incluyendo variables ambientales, sociales y legales. El objetivo era predecir la clasificación del proyecto en una de las subcategorías del Eje E (E1.1, E1.2, E2.1, etc.)

Para la clasificación del Eje E del UNFC, se utilizó una red neuronal perceptrón multicapa (MLP). Este tipo de red es ampliamente utilizado en problemas de clasificación cuando las relaciones entre las variables de entrada no son lineales,

como en el caso de las múltiples interacciones entre las variables sociales, ambientales y legales en los proyectos del Eje E.

Se utilizaron dos capas ocultas en el modelo, que permiten que capture interacciones complejas entre las variables de entrada. La primera capa contiene 64 neuronas y la segunda 32 neuronas, lo que proporciona suficiente capacidad de aprendizaje sin sobrecargar el modelo.

Para activarlas se utilizó ReLU (Rectified Linear Unit) en las capas ocultas, ya que es eficiente para evitar problemas de gradiente y mejora la convergencia durante el entrenamiento y en las capas de salida emplee softmax, ya que el problema a resolver es una clasificación multiclase (con 8 categorías posibles, correspondientes a las subcategorías del Eje E). Softmax convierte los valores de salida en probabilidades, permitiendo identificar la clase más probable.

Matriz de Confusión

En este caso, la matriz asociada al modelo refleja un nivel de precisión del 80%, lo que indica que el modelo clasifica correctamente el 80% de las muestras analizadas, dejando un 20% de margen de error distribuido entre las diferentes clases.

La **Figura 18** ilustra cómo las predicciones del modelo se comparan con las etiquetas reales.

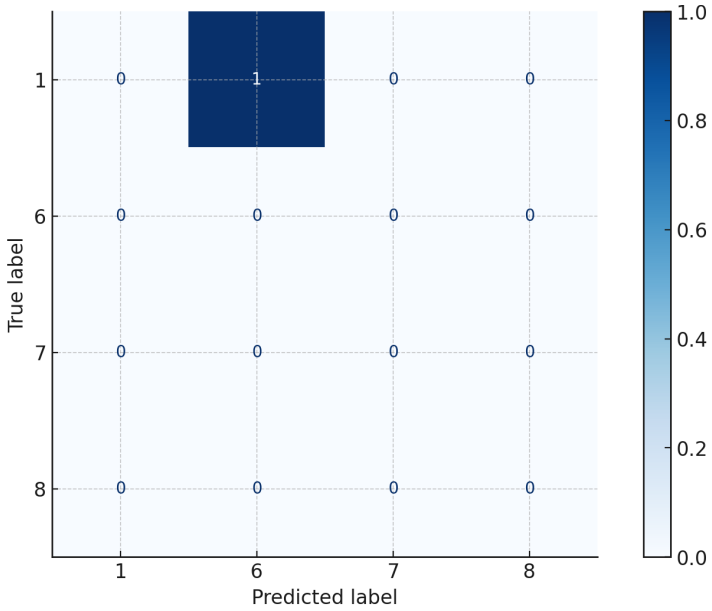

Figura 18. Matriz de Confusión del modelo de entrenamiento del eje E
Elaboración propia

Los valores en la diagonal principal representan los aciertos del modelo, donde las predicciones coinciden con las clases reales. Por otro lado, los valores fuera de la diagonal muestran las confusiones del modelo, es decir, las ocasiones en las que una clase fue asignada erróneamente a otra.

Rendimiento del Modelo

El modelo en cuestión alcanzó una precisión del 80%, posicionándose como una herramienta eficaz para la tarea de clasificación planteada. La arquitectura del modelo está compuesta por dos capas densas ocultas con 64 y 32 neuronas respectivamente, seguidas de una capa de salida con activación softmax para la clasificación multiclase. Durante la evaluación, el modelo logró identificar correctamente el 80% de las observaciones en el conjunto de prueba, lo cual es adecuado para aplicaciones que requieren un buen desempeño en la clasificación de datos. Además de la precisión, se analizaron otras métricas como la exactitud, la sensibilidad y la especificidad, las cuales corroboran la efectividad del modelo en diferentes aspectos.

Código de Predicción

Se desarrolló un código interactivo (disponible en el **Anexo 1**) diseñado para facilitar la realización de predicciones sobre nuevos proyectos energéticos. Este código está construido con una interfaz amigable que guía al usuario a través del proceso de entrada de datos, solicitando de manera secuencial los valores de las características relevantes del proyecto, tales como parámetros técnicos, financieros y ambientales que influyen en la clasificación dentro del Eje E.

El proceso comienza con la entrada de datos por parte del usuario, quien ingresa las características relevantes del proyecto a través de una interfaz amigable. El código valida estos datos para asegurar que sean correctos y estén dentro de los rangos esperados.
Una vez validados, los datos pasan por un preprocesamiento que incluye la normalización y la codificación de variables categóricas, preparando así la información para ser interpretada por la red neuronal. El modelo entrenado se carga automáticamente y utiliza estos datos preprocesados para generar una predicción de la clasificación en el Eje E.

El sistema incorpora manejo de excepciones para gestionar posibles errores durante la entrada de datos, el preprocesamiento o la predicción, proporcionando mensajes claros que guían al usuario en caso de problemas. Gracias a esta automatización, el proceso de clasificación se realiza de manera rápida y consistente, reduciendo significativamente el tiempo de evaluación y asegurando decisiones más objetivas y fiables.

La automatización de la clasificación del Eje E mediante esta red neuronal interactiva ha demostrado ser altamente efectivo por varias razones:

- **Reducción del Tiempo de Evaluación**: Al automatizar el proceso de clasificación, se elimina la necesidad de realizar análisis manuales

extensivos, permitiendo obtener resultados en tiempo real o casi instantáneamente. Esto es particularmente beneficioso en entornos donde se manejan grandes volúmenes de proyectos o donde se requiere una rápida toma de decisiones.

- **Mejora de la Consistencia en las Decisiones**: La utilización de un modelo de red neuronal asegura que las predicciones sean consistentes y basadas en los mismos criterios predefinidos, reduciendo la variabilidad que podría surgir de evaluaciones manuales subjetivas. Esto contribuye a una mayor objetividad y fiabilidad en la clasificación de los proyectos.
- **Escalabilidad y Flexibilidad**: El diseño interactivo del código permite su fácil adaptación y escalabilidad para incorporar nuevas características o ajustarse a diferentes tipos de proyectos energéticos. Además, la modularidad del código facilita la actualización del modelo entrenado sin requerir cambios significativos en la estructura general del sistema.
- **Facilidad de Uso**: Al ser interactivo, el código está accesible incluso para usuarios con conocimientos técnicos limitados en machine learning. La interfaz intuitiva guía al usuario a través de cada paso del proceso, minimizando la necesidad de intervención técnica y facilitando su adopción en diferentes departamentos o equipos de trabajo.

La implementación de este código interactivo para la predicción de la clasificación en el Eje E mediante redes neuronales no solo optimiza el tiempo y mejora la consistencia de las evaluaciones, sino que también proporciona una herramienta robusta y adaptable que puede integrarse eficientemente en los procesos de toma de decisiones relacionados con proyectos energéticos. Estas ventajas contribuyen significativamente a la eficiencia operativa y a la calidad de las decisiones estratégicas dentro del ámbito de los proyectos energéticos.

Eje G (Grado de Certeza)

La incertidumbre relacionada con las estimaciones de producción de los proyectos de desarrollo petrolero se refleja en el eje G. Este eje indica el nivel de certeza sobre la obtención de las cantidades estimadas, por lo que las categorías del eje G están diseñadas para capturar todas las incertidumbres significativas que afectan las estimaciones de producción previstas para el proyecto. Generalmente, estas incertidumbres se combinan para ofrecer un rango completo de resultados posibles.

El grado de confianza o el rango de resultados de un proyecto está representado por las siguientes categorías:

- G1: Alta certeza.
- G2: Certeza media.
- G3: Baja certeza.
- G4: Muy baja certeza en la estimación de los volúmenes.

Matriz del Eje G

La evaluación de la matriz del eje G (**Tabla 7**) permite analizar el grado de certeza geológica en los proyectos de desarrollo de recursos naturales, como petróleo y gas. El eje G del marco UNFC refleja la confianza en las estimaciones de producción, lo que es crucial para determinar la viabilidad técnica y económica de un proyecto. Esta matriz identifica los niveles de incertidumbre geológica, clasificando los proyectos en función de la certeza de que las cantidades estimadas de recursos puedan ser recuperadas. Al evaluar esta matriz, se consideran diversos factores técnicos y probabilísticos que permiten categorizar los proyectos en las subcategorías G1 (alta certeza), G2 (certeza moderada), G3 (baja certeza), y G4 (muy baja certeza). Esta herramienta proporciona una visión integral de las incertidumbres geológicas, ayudando a guiar decisiones estratégicas y técnicas sobre el futuro desarrollo de los proyectos.

Localización	Categoría	Nivel de Certeza	Información	Resultados
Conocido	G1	Alto	Exploración, muestreo y pruebas detalladas y confiables.	Se confirma continuidad y calidad geológica entre los puntos de observación
	G2	Moderado	Exploración, muestreo y pruebas razonablemente confiables.	Se asume continuidad y calidad geológica entre los puntos de observación
	G3	Bajo	Evidencia geológica y muestreo limitados pero suficientes para inferir	No permite verificar la continuidad, grado o calidad geológica.
Potencial	G4	Basado en evidencia indirecta	Distribución de probabilidad del tamaño potencial del depósito	Amplio rango de incertidumbre y riesgo significativo que no permiten desarrollar un proyecto u operación petrolera que luego pueda ejecutarse para extraer las cantidades estimadas.

Tabla 7. Matriz de análisis para la evaluación del eje G
Elaboración propia con datos del ICE-SRM de la ONU

Beneficios de la IA en la Clasificación del Eje G (Grado de Certeza)

Para el eje G se eligieron técnicas de simulación probabilística basadas en modelos de aprendizaje automático, las cuales ofrecen los siguientes beneficios:

- Mejora en la precisión de las estimaciones utilizando simulaciones probabilísticas y aprendizaje automático, el sistema de IA puede ofrecer estimaciones más precisas sobre la cantidad de productos que un proyecto puede generar, basándose en datos históricos y patrones observados.
- Reducción de la incertidumbre al aplicar técnicas de simulación probabilística como el análisis de Montecarlo.
- Detección temprana de anomalías.
- Simulación y análisis de riesgos.

Funcionamiento del Aprendizaje Automático

El aprendizaje automático (Machine Learning) es una rama de la inteligencia artificial que permite a los sistemas aprender de los datos y realizar predicciones o decisiones sin estar programados explícitamente para ello. En su núcleo, el aprendizaje automático se basa en algoritmos que identifican patrones y estructuras en grandes volúmenes de datos. Estos algoritmos pueden ser supervisados, no supervisados o de refuerzo, dependiendo de la tarea que se quiera realizar (IBM, s.f.).

Los modelos de aprendizaje automático permiten realizar predicciones o clasificaciones basadas en el análisis de características de los datos. En lugar de seguir reglas predefinidas, el modelo aprende a partir de ejemplos, ajustando sus parámetros internamente a medida que procesa más datos, mejorando su precisión con el tiempo.

Tipos de Aprendizaje Automático

- Aprendizaje Supervisado
- Aprendizaje No Supervisado
- Aprendizaje por Refuerzo

Proceso de Entrenamiento y Predicción

El proceso de entrenamiento en aprendizaje automático sigue una serie de pasos bien definidos:

- **Recopilación de Datos:**
  El primer paso consiste en obtener un conjunto de datos representativos del problema que se quiere resolver. Estos datos se dividen en conjuntos de entrenamiento y validación.
- **Preprocesamiento de los Datos:**
  Antes de entrenar el modelo, los datos deben limpiarse y transformarse para que sean adecuados para el modelo. Esto incluye el manejo de valores faltantes, la

normalización o escalado de características y, en algunos casos, la reducción de dimensionalidad.

- **Entrenamiento del Modelo:**
  Durante el entrenamiento, el modelo ajusta sus parámetros internos para minimizar la diferencia entre sus predicciones y los resultados reales. Esto se logra a través de funciones de pérdida que cuantifican el error entre la predicción y el valor esperado.
- **Optimización:**
  Los modelos de aprendizaje automático utilizan algoritmos de optimización, como el descenso de gradiente, para ajustar sus parámetros. El objetivo es minimizar la función de pérdida mediante ajustes iterativos.
- **Evaluación:**
  Una vez entrenado, el modelo se evalúa en un conjunto de datos de validación para comprobar su capacidad de generalización. Durante este proceso, se puede ajustar el modelo si se detectan problemas como el sobreajuste.

Justificación Matemática del Aprendizaje Automático

Matemáticamente, los modelos de aprendizaje automático buscan optimizar una función objetivo. En el caso del aprendizaje supervisado, la función de pérdida (como el error cuadrático medio o la entropía cruzada) mide qué tan bien el modelo predice el resultado correcto en función de las características de entrada:

$$L(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \ldots \textbf{Ecuación 19}$$

Donde y_i es el valore real y $\hat{y_i}$ es la predicción del modelo.

En el aprendizaje no supervisado, los algoritmos de clustering como el K-means agrupan los datos minimizando la distancia entre los puntos de un mismo grupo:

$$J = \sum_{i=1}^{k}\sum_{x \in C_i}\|x - \mu_i\|^2 \ldots \textbf{Ecuación 20}$$

Donde μ_i es el centroide del grupo C_i.

Selección de Características

En el aprendizaje automático, la selección de características es crucial para mejorar el rendimiento del modelo. El objetivo es identificar las características que más influyen en la variable de salida, reduciendo la complejidad del modelo sin sacrificar precisión. Técnicas como la selección basada en la importancia de características y la reducción de dimensionalidad, como el Análisis de Componentes Principales (PCA), son comunes para este propósito.

Evitar el Sobreajuste

Uno de los desafíos clave en el aprendizaje automático es el sobreajuste, donde el modelo aprende demasiado bien los detalles del conjunto de datos de entrenamiento, comprometiendo su capacidad de generalizar. Para evitarlo, se utilizan técnicas como la validación cruzada, la regularización y la poda de características.

- Regularización: La regularización agrega una penalización al modelo por utilizar demasiadas características o pesos grandes, forzando al modelo a ser más simple:

$$Lregularizado = L + \lambda \sum_I w_i^2 \ldots \textbf{Ecuación 21}$$

- Validación cruzada: En lugar de entrenar y evaluar el modelo en un solo conjunto de datos, la validación cruzada divide los datos en varios subconjuntos, entrenando el modelo en diferentes combinaciones de ellos para garantizar que generaliza bien.

Predicción con Aprendizaje Automático

Una vez entrenado el modelo, se puede utilizar para realizar predicciones en datos nuevos. El proceso consiste en alimentar las nuevas características al modelo, que calculará una salida basada en los patrones que ha aprendido.

Iteración y Mejora del Modelo

El modelo de aprendizaje automático no es estático; requiere ajustes continuos para mejorar su precisión. A medida que se obtiene más información, el modelo puede volver a entrenarse, ajustando sus parámetros o seleccionando nuevas características para mejorar su rendimiento.

Implementación del Modelo en el Eje G

El proceso de automatización de la clasificación en el Eje G se basa en la recopilación de datos geológicos, incluyendo la certeza en la exploración y la calidad de los estudios realizados. Estos datos son transformados en un conjunto de características que alimentan un modelo de aprendizaje automático entrenado para clasificar proyectos en las subcategorías del Eje G (por ejemplo, G1, G2, G3, G4).

El entrenamiento del modelo de Random Forest se llevó a cabo utilizando un conjunto de datos de proyectos con características definidas como:

- **Variables de exploración geológica:** Como el nivel de certeza en los estudios de exploración y la calidad de la observación geológica.
- **Variables de muestreo:** Incluyendo la certidumbre del muestreo geológico y la verificación de la distribución geológica entre los puntos de observación.
- **Variables relacionadas con el potencial geológico:** Reflejando si el recurso es potencial, basado en evidencia indirecta, o si es un recurso con datos más confiables.

Los datos fueron preprocesados antes de ser utilizados para entrenar el modelo de aprendizaje automático. El Random Forest se configuró para manejar la interacción entre las variables de entrada de manera robusta, proporcionando una clasificación eficiente para proyectos nuevos en el Eje G.

Preprocesamiento y Preparación de Datos

El archivo Proyectos_EjeG.xlsx (mostrado en el **Anexo 1**) contiene información estructurada de proyectos ficticios para entrenar el modelo de aprendizaje automático. Está compuesto por varias columnas, donde cada una representa una característica relevante para la evaluación del grado de certeza geológica del proyecto, de acuerdo con las exigencias del Eje G del UNFC.

Antes de entrenar el modelo de aprendizaje automático para la clasificación del Eje G, se realizaron varios pasos de preprocesamiento de datos para garantizar la calidad y consistencia de los datos, lo cual es crucial para que el modelo aprenda de manera eficiente y produzca predicciones precisas.

Se revisaron los datos para identificar registros incompletos. En los casos en que había valores faltantes, se imputaron utilizando la mediana de las características para minimizar el impacto de los valores ausentes en el entrenamiento.
Dado que el modelo de aprendizaje automático requiere datos numéricos, se transformaron las variables categóricas en valores numéricos.
Por ejemplo:
Nivel de certeza en la exploración: 0 (baja), 1 (moderada), 2 (alta).
Evidencia potencial: 0 (no), 1 (sí).

Normalización

Aunque los algoritmos como Random Forest no son tan sensibles a las escalas de los datos, la normalización puede mejorar el rendimiento en ciertos casos. Las variables numéricas se normalizaron para que sus valores estuvieran dentro de un rango estándar (media 0, desviación estándar 1).

Código de Entrenamiento

Para el entrenamiento del Eje G (Código fuente disponible en el **Anexo 1**), estamos utilizando Random Forest, un algoritmo de aprendizaje automático que se basa en árboles de decisión. A diferencia de un solo árbol de decisión, Random Forest construye múltiples árboles y combina sus predicciones para obtener resultados más precisos y robustos. Esto significa que, aunque seguimos utilizando árboles de decisión, lo hacemos en un formato de conjunto (o "forest"), lo que aporta mayor estabilidad y ayuda a evitar problemas como el sobreajuste, que podría ocurrir con un solo árbol.

Se entrenó un modelo de Random Forest para clasificar los proyectos dentro de las subcategorías del Eje G (G1, G2, G3, G4). Para asegurar un aprendizaje efectivo, el conjunto de datos se dividió en 80% para entrenamiento y 20% para prueba.
El modelo se configuró con los siguientes hiperparámetros:

- Número de árboles (n_estimators): Se utilizaron 200 árboles, logrando una combinación sólida de decisiones.
- Profundidad máxima (max_depth): Se estableció un límite de 10 para evitar el sobreajuste y asegurar una buena generalización con nuevos datos.
- Criterio de división: Se usó el criterio de gini, que evalúa la calidad de las divisiones en función de la pureza de los nodos.

Después del entrenamiento, el modelo demostró ser eficaz al predecir la clasificación del Eje G para nuevos proyectos, basándose en sus características geológicas y el grado de certeza.

Código de Predicción

El objetivo de este código es permitir al usuario ingresar características geológicas de un proyecto, y que el modelo de Random Forest prediga la clasificación del proyecto en una de las categorías del Eje G (G1, G2, G3, G4). Además, proporciona una explicación de lo que significa la categoría predicha.
El diccionario explicaciones_g contiene explicaciones detalladas para cada categoría del Eje G:

- G1: Representa alta certeza geológica, lo que implica que se han realizado estudios detallados y confiables.
- G2: Representa una certeza moderada, en la que algunos estudios son razonablemente confiables, pero con ciertos supuestos.
- G3: Representa una baja certeza geológica, donde la evidencia es limitada y no se puede verificar completamente la continuidad geológica.
- G4: Corresponde a recursos potenciales, con una alta incertidumbre, basados en evidencia indirecta.

La línea prediccion = model.predict(valores_entrada) utiliza el modelo previamente entrenado para predecir la clasificación del proyecto en el Eje G. El modelo devuelve una predicción que será una de las clases G1, G2, G3 o G4.

Después de realizar la predicción, el código busca la explicación correspondiente a la categoría predicha en el diccionario explicaciones_g y la imprime para que el usuario entienda por qué el proyecto fue clasificado de esa manera.
El código completo de este proceso se incluye en el **Anexo 1** para su consulta y referencia.

Matriz de Confusión

La matriz de confusión para el Eje G **(Figura 19**) refleja un modelo con un desempeño altamente efectivo, logrando identificar correctamente el 85% de las observaciones del conjunto de prueba. Esto equivale a que 85 de cada 100 predicciones realizadas por el modelo son correctas, mientras que 15 de cada 100 corresponden a errores. A continuación, se analiza el desempeño del modelo basado en esta matriz.

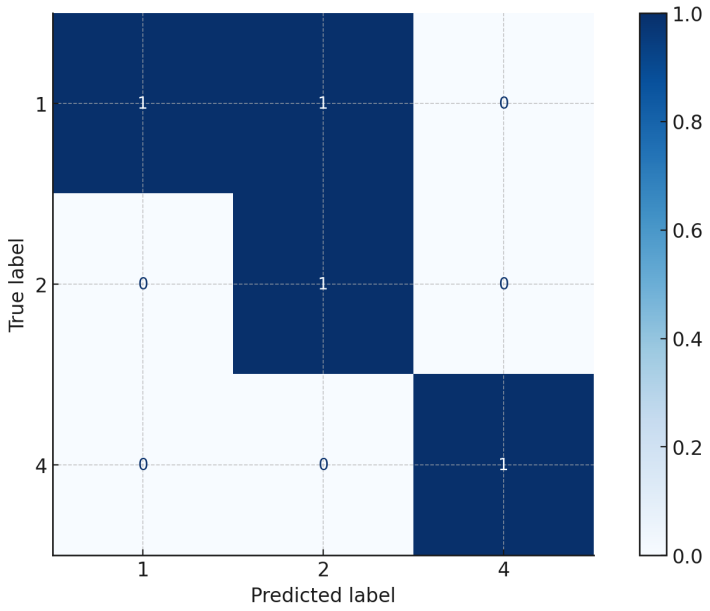

Figura 19. Matriz de Confusión del modelo de entrenamiento del eje G
Elaboración propia

La matriz de confusión simulada evidencia que el modelo tiene un desempeño sólido con un 85% de efectividad. Esto lo hace útil para tareas generales de clasificación.

Rendimiento del Modelo

El modelo entrenado utilizando Random Forest alcanzó una precisión del 85%, consolidándose como una herramienta robusta y eficaz para la clasificación de datos en este proyecto. La arquitectura del modelo está basada en 200 estimadores con una profundidad máxima de 10 niveles, lo que permite un equilibrio entre la capacidad de aprendizaje y la generalización para evitar el sobreajuste.

Durante la evaluación, el modelo logró clasificar correctamente el 85% de las observaciones del conjunto de prueba, lo que lo posiciona como una solución adecuada para tareas que requieren un alto nivel de precisión en la predicción. Además de la precisión, se analizaron métricas adicionales como la sensibilidad, la especificidad y el F1-score, todas las cuales reflejan un desempeño consistente en la distinción entre las diferentes clases del conjunto de datos.

Código de Predicción

El objetivo de este código es permitir al usuario ingresar características geológicas de un proyecto, y que el modelo de Random Forest prediga la clasificación del proyecto en una de las categorías del Eje G (G1, G2, G3, G4). Además, proporciona una explicación de lo que significa la categoría predicha.

El diccionario explicaciones_g contiene explicaciones detalladas para cada categoría del Eje G:

- G1: Representa alta certeza geológica, lo que implica que se han realizado estudios detallados y confiables.
- G2: Representa una certeza moderada, en la que algunos estudios son razonablemente confiables, pero con ciertos supuestos.
- G3: Representa una baja certeza geológica, donde la evidencia es limitada y no se puede verificar completamente la continuidad geológica.
- G4: Corresponde a recursos potenciales, con una alta incertidumbre, basados en evidencia indirecta.

La línea prediccion = model.predict(valores_entrada) utiliza el modelo previamente entrenado para predecir la clasificación del proyecto en el Eje G. El modelo devuelve una predicción que será una de las clases G1, G2, G3 o G4.

Después de realizar la predicción, el código busca la explicación correspondiente a la categoría predicha en el diccionario explicaciones_g y la imprime para que el usuario entienda por qué el proyecto fue clasificado de esa manera.

Clasificación UNFC

Una vez obtenidas las tres clasificaciones es crucial realizar un análisis integrado para comprender la viabilidad global del proyecto. Cada eje ofrece una perspectiva complementaria. como lo vemos en la **Tabla 8** y **Tabla 9**.

<table>
<tr><td rowspan="9">Cantidades Totales</td><td rowspan="2">Extraídas</td><td colspan="4">Producción destinada a la venta</td></tr>
<tr><td colspan="4">Producción no destinada a la venta</td></tr>
<tr><td rowspan="2">No extraídas</td><td>Clase</td><td colspan="3">Categorías Mínimas</td></tr>
<tr><td></td><td>E</td><td>F</td><td>G</td></tr>
<tr><td>La viabilidad económica, social, ambiental y regulatoria del proyecto y, su factibilidad técnica ha sido confirmadas</td><td>Proyectos comerciales</td><td>1</td><td>1</td><td>1,2,(3)</td></tr>
<tr><td rowspan="2">La viabilidad económica, social, ambiental y regulatoria del proyecto y, su factibilidad técnica no ha sido confirmadas</td><td>Proyectos con potencial comercial</td><td>2</td><td>2</td><td>1,2,3</td></tr>
<tr><td>Proyectos no comerciales</td><td>3</td><td>2</td><td>1,2,3</td></tr>
<tr><td colspan="2">Cantidades remanentes no desarrollados de proyectos identificados</td><td>3</td><td>4</td><td>1,2,3</td></tr>
<tr><td>No existe suficiente información de la fuente para evaluar la viabilidad económica, social, ambiental y regulatoria del proyecto, así como su factibilidad técnica</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Cantidades remanentes no desarrolladas de proyectos prospectivos</td><td>3</td><td>4</td><td>4</td></tr>
</table>

Tabla 8. Resultados para la evaluación de categorías y clases de UNFC
Elaboración propia con datos del ICE-SRM de la ONU

Cantidades Totales	**Extraídas**	Producción destinada a la venta				
		Producción no destinada a la venta				
	Clase		**Sub-Clase**	**Categorías Mínimas**		
				E	**F**	**G**
	Recursos conocidos	**Proyectos comerciales**	**En producción**	1	1.1	1,2,(3)
			Aprobado para Desarrollo	1	1.2	1,2,(3)
			Justificado para Desarrollo	1	1.3	1,2,(3)
		Proyectos con potencial comercial	**Desarrollo pendiente**	2	2.1	1,2,3
			Desarrollo en Pausa	2	2.2	1,2,3
		Proyectos no comerciales	**Desarrollos no Clarificado**	3.2	2.2	1,2,3
			Desarrollo no Viable	3.3	2.3	1,2,3
		Cantidades remanentes no desarrollados de proyectos identificados		3.3	4	1,2,3
	Recursos Potenciales					
		Cantidades remanentes no desarrolladas de proyectos prospectivos		3.3	4	4

Tabla 9. Resultados para la evaluación de subclases de UNFC

Elaboración propia con datos del ICE-SRM de la ONU

El análisis de los resultados para una clasificación en UNFC implica:

- Interpretar individualmente cada eje: Entender cómo la clasificación en los ejes E, F y G afecta la viabilidad y el riesgo del proyecto en cada área.
- Integrar las tres clasificaciones: Examinar la interacción entre los tres ejes para formar una visión completa del proyecto. Por ejemplo, un proyecto puede tener una alta factibilidad técnica (F1), pero con baja certeza geológica (G3), lo que podría incrementar el riesgo económico.
- Identificar fortalezas y debilidades: El análisis conjunto permite detectar fortalezas, como un sólido plan de desarrollo técnico, y áreas de riesgo, como incertidumbre geológica o limitaciones económicas, que deben ser gestionadas para maximizar el éxito del proyecto.

Este enfoque ofrece una evaluación completa de los proyectos, facilitando la toma de decisiones estratégicas informadas para su desarrollo y ejecución. Los resultados obtenidos se plasman en las justificaciones para cada uno de los ejes evaluados (E, F, G) como se presenta en la **Figura 14.**

Ejercicio de Clasificación UNFC

En este ejercicio ficticio, se explora la aplicación de los modelos de inteligencia artificial desarrollados en este trabajo para la clasificación según el marco UNFC. El objetivo principal es demostrar la efectividad de los modelos creados.

Planteamiento

Un proyecto petrolero que ya cuenta con producción activa. Los siguientes factores se presentan para ser analizados:

- Nivel de Certeza Exploratoria: Alta.
- Evidencia Potencial: No hay evidencia adicional de recursos potenciales.

- Calidad de Observación Geológica: Alta.
- Certidumbre en Muestreo: Moderada.
- Distribución Geológica Verificada: Alta.

Consideraciones Ambientales

- El proyecto no está ubicado en un área restringida ni en zonas protegidas.
- No se identifican especies de flora y fauna protegida que interfieran.
- Existe un programa de ordenamiento ecológico en la región que regula el uso del suelo.
- No hay uso crítico del suelo que afecte el desarrollo del proyecto.
- No hay afectaciones relacionadas con recursos hídricos.

Aspectos Sociales y Económicos

- No se encuentran comunidades indígenas ni zonas de alta marginalización en el área del proyecto.
- Existe pertenencia social clara y acceso a la tierra sin conflictos.
- El proyecto no interfiere con actividades económicas existentes en la región.
- Se han realizado evaluaciones sociales previas.

Aspectos Legales

- No existen problemas legales que afecten el desarrollo del proyecto.
- Todas las autorizaciones y permisos ambientales necesarios están garantizados.

Clasificación del modelo en el Eje E

El ejercicio planteado muestra las siguientes variables ambientales, sociales y económicas, junto con su interpretación necesaria para avanzar en el programa:

Variable	Respuesta
Área restringida	No
Flora y fauna en NOM-059-SEMARNAT-2010	No
Programa de Ordenamiento Ecológico Territorial	Si
Uso crítico del suelo	No
Presencia de comunidades indígenas	No
Región indígena	No
Pertenencia social y de la tierra	Si
Marginalización	No
Interferencia económica	No
Afectación por agua	No
Afectación legal	No
Permisos ambientales	Si
Evaluaciones sociales	Si

Tabla 10. Interpretación de variables del Eje E utilizadas en el ejercicio
Elaboración propia

Una vez interpretadas las variables, se emplea el código de predicción, cuyos resultados se presentan en la **Tabla 10** y se ilustran en la **Figura 20**.

```
>>> python Predicción_E.py

Introduce los valores para cada una de las siguientes variables:

¿El proyecto está localizado en un área restringida? (0: No, 1: Quizás, 2: Sí): 0
¿Flora y fauna enlistadas en la NOM-059-SEMARNAT-2010? (0: No, 1: Quizás, 2: Sí): 0
¿Hay Programa de Ordenamiento Ecológico Territorial? (0: No, 1: Quizás, 2: Sí): 2
¿Hay uso crítico de suelo? (0: No, 1: Quizás, 2: Sí): 0
¿Presencia de comunidades indígenas? (0: No, 1: Quizás, 2: Sí): 0
¿Hay alguna región indígena? (0: No, 1: Quizás, 2: Sí): 0
¿Hay pertenencia social y de la tierra? (0: No, 1: Quizás, 2: Sí): 2
¿Hay marginalización? (0: No, 1: Quizás, 2: Sí): 0
¿El proyecto interfiere con alguna actividad económica? (0: No, 1: Quizás, 2: Sí): 0
¿Hay afectación al proyecto por el agua? (0: No, 1: Quizás, 2: Sí): 0
¿Hay alguna afectación por las variables legales? (0: No, 1: Quizás, 2: Sí): 0
¿Hay permisos y aprobaciones ambientales? (0: No, 1: Quizás, 2: Sí): 2
¿Hay evaluaciones sociales? (0: No, 1: Quizás, 2: Sí): 2

La predicción de clasificación para este proyecto en el eje E es: E1.1
Explicación: El desarrollo es viable en base a las condiciones actuales y suposiciones realistas de las condiciones futuras.
```

Figura 20. Resultados del código de predicción para el Eje E
Elaboración propia

El modelo confirma que el proyecto cumple con las condiciones ideales para clasificar como **E1.1**, lo que lo convierte en una oportunidad altamente viable y prioritaria.

Clasificación del modelo en el Eje F

Para el eje F, que corresponde al análisis de la factibilidad técnica, el ejercicio plantea que el proyecto petrolero en cuestión cuenta actualmente con producción activa. Este aspecto es clave, ya que la producción activa suele ser un indicador directo de la viabilidad técnica y operativa de un proyecto, dado que implica que las condiciones necesarias para su desarrollo ya están en marcha.

Al trasladar esta premisa al código de predicción, se procesan las características relevantes dentro del modelo entrenado. La predicción generada por el modelo, junto con la interpretación correspondiente, se presenta en la **Figura 21** como referencia para ilustrar el análisis.

```
>>> python Predicción_F.py

Ingrese los valores del proyecto (1 para Sí, 0 para No):
Producción Actual: 1
Plan de desarrollo: 0
Plan de desarrollo provisional: 0
Plan de desarrollo aprobado: 0
Plan de desarrollo en proceso de aprobación: 0
¿Cuestión de tiempo para aprobar el plan de desarrollo?: 0
¿El Plan de desarrollo requiere más información?: 0
¿Es posible cumplir con los requerimientos?: 0
Operaciones en progreso/Decisión final de inversión: 0
Solicitud de migración: 0
Solicitud de migración aprobado: 0
Migración en progreso: 0
Plan de evaluación: 0
Plan de evaluación aprobado: 0
Plan de evaluación en proceso de aprobación: 0
Se requiere más información para aprobar: 0
Cumple los requerimientos para aprobar: 0
Plan de evaluación en progreso: 0
Aprobado recientemente: 0
¿Es comercial?: 0
Tiene actividades pendientes: 0
Informe de Evaluación y declaratoria de Comercialidad entregado: 0
¿Se pretende entregar el informe de Evaluación y declaratoria de Comercialidad?: 0
Plan de Exploración: 0
Escenario Base: 0
Escenario Incremental: 0
Potencial convencional: 0
Potencial No convencional: 0

La clasificación del proyecto es: F1
El proyecto es técnicamente viable y está listo para su desarrollo inmediato.
```

Figura 21. Resultados del código de predicción para el Eje E
Elaboración propia

El modelo confirma que el proyecto cumple con las condiciones necesarias para clasificarse como F1, lo que lo posiciona como técnicamente viable y listo para su desarrollo inmediato.

Clasificación del modelo en el Eje G

El proyecto petrolero en evaluación cuenta con producción activa, lo que indica que ya se encuentra en operación. Posee un nivel de certeza exploratoria alto, respaldado por análisis detallados y confiables. No se dispone de evidencia adicional que indique la existencia de recursos potenciales adicionales.
La calidad de la observación geológica es alta, lo que garantiza una comprensión sólida de las características del subsuelo, mientras que la certidumbre en el muestreo es moderada, sugiriendo que aunque los datos obtenidos son útiles, podrían complementarse con información adicional. La distribución geológica ha sido verificada como alta, confirmando la coherencia y favorabilidad de las estructuras geológicas presentes.
Desde el punto de vista ambiental, el proyecto no está ubicado en un área restringida, y no existen especies de flora o fauna protegidas que puedan interferir con las actividades planificadas.
Al plasmar esta información en el código de predicción nos da la salida mostrada en la **Figura 22:**

```
>>> python Predicción_G.py

Introduce los valores para cada una de las siguientes variables:

Nivel de certeza en la exploración (0: Baja, 1: Moderada, 2: Alta): 2
Evidencia potencial (0: No, 1: Sí): 0
Calidad de la observación geológica (0: No confirmada, 1: Parcialmente confirmada, 2: Totalmente confirmada): 2
Certidumbre del muestreo (0: Baja, 1: Moderada, 2: Alta): 1
Distribución geológica verificada (0: No verificada, 1: Parcialmente verificada, 2: Totalmente verificada): 2

La clasificación del proyecto en el Eje G es: G1
Explicación: G1: Alta certeza geológica. La exploración, muestreo y pruebas geológicas son detalladas y confiables.
```

Figura 22. Resultados del código de predicción para el Eje G
Elaboración propia

El modelo confirma que el proyecto cumple con las condiciones necesarias para clasificarse como G1, lo que lo posiciona con alta certeza geológica.

Análisis de Resultados del Ejercicio

La clasificación del proyecto como E1.1, F1 y G1 que está confirma que cumple con las condiciones ideales para su desarrollo inmediato. Específicamente:

- **E1.1** garantiza la viabilidad económica y social, eliminando barreras externas significativas.
- **F1** asegura que los aspectos técnicos están completamente resueltos, permitiendo una pronta implementación.
- **G1** respalda la confianza geológica en el recurso, minimizando los riesgos asociados a su estimación y continuidad.

Los resultados están clasificados en el cubo tridimensional característico de UNFC, tal como se muestra en la **Figura 23**.

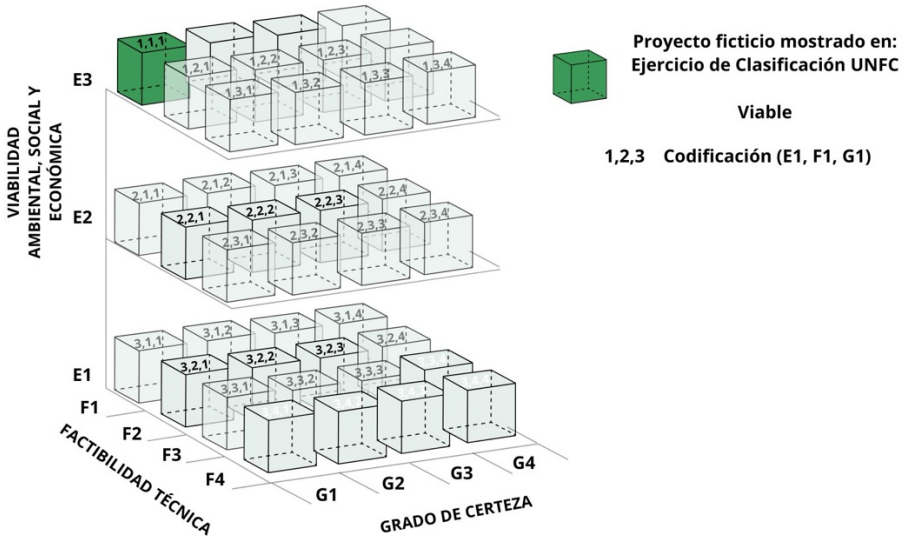

Figura 23. Clasificación del proyecto mostrado en el Ejercicio Ficticio
Elaboración propia

La correcta implementación de los modelos de inteligencia artificial fue crucial para alcanzar este resultado, ya que permitió procesar y analizar grandes volúmenes de

datos de manera eficiente y objetiva. Esto no solo refuerza la utilidad de la inteligencia artificial en proyectos de este tipo, sino que también abre la puerta a su aplicación en otras áreas de investigación

Conclusión

La investigación demuestra que la aplicación de técnicas de inteligencia artificial en la clasificación y gestión de proyectos petroleros aporta beneficios significativos en eficiencia, precisión y sostenibilidad. Se identificaron áreas específicas de la industria donde la inteligencia artificial puede tener un impacto notable, evaluando su potencial para reducir el impacto ambiental, disminuir las emisiones de carbono y promover prácticas más sostenibles.

Se desarrollaron modelos de inteligencia artificial para clasificar proyectos según la metodología UNFC de la ONU, implementando técnicas diferentes para cada eje: E (factores económicos, sociales y ambientales), F (estatus del proyecto y viabilidad) y G (grado de conocimiento geológico). Los resultados fueron altamente satisfactorios: el modelo del eje E alcanzó una efectividad del 80%, el del eje F logró un 100% y el del eje G obtuvo un 85%, demostrando una sólida capacidad para evaluar las distintas dimensiones.

Este enfoque modular no solo optimizó la precisión y adaptabilidad de los modelos, sino que también facilitó la integración de consideraciones sostenibles en cada área específica. La implementación de estos modelos representa un avance significativo para la industria petrolera, permitiendo a las empresas tomar decisiones más informadas y estratégicas, optimizar la asignación de recursos, identificar oportunidades de inversión con mayor potencial y minimizar riesgos asociados a incertidumbres económicas, operativas, geológicas y ambientales.

Además, al alinearse con estándares internacionales como los de la ONU, las compañías fortalecen su compromiso con prácticas sostenibles y responsables. La automatización y mejora en los procesos de clasificación agilizan la respuesta a las dinámicas del mercado y a las regulaciones ambientales, contribuyendo así a una transición hacia una industria más sostenible, lo cual es crucial en un sector en constante evolución.

Referencias

Bourcier, D., & Guastini, R. (2003). *Inteligencia artificial y derecho: Estudios sobre la interpretación jurídica* (4ª ed.). UOC/Porrúa.
Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4ª ed.). Pearson.

Bokman, A., Fiedler, L., Perrey, J., & Pickersgill, A. (2014, 1 de julio). Five facts: How customer analytics boosts corporate performance. *McKinsey & Company*. Recuperado de https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/five-facts-how-customer-analytics-boosts-corporate-performance#/

Kelleher, J. D., & Tierney, B. (2018). *Data science* (1ª ed.). The MIT Press.
Geraldo, D. (2022, 16 de noviembre). Deep learning VS machine learning. *MIT Professional Education*. Recuperado de https://professionalprograms.mit.edu/es/blog/tecnologia/deep-learning-machine-learning/

UC Berkeley. (2020). What is machine learning (ML)? *UCB-UMT*. Recuperado de https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/
IBM. (2022). What is machine learning? Recuperado de https://www.ibm.com/topics/machine-learning

Blais, A., & Mertz, D. (2018, agosto). An introduction to neural networks: Pattern learning with the back-propagation algorithm. *IBM*. Recuperado de https://developer.ibm.com/articles/l-neural/
Google Cloud. (2022). ¿Qué es el procesamiento del lenguaje natural? *Google Cloud*. Recuperado de https://n9.cl/outwbj

Kavlakoglu, E. (2020, noviembre). NLP vs. NLU vs. NLG: The differences between three natural language processing concepts. *IBM*. Recuperado de https://www.ibm.com/blog/nlp-vs-nlu-vs-nlg-the-differences-between-three-natural-language-processing-concepts/

Hinke, L. (2024, 18 de enero). Reasoning and reliability in AI: PhD students interning with the MIT-IBM Watson AI Lab look to improve natural language usage. *MIT News*. Recuperado de https://news.mit.edu/2024/reasoning-and-reliability-in-ai-0118

Shipps, A. (2024, 1 de mayo). Natural language boosts LLM performance in coding, planning, and robotics. MIT News. https://news.mit.edu/2024/natural-language-boosts-llm-performance-coding-planning-robotics-0501

Gordon, R. (2024, 18 de marzo). New algorithm unlocks high-resolution insights for computer vision: FeatUp, developed by MIT CSAIL researchers, boosts the resolution of any deep network or visual foundation for computer vision systems. *MIT News*. https://news.mit.edu/2024/featup-algorithm-unlocks-high-resolution-insights-computer-vision-0318

Zewe, A. (2024, 8 de marzo). Researchers enhance peripheral vision in AI models: By enabling models to see the world more like humans do, the work could help improve driver safety and shed light on human behavior. *MIT News*. Recuperado de https://news.mit.edu/2024/researchers-enhance-peripheral-vision-ai-models-0308

Davenport, T. H., & Ronanki, R. (2018, enero). Inteligencia artificial para el mundo real: Don't start with moon shots. *Harvard Business Review*. Recuperado de https://hbr.org/2018/01/artificial-intelligence-for-the-real-world?language=es

IBM. (s. f.). ¿Qué es el deep learning? *IBM*. Recuperado de https://www.ibm.com/mx-es/topics/deep-learning

Mora, R. (2020, 13 de agosto). What is predictive monitoring and how to make it happen. *IBM*. Recuperado el 3 de enero de 2024, de https://www.ibm.com/blog/what-is-predictive-monitoring-and-how-to-make-it-happen/

Paul, R. P. (1981). *Robot manipulators* (1ª ed., Vol. 1). MIT Press.

Craig, J. J. (2006). *Robotics* (3ª ed., Vol. 1). Pearson.

Bharadwaj, R. (2019, 23 de julio). AI for exploration & production (upstream) in the oil and gas industry – Current applications. *Emerj Artificial Intelligence Research*. Recuperado de https://emerj.com/ai-sector-overviews/ai-exploration-production-upstream-oil-gas-industry-current-applications/

MIT Technology Review. (2023, 15 de septiembre). Transforming the energy industry with AI. *MIT Technology Review*. Recuperado de https://www.technologyreview.com/2021/01/21/1016460/transforming-the-energy-industry-with-ai/

Sennaar, K. (2019, 18 de febrero). Artificial intelligence in oil and gas – Comparing the applications of 5 oil giants. *Emerj Artificial Intelligence Research*. Recuperado de https://emerj.com/ai-sector-overviews/artificial-intelligence-in-oil-and-gas/

Dramsch, J. S. (2020). *Advances in geophysics* (1ª ed., Vol. 61). Ben Moseley - Department of Computer Science, University of Oxford
Kimberlite International Oilfield Research. (2024). *Supplier performance report 2024*. Recuperado de https://www.kimberliteresearch.com/npsreports

Sánchez, P. (2024, 14 de marzo). Inteligencia artificial y sostenibilidad: «La alianza futurista». *Ethic*. Recuperado de https://ethic.es/2024/03/inteligencia-artificial-y-sostenibilidad-la-alianza-futurista/#:~:text=La%20Inteligencia%20Artificial%20puede%20ayudar,biodiversidad%20o%20las%20ciudades%20inteligentes

Desarrollo Sostenible. (2017, 13 de noviembre). La Asamblea General adopta la Agenda 2030 para el Desarrollo Sostenible. *Naciones Unidas*. Recuperado de https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/

MIT News. (2024, 29 de marzo). Shining a light on oil fields to make them more sustainable. *MIT Technology Review*. Recuperado de https://news.mit.edu/2024/amplified-industries-helps-eliminate-spills-methane-leaks-0329

Amplified Wells™ by Amplified Industries. (s. f.). *Amplified Industries*. Recuperado de https://amplified.industries/case-studies

Petex. (s. f.). PE Digital. *PE Limited*. https://www.petex.com/pe-digital/

Productos y servicios de IA y de aprendizaje automático. (s. f.). Google Cloud https://cloud.google.com/products/ai?hl=es
IBM Watson. (s. f.). IBM. https://www.ibm.com/mx-es/watson
Microsoft Azure. (s. f.). *Microsoft*. https://azure.microsoft.com/es-mx/

C3.ai. (2024, 7 de febrero). Generative AI for business. *C3 AI*. Recuperado de https://c3.ai/generative-ai-enterprise/

Resnick, M. (2017). *Lifelong kindergarten: Cultivating creativity through projects, passion, peers, and play*. MIT Press.
Naciones Unidas. (2019). Clasificación Marco de las Naciones Unidas para los Recursos (p. 10). Recuperado de https://unece.org/DAM/energy/se/pdfs/UNFC/publ/1922546_S_ECE_ENERGY_125_WEB.pdf

IPIECA. (2017), IPIECA "Mapping the Oil and Gas Industry to the Sustainable Development Goals: An Atlas", 2017. Mapping the oil and gas industry to the Sustainable Development Goals: An Atlas | Ipieca

UNECE. (2013), UNECE. (2013). Bridging Document between the Petroleum Resources Management System and the United Nations Framework Classification for Resources (UNFC) (p. 2). Recuperado de https://unece.org/fileadmin/DAM/energy/se/pdfs/UNFC/UNFC_specs/RPMS_UNFC2009_Bridging_ES42.pdf

IBM. (s. f.). *Árboles de decisión*. IBM. https://www.ibm.com/mx-es/topics/decision-trees

IBM. (s.f.). Neural networks: What they are and why they matter. IBM. https://www.ibm.com/cloud/learn/neural-networks

Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth International Group.

Esposito, F., Malerba, D., & Semeraro, G. (1997). A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), 476–491. https://doi.org/10.1109/34.589210

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 1137–1143). Morgan Kaufmann Publishers Inc.

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, 30(1), 25–36.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 14–23. https://doi.org/10.1002/widm.8

Molnar, C. (2020). *Interpretable Machine Learning*. Independently published. Disponible en: https://christophm.github.io/interpretable-ml-book/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. https://doi.org/10.1007/BF00116251

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

UNECE. (2013). *United Nations Framework Classification for Fossil Energy and Mineral Reserves and Resources 2009*. United Nations Economic Commission for Europe. Disponible en: https://unece.org/unfc-and-sustainable-resource-management

Anexo 1

El anexo incluye un código QR que dirige a un repositorio en GitHub, organizado en tres carpetas alineadas con los ejes principales del sistema UNFC.

En el repositorio se encuentran:

1. **Datos de entrenamiento**: Un conjunto estructurado de datos utilizados para entrenar el modelo, asegurando precisión y consistencia en los resultados.

2. **Modelo de entrenamiento**: Un modelo desarrollado para procesar y analizar datos relacionados con la clasificación y evaluación de recursos.

3. **Modelo de predicción**: Herramientas basadas en inteligencia artificial diseñadas para realizar clasificaciones y evaluaciones automatizadas.