

Estudio exploratorio de la aplicación de la ciencia de datos en el área de ingeniería de producción de hidrocarburos del ciclo profesional del pensum de la carrera de ingeniería de petróleos de la Universidad Industrial de Santander

Silvia Juliana Franco Ardila y Kel Guzmán Bocanegra

Trabajo de Grado para Optar por el Título de ingenieros de petróleos

Director

Fernando Enrique Calvete González

Magíster en Informática



Escuela de Ingeniería
de Petróleos

Universidad Industrial de Santander
Facultad de Ingeniería Físicoquímicas
Escuela de Ingeniería de Petróleos
Bucaramanga

2024

Agradecimientos

Agradezco a Dios por todo, por proveer los recursos para culminar esta etapa de mi vida y por permitirme tener a mi madre Ingrid Bocanegra Mena y a mi Abuelo José Luis Bocanegra Quila que siempre me brindan su amor y apoyo incondicional para cumplir mis objetivos, gracias a ellos soy un hombre con principios. Sin su apoyo y enseñanzas esto no hubiese sido posible.

Le doy gracias a mi hermano Ian Guzman Bocanegra y a mi prima Jennifer Vanessa Umaña por acompañarme en cada paso de mi vida.

A mis tías Irina Bocanegra, Tamara Inés Bocanegra, María del Mar Bocanegra y a mi tío José Luis Bocanegra agradecerles por los consejos y enseñanzas brindadas, al igual que al resto de mis familiares. Sé que están contentos de saber que con sacrificio y esfuerzo estoy logrando culminar mi etapa universitaria.

Muchas gracias a mi esposa Ayra Valentina Chaparro por brindarme su amor y soporte en lo transcurrido de esta etapa y por tener siempre las palabras adecuadas para aconsejarme.

Agradezco enormemente al Ingeniero Fernando Enrique Calvete, por su disposición, enseñanzas, conocimiento y para llevar a cabo este proyecto, es una persona de grandes valores el cual merece que todo lo bonito le suceda.

Por último, agradecerle a mi compañera de Tesis Silvia Juliana Franco Ardila por la disposición y dedicación para finalizar con este trabajo, le deseo muchos éxitos y bendiciones en esta etapa que se viene.

Kel Guzman Bocanegra

Agradecimientos

Agradezco en primer lugar a Dios por guiarme en cada momento de dificultad y siempre iluminar mi camino. Quiero expresar mi profunda gratitud a mis padres, Gloria Ardila Carrero y Alberto Franco Ferreira por el amor, paciencia, comprensión y apoyo, así como ellos sienten orgullo por este logro, yo me siento orgullosa de brindarles uno de los tantos logros que quiero para mi vida. A mi hermano mayor, que vivió y sufrió conmigo travesías para lograr esta ingeniería. Gracias por enseñarme a ser valiente y libre, nunca me faltaron palabras de motivación gracias a él. asimismo, agradezco a mi compañero de grado Kel Guzmán Bocanegra, compañero, amigo y colega, por tolerar mis pasos, por armar lo que teníamos en mente y su comprensión, compartir este camino contigo ha enriquecido mi experiencia.

Agradezco al ingeniero Fernando Enrique Calvete por su orientación en esta etapa sabiduría durante todo el proceso de elaboración del trabajo de grado, ya que con su gran corazón exalta su vocación con su aprendizaje. Finalmente, agradezco a mis profesores, amigos, familiares y a cada persona que estuvo en mi proceso apoyándome, animándome e invitándome a soñar, a Marcela Franco, Armando cortes y Paula cortes ¡muchas gracias!

Silvia Juliana Franco Ardila

Tabla de contenido

Introducción	12
1. Objetivos	14
1.1. Objetivo General.....	14
1.2. Objetivos Específicos.....	14
2. Marco teórico	15
2.1. Ciencia de datos	15
2.2. Ciencia de datos e inteligencia artificial	16
2.2.1. Inteligencia artificial	16
2.2.2. Aplicaciones de la inteligencia artificial	17
2.3. Subconjuntos de la inteligencia artificial	19
2.3.1. Aprendizaje automático ML.....	20
2.3.2. Minería de datos	20
2.3.3. Robótica inteligente.....	20
2.3.4. Procesamiento del lenguaje natural.....	21
2.3.5. Sistemas expertos	21
2.3.6. Sistemas de visión	21
2.3.7. Planificación y programación.....	22
2.3.8. Búsqueda y optimización	22
2.3.9. Redes neuronales artificiales	24
2.3.10. Agentes inteligentes y sistemas multiagente	24
2.3.11. Lógica difusa.....	25
2.4. Tipos de inteligencia artificial	25
2.5. Técnicas de la ciencia de datos	25
2.5.1. Big data	26
2.5.2. Minería de datos	28
2.5.3. Lógica difusa.....	28

2.5.4.	Subconjuntos machine learning	29
2.5.5.	Algoritmos de machine learning utilizados en los sistemas de producción de la industria petrolera	34
2.5.6.	Otros conceptos relacionados con la ciencia de datos.....	41
3.	Metodología para la selección en los procedimientos de diseño en el área de ingeniería de producción de hidrocarburo del ciclo profesional de ingeniería de petróleos de la Universidad Industrial de Santander	51
3.1.	Contenido de las asignaturas métodos de producción, facilidades de superficie e ingeniería del gas del ciclo profesional de ingeniería de petróleos.....	52
3.2.	Procedimientos de diseño de ingeniería.....	55
3.3.	Selección de procedimientos de diseño de ingeniería en las asignaturas métodos de producción, ingeniería del gas y facilidades de superficie del área de hidrocarburos.....	57
3.4.	Procedimientos de ingeniera presentes en el plan de estudios del área de producción de los hidrocarburos.....	58
4.	Descripción de la metodología para la implementación de la ciencia de datos en el área de ingeniería de producción basados en ejemplos seleccionados.	65
4.1.	Análisis bibliométrico	65
4.1.1.	Scopus	66
4.1.1.1.	Indicadores básicos	66
4.2.	Indicadores relevantes para la selección de ejemplos	69
4.2.1.	Ejemplos métodos de producción.....	70
4.2.2.	Ejemplos ingeniería del gas.....	87
4.2.3.	Ejemplos facilidades de superficie.....	95
5.	Comparación del procedimiento obtenido en la aplicación de las técnicas de la ciencia de datos respecto a la forma tradicional de los ejemplos seleccionados	102
5.1.	Un enfoque basado en redes neuronales de inteligencia artificial para predecir la producción en yacimientos no convencionales: análisis comparativo con la curva de declive	104
5.2.	Aplicación de técnicas de inteligencia artificial para predecir la productividad de los pozos de espina de pescado	106

5.3. Predicción de la optimización del sistema de producción de petróleo a través de modelos de regresión lineal y herramientas Big Data	107
5.4. Inteligencia artificial para predecir el IPR de pozos verticales en yacimientos derivados de gas en solución: un nuevo enfoque	108
5.5. Gestión de la integridad de las tuberías de CO ₂ : Un enfoque gemelo digital	109
5.6. Una nueva correlación para predecir los caudales de gas utilizando inteligencia artificial: un enfoque industrial 4.0	110
5.7. Aplicación de técnicas de computación blanda a la medición de flujo multifásico: una revisión 111	
6. Conclusiones	114
7. Recomendaciones	115
Bibliografía.....	116
Anexos.....	125

Tabla de figuras

Figura 2.1. Diferentes campos de la ciencia de datos.	16
Figura 2.2. Aplicaciones de la inteligencia artificial.....	17
Figura 2.3. Subconjuntos de la inteligencia artificial.....	19
Figura 2.4 Tipos de inteligencia artificial	26
Figura 2.5. Características de la Big Data.....	27
Figura 2.6. Clasificación de la minería de datos	29
Figura 2.7. Subconjuntos de machine learning	30
Figura 2.8. Tipos de algoritmos del aprendizaje supervisado	31
Figura 2.9. Tipos de algoritmos de Aprendizaje no supervisado	32
Figura 2.10. Neurona biológica y neurona artificial	35
Figura 2.11. Funciones de activación.....	36
Figura 2.12. Arquitecturas de las redes neuronales artificiales. 12.a. Red monocapa. 12.b. Red multicapa. 12.c. Red recurrente.....	37
Figura 2.13. Tipos de redes neuronales.....	38
Figura 2.14. Modelo digital del proceso de producción, modelo cíclico	43
Figura 2.15. Red neuronal neuro-difusa (ANFIS).....	47
Figura 2.16. Arquitectura típica de Redes de función de base radial RBF	48
Figura 4.1. Desarrollo del tema con el pasar de los años en Scopus.	67
Figura 4.2. Área temática en Scopus.....	68
Figura 4.3. Palabras clave en Scopus	69
Figura 4.4. Modelo ANN, según ejemplo	72
Figura 4.5. Gráfica coeficiente de correlación R, VS N – número de laterales, D – Distancia entre laterales, L – Longitud lateral, Kh/Kv – Relación de permeabilidad, Pwf – Presión de fondo de pozo que fluye.....	75
Figura 4.6. Modelo ANN ejemplo de producción.....	76
Figura 4.7. Comparación de las diferentes técnicas de inteligencia artificial.	78
Figura 4.8. Modelo CRISP (Proceso Estándar Interindustrial Cruzada).....	81
Figura 4.9. Modelo de predicción actual vs modelo por progresión lineal	84
Figura 4.10. IPR de campo comparados con el método de Fetkovich y AI.....	87
Figura 4.11. Bloques de construcción de gemelo digital.....	89
Figura 4.12. Modelo de ANN para predecir los caudales de gas	93
Figura 5.1. Ventajas y desventajas del área de producción de la ciencia de datos.....	103
Figura 5.2. Método de ciencia de datos para predecir la producción en yacimientos convencionales	104

Figura 5.3. Método convencional para predecir la producción en yacimientos convencionales	105
Figura 5.4. Método de ciencia de datos para predecir la productividad de los pozos de espina de pescado	106
Figura 5.5. Método convencional para predecir la productividad de los pozos de espina de pescado.....	106
Figura 5.6. Método de la ciencia de datos vs método convencional para la predicción.....	107
Figura 5.7. Método de ciencia de datos para predecir el IPR de pozos verticales	108
Figura 5.8. Método convencional para predecir el IPR de pozos verticales	109
Figura 5.9. Método de ciencia de datos para la gestión de la integridad de las tuberías de CO ₂	109
Figura 5.10. IPR de campo comparados con el método de Fetkovich y AI	110
Figura 5.11. Método de ciencia de datos para predecir los caudales de gas	110
Figura 5.12. Método convencional para predecir los caudales de gas	111
Figura 5.13. Método de ciencia de datos para la medición de flujo multifásico	111
Figura 5.14. Método convencional para la medición de flujo multifásico	112
Figura 5.15. Método convencional para detectar fugas en gasoductos	113
Figura 5.16. Método convencional para detectar fugas en gasoductos	113

Lista de tablas

Tabla 2.1. Funciones de la ANN	39
Tabla 3.1. Contenido de las asignaturas métodos de producción, facilidades de superficie e ingeniería del gas del ciclo profesional de ingeniería de petróleos	54
Tabla 3.2. Selección áreas de producción.....	58
Tabla 3.3. Contenido de los procedimientos de ingeniería de producción de hidrocarburos	59
Tabla 3.4. Procedimientos de diseño métodos de producción.....	62
Tabla 3.5. Procedimientos de diseño de ingeniería del gas	64
Tabla 3.6. Procedimientos de facilidades de superficie.....	65
Tabla 4.1. Análisis estadístico del ejemplo	75
Tabla 4.2. Análisis estadístico del ejemplo	75
Tabla 4.3. Resultados de los diversos escenarios del ejemplo	77
Tabla 4.4. Resultados del sistema de lógica difusa, según el ejemplo	77
Tabla 4.5. Resultados del sistema de lógica difusa, según el ejemplo	77
Tabla 4.6. Resumen de parámetros y resultados del modelo de red de función de base radial	78

Lista de ecuaciones

Ecuación 4-1. Número de capas ocultas.....	72
Ecuación 4-2. Ecuación numérica para el escalado o normalización de datos.....	82
Ecuación 4-3. Valor de error cuadrático medio RMSE.....	83

Resumen

Título: Estudio exploratorio de la aplicación de la ciencia de datos en el área de ingeniería de producción de hidrocarburos del ciclo profesional del pensum de la carrera de ingeniería de petróleos de la Universidad Industrial de Santander.

Autor: Silvia Juliana Franco Ardila y Kel Guzmán Bocanegra.

Palabras clave: Data science, Machine Learnig (ML), Deep Learning (DL), Big Data Analytics, Data mining, Internet of Things (IoT), Redes neuronales artificiales, Algoritmos genéticos.

Descripción: La ciencia de datos o también conocida como data Science se define como “el arte y la ciencia de adquirir conocimiento a través de datos” (Ozdemir, 2016). Es por ello, que la ciencia de datos busca combinar herramientas, métodos y tecnologías que generen un significado a partir de un almacenamiento de datos evitando situaciones inmediatas a problemas y/o anticipar sucesos que afecten las operaciones, de tal forma, que un estudio a la aplicación de la ciencia de datos en el área de ingeniería de producción de hidrocarburos permitirá conocer de forma más precisa y detallada las técnicas que se podrán usar y de acuerdo con estudios hechos dar a conocer cual sería la más viable de aplicar en el área de ingeniería de producción de hidrocarburos.

El actual proyecto de investigación se basa en desarrollar un estudio exploratorio que permita definir la aplicación de las diferentes técnicas de la ciencia de datos en las áreas de la ingeniería de producción de hidrocarburos del pensum académico de la carrera de ingeniería de petróleos de la Universidad Industrial de Santander. Como lo son “ingeniería del gas, Facilidades de superficie y Métodos de producción”. Para ello, se inició la investigación de diferentes técnicas de la ciencia de datos aplicadas a la ingeniería de producción de hidrocarburos actuales para así definir los procedimientos de diseño que se pueden aplicar en el área de ingeniería de producción, después, se describió metodológicamente la aplicación de técnicas de la ciencia de datos en el área de ingeniería de producción para comparar los resultados obtenidos en la aplicación de la ciencia de datos respecto a la forma tradicional como se efectúan. Todo esto para dar a conocer de manera clara y concisa las técnicas que se podrían usar y, según estudios realizados, dar a conocer la más viable de aplicar en el área de ingeniería de producción de hidrocarburos.

Abstract

Title: Exploratory study of the application of data science in the area of hydrocarbon production engineering of the professional cycle of the curriculum of the petroleum engineering career of the Industrial University of Santander.

Authors(s): Silvia Juliana Franco Ardila, Kel Guzmán Bocanegra.

Key Words: Data science, Machine Learning (ML), Deep Learning (DL), Big Data Analytics, Data mining, Internet of Things (IoT), Artificial neural networks, Algorithms.

Description: Data science or also known as data Science is defined as "the art and science of acquiring knowledge through data" (Ozdemir, 2016). That is why data science seeks to combine tools, methods and technologies that generate meaning from data storage avoiding immediate situations to problems and / or anticipate events that affect operations, so that a study of the application of data science in the area of hydrocarbon production engineering will allow to know more precisely and in detail the techniques that can be used and according to studies made to make known which would be the most viable to apply in the area of hydrocarbon production engineering.

The current research project is based on developing an exploratory study to define the application of the different techniques of data science in the areas of hydrocarbon production engineering of the academic curriculum of the petroleum engineering career of the Industrial University of Santander. Such as "gas engineering, surface facilities and production methods". For this purpose, the investigation of different techniques of data science applied to the current hydrocarbon production engineering was initiated in order to define the design procedures that can be applied in the area of production engineering, then, the methodological description of the application of data science techniques was described.

Introducción

Hoy en día el uso de la ciencia de datos ha ganado creciente relevancia, transformando la manera en que se abordan los desafíos y se toman decisiones estratégicas en el sector petrolero. La ciencia de datos se encarga de extraer información y conocimiento a partir de datos. El área de ingeniería de producción es una de las áreas más importantes de la industria petrolera, ya que es la encargada del manejo y la conducción de Hidrocarburos del yacimiento a las áreas de donde se van a comercializar, de esta manera también trata del método de explotación, diseño de instalaciones de superficie, y optimización de los pozos y no siendo menos importante el tiempo de explotación de los yacimientos para no generar pérdidas económicas, es por eso que este proyecto se enfoca en buscar como la ciencia de datos puede mejorar dichas técnicas convencionales combinando herramientas, métodos y tecnologías que generen un significado a partir de un almacenamiento de datos evitando situaciones de riesgo, de modo que, por medio de esta tecnología innovadora se busca dar soluciones inmediatas a problemas y/o anticipar sucesos que afecten las operaciones.

Esta investigación, se enfoca en realizar un análisis bibliométrico de las diferentes técnicas de la ciencia de datos aplicadas a tres asignaturas del pensum académico de ingeniería de petróleos en la Universidad Industrial de Santander, estas asignaturas son: métodos de producción, Facilidades de superficie e ingeniería del gas.

En este sentido, la investigación se llevó a cabo en cuatro pasos: El primer paso consiste en hacer una revisión bibliográfica y bibliométrica de las diferentes técnicas de la ciencia de datos que existen en la actualidad con enfoque en el área de producción; El Segundo paso, define los procedimientos de diseño en dicha área, con la finalidad de dar a conocer donde se puede

aplicar la ciencia de datos, tomando como guía el contenido de las materias del área de producción del pensum académico de ingeniería de petróleos de la Universidad Industrial de Santander; El tercer paso, muestra una recopilación de ejemplos de los diferentes procesos de diseño de la aplicación de las técnicas de ciencia de datos a los procedimientos de ingeniería de las asignaturas de Métodos de producción, Facilidades de superficies e ingeniería del Gas; Por último, se seleccionaron los ejemplos con más influencia hasta el momento y se realizó una descripción metodológica de la aplicación de las técnicas con mayor relevancia en el área de producción de hidrocarburos. Así mismo, se compararon los resultados con las técnicas convencionales de estos temas. Se concluyó que las técnicas más usadas en la industria petrolera son las técnicas de aprendizaje automático y redes neuronales, estas presentan un escenario más dinámico, centrado en lograr resultados cada vez más exactos y eficientes, maximizando la producción de manera sostenible y, por último, se realizaron algunas recomendaciones para estudios futuros.

1. Objetivos

1.1. Objetivo General

Desarrollar un estudio exploratorio que permita definir la aplicación de las diferentes técnicas de la ciencia de datos en las áreas de la ingeniería de producción de hidrocarburos del pensum académico de la carrera de ingeniería de petróleos de la Universidad Industrial de Santander.

1.2. Objetivos Específicos

Investigar las diferentes técnicas de la ciencia de datos aplicadas a la ingeniería de producción de hidrocarburos que existen en la actualidad.

Definir los procedimientos de diseño en el área de ingeniería de producción del pensum académico de la carrera ingeniería de petróleos en la universidad Industrial de Santander.

Describir metodológicamente la aplicación de las técnicas de la ciencia de datos en el área de ingeniería de producción.

Realizar la comparación de los resultados obtenidos en la aplicación de las técnicas de la ciencia de datos respecto a la forma tradicional como se efectúan.

2. Marco teórico

Basado en la tesis de la universidad industrial de Santander “Estudio exploratorio de la aplicación de la ciencia de datos en las materias del ciclo profesional del pénsum de la carrera de ingeniería de petróleos de la Universidad Industrial de Santander” (Ballesteros, B. Sánchez, W. 2021), se decide continuar con la investigación enfocándola en las técnicas aplicadas en el área de producción del pensum de la universidad industrial de Santander, para ello se revisa el marco de referencia que se plantea en su trabajo y tomarlo como guía y así, continuar con el estudio bibliométrico y/o bibliográfico ya enfocado en el área de producción de la Universidad Industrial de Santander.

2.1. Ciencia de datos

La ciencia de datos o también conocida como data Science se define como “el arte y la ciencia de adquirir conocimiento a través de datos” (Ozdemir, S. 2016). La ciencia de datos trata de como tomar los datos, para luego adquirir conocimientos y en base a ello tomar decisiones, predecir el futuro, comprender el pasado y/o presente, y, para crear nuevas industrias y productos. Es importante señalar que La ciencia de datos no pretende sustituir la capacidad del cerebro humano, sino complementar ideas al colaborar juntamente con él.

La ciencia de datos abarca una amplia gama de campos y disciplinas, como lo son la inteligencia artificial, la minería de datos, estadística, entre otros. Lo anterior mencionado puede evidenciarse en la siguiente figura 2-1.

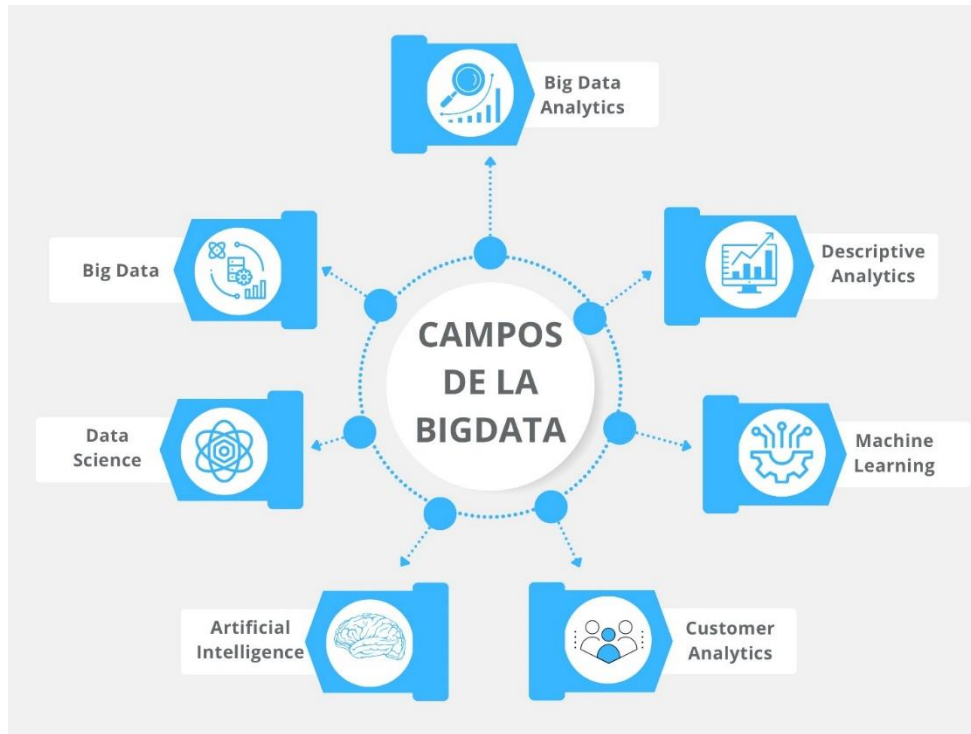


Figura 2.1. Diferentes campos de la ciencia de datos.

2.2. Ciencia de datos e inteligencia artificial

La ciencia de datos se enfoca en el análisis y la interpretación de datos para obtener información valiosa, mientras que la inteligencia artificial se enfoca en la creación de sistemas informáticos capaces de realizar tareas que requieren inteligencia humana. Ambas disciplinas son importantes para la tecnología y la informática, y se utilizan juntas en muchos campos para lograr objetivos específicos. Por lo tanto, en esta sección, se dirigirá la atención hacia el ámbito de la inteligencia artificial con el propósito de ofrecer una perspectiva general de este campo.

2.2.1. Inteligencia artificial

La inteligencia artificial IA (o artificial intelligence) es una realidad múltiple y compleja. Una definición concisa de IA es la que ofrecen (Keith Frankish & William M. Ramsey, 2014), donde indican que la IA es un acercamiento interdisciplinario para comprender, modelar y

replicar los procesos inteligentes y cognitivos a partir de diversos principios y dispositivos computacionales, matemáticos, lógicos, mecánicos e incluso biológicos. Esto con el fin de buscar que la inteligencia artificial pueda desarrollar máquinas que sea capaces de hacer las cosas que hacen las mentes humanas.

2.2.2. Aplicaciones de la inteligencia artificial

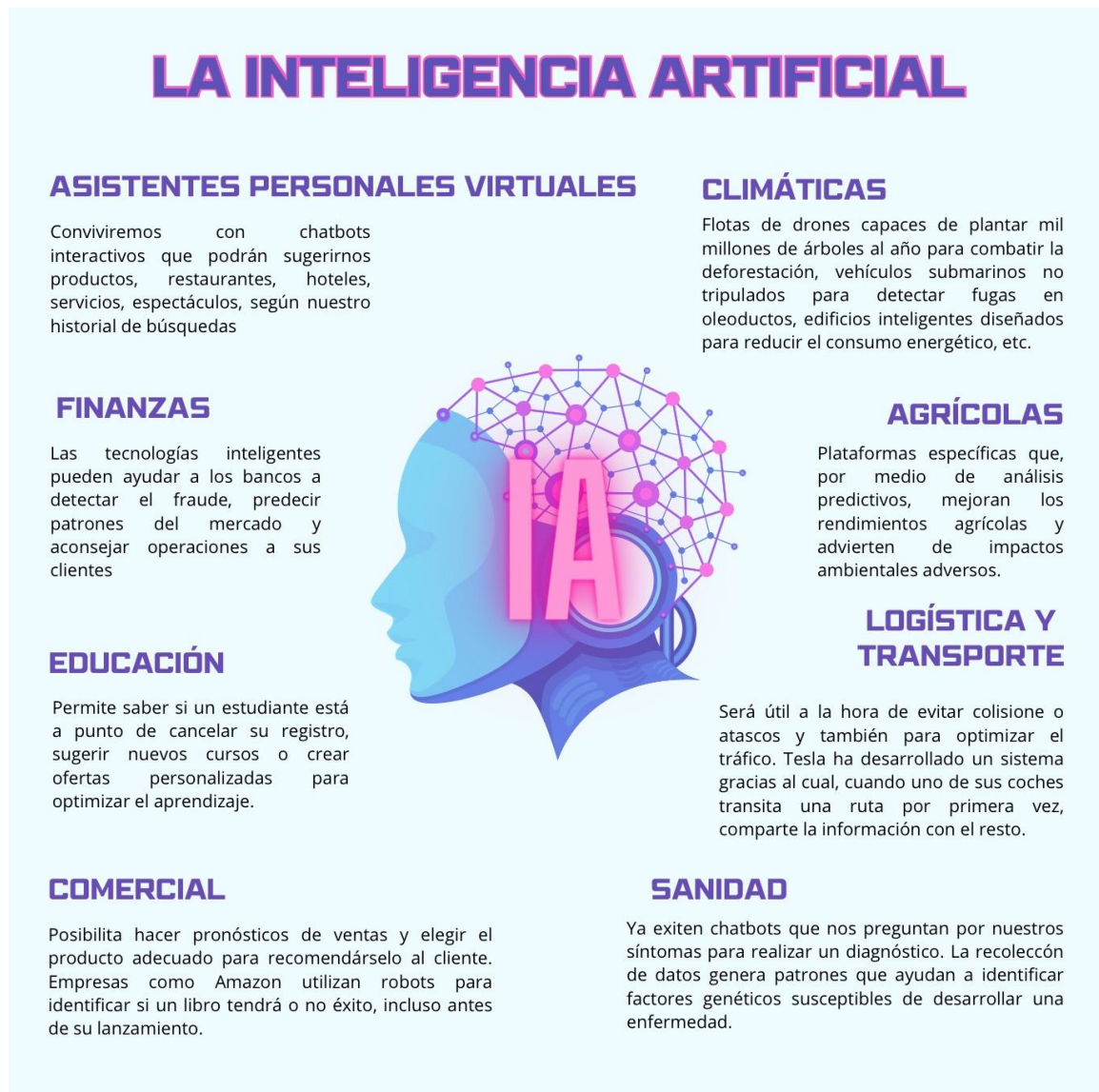


Figura 2.2. Aplicaciones de la inteligencia artificial

Nota: adaptado de Iberdrola <https://acortar.link/a8PmAh>

Existe una gran cantidad de aplicaciones de la IA en el mundo real, donde el papel clave es generar un sistema de cómputo más efectivo, más eficiente y mejor. Esto se debe al aumento de datos almacenados, lo que ha generado una necesidad urgente de enfoques innovadores y herramientas automatizadas capaces de transformar grandes volúmenes de datos en información valiosa y conocimiento inteligente. Algunas de las aplicaciones se representan en la figura 2-2.

Por lo tanto, basándose en el área de producción de hidrocarburos, la inteligencia artificial (IA) puede ayudar con diversas aplicaciones. A continuación, se mencionan algunas de ellas:

Optimización de la producción: la IA se utiliza para analizar grandes cantidades de datos y proporcionar recomendaciones para maximizar la de petróleo. Los algoritmos de aprendizaje automático pueden identificar patrones en los datos y predecir los cambios en las condiciones del pozo para que los ingenieros puedan ajustar los procesos de producción en consecuencia.

Mantenimiento predictivo: Puede ayudar a predecir y prevenir fallas en los equipos de producción. Al monitorear continuamente los datos de los sensores, la IA puede detectar patrones que indiquen posibles fallas en los equipos.

Planificación y programación de pozos: mediante algoritmos de optimización se identifican los mejores lugares para perforar nuevos pozos y la mejor forma de programar trabajos de mantenimiento y reparación.

Análisis de imágenes sísmicas: la IA también se utiliza para analizar imágenes sísmicas para identificar la ubicación de depósitos de petróleo y gas. Los algoritmos de procesamiento de imágenes pueden identificar patrones y características que indiquen la presencia de hidrocarburos en el subsuelo.

2.3. Subconjuntos de la inteligencia artificial

Cuando se habla de las ramas de la inteligencia artificial se debe empezar por aclarar que es una rama de la informática que se ocupa del diseño y la construcción de sistemas que realizan las tareas relacionadas con la inteligencia humana.

La figura 2-3 es una modificación de las principales ramas de la inteligencia artificial que se encontró en el libro *Practical Machine Learning With Python* (Sarkar, D., Bali, R., & Sharma, 2018), donde se agregó nuevos subconjuntos de la IA que se encontraron a partir de la revisión bibliométrica.

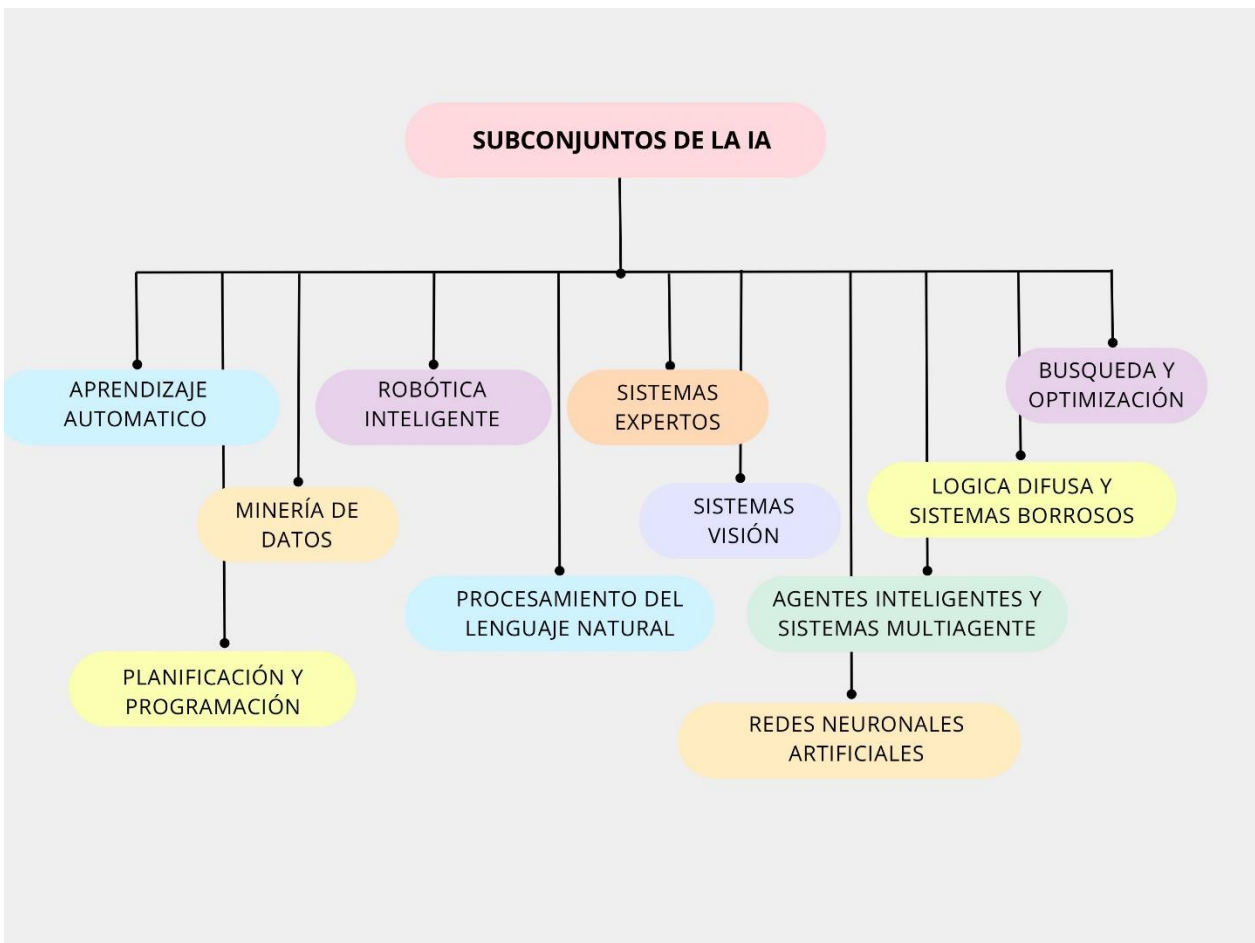


Figura 2.3. Subconjuntos de la inteligencia artificial

2.3.1. Aprendizaje automático ML

El aprendizaje automático (o Machine learning ML) es también conocido como “análisis predictivo” o “aprendizaje estadístico” (Müller & Guido, 2016). Este permite realizar procesos de aprendizaje desde un ordenador sin haber sido previamente programado, por ende, esta técnica se enfoca en procesar grandes volúmenes de datos imposibles de procesar por el cerebro humano.

2.3.2. Minería de datos

La minería de datos hace énfasis en descubrir patrones, modelos y otros tipos de conocimientos en grandes conjuntos de datos, esto consiste en extraer la mayor cantidad de información y conocimiento a partir de datos como: bases de datos, depósitos de datos, la Web, otros repositorios de información o datos que se transmiten dinámicamente al sistema (Han, Kamber, & Pei, 2011). Por ello, la minería de datos es un proceso iterativo en el que se enfoca en el descubrimiento, ya sea mediante métodos manuales o automáticos en los que busca obtener información concreta y mejorar las actividades a analizar.

2.3.3. Robótica inteligente

La robótica es un campo científico interdisciplinario relacionado con el diseño, desarrollo, operación, y evaluación de dispositivos electromecánicos utilizados para realizar tareas que normalmente realizan los humanos. Los robots, por ejemplo, son utilizados en líneas de montaje industrial para realizar tareas repetitivas. Los robots normalmente constan de al menos tres partes: una estructura metálica (a menudo un brazo que ayuda a la ejecución de diferentes labores) que es la que le permite al robot relacionarse con el entorno; sensores que recopilan información sobre las propiedades físicas tales como sonido, temperatura, movimiento y presión; y algún tipo de sistema de procesamiento que transforma datos adquiridos por los sensores en instrucciones sobre qué acciones realizar. (Franceschetti, 2018. p, 250 y p, 265)

2.3.4. Procesamiento del lenguaje natural

El procesamiento de lenguaje natural en sus siglas NLP (Natural Language Processing) Se trata de un campo interdisciplinario que fusiona principios de lingüística computacional, informática y la inteligencia artificial. Su objetivo es facilitar la interacción entre máquinas y entornos naturales en constante evolución. Algunas de las aplicaciones clave del procesamiento de lenguaje natural incluyen la traducción automática, el reconocimiento de voz, sistemas de respuesta a preguntas, identificación y resolución de contexto, resumen de texto, clasificación de contenido textual, extracción de información, análisis de emociones y opiniones, así como segmentación de temas. (Sarker et al., 2018).

2.3.5. Sistemas expertos

Son sistemas basados en programas de computadora que facilitan soluciones de nivel experto a problemas importantes, este es heurístico, transparente y flexible (Buchanan, B. Duda, R. 1983). Los sistemas expertos se están desarrollando en campos que requieran supervisión o experiencia humana considerable para su solución, por ende, la mayoría de los sistemas expertos actuales son capaces de lidiar con áreas problemáticamente restringidas, algunos ejemplos de estas áreas problemáticas son: Diagnostico medico de alguna enfermedad, análisis financiero, diseño de producto, etc. Sin embargo, en áreas restringidas estos sistemas necesitan grandes cantidades de conocimiento para llegar a un rendimiento comparable al de los expertos humanos en el campo (Peter J, F. Lucas y Linda C. Van der Gaag. 1991).

2.3.6. Sistemas de visión

Los sistemas de visión artificial se basan en sensores digitales protegidos dentro de cámaras industriales con ópticas especializadas en adquirir imágenes, para que el hardware y el software puedan procesar, analizar y medir diferentes características para tomar decisiones. Esta

funciona de manera parecida a la visión humana, excepto que la visión humana tiene la ventaja de la experiencia y los contextos aprendidos para diferenciar entre los objetos, qué tan lejos están, si se están moviendo o si hay algo mal en una imagen. De tal forma, en un sistema de visión artificial existen dispositivos con sensores que hacen las veces del ojo humano, tales como son: una cámara, un radar, rayos X o cualquier otro dispositivo que proporciones una escena completa del entorno donde se ejecutara la tarea. Es por eso, que el sistema cuenta con un poderoso algoritmo que imita las funciones del cerebro y se encarga de interpretar y clasificar el contenido de la imagen (Elgendy, M. 2019).

2.3.7. Planificación y programación

La planificación y programación hacen referencia a la realización de estrategias o secuencias de acción, que se usan para la ejecución por agentes inteligentes, robots autónomos y vehículos no tripulados. El objetivo es reducir costos, por ejemplo, en una empresa de preparación o de fabricación, minimizar tiempos, cumplir con el 100% de las tareas, utilización óptima de la capacidad a realizar, etc. Es por ello, que la programación se entiende como la asignación de dichas actividades o acciones a lo largo del tiempo de acuerdo con ciertos criterios de desempeño (Spyropoulos, C. 2000).

2.3.8. Búsqueda y optimización

Los algoritmos de búsqueda desempeñan un papel esencial en diversas aplicaciones de la inteligencia artificial, y una de sus funciones primordiales es hallar rutas para resolver problemas. Por consiguiente, cuando se aborda un problema en los ámbitos de la informática y la inteligencia artificial, suele considerarse como un proceso de exploración a través del conjunto de soluciones posibles. Por otro lado, en ingeniería y matemáticas, se interpreta como un proceso de

optimización, con el propósito de descubrir la mejor solución o la solución óptima para un problema, lo que permite abordar desafíos de manera eficiente y veloz. (Chandel y Sood, 2014)

2.3.8.1. Algoritmos genéticos (GA)

Los algoritmos genéticos (o genetic algorithm según sus siglas en inglés), son una técnica de optimización y búsqueda inspirada en la evolución biológica, donde opera en función del tamaño de población fijo llamado individuos que evoluciona con el tiempo (Binu, D. Rajakumar, B. 2021). Los algoritmos genéticos usan tres operadores genéticos: operador de cruce, operador de mutación y operador de selección, donde el individuo más fuerte de la población puede crear descendencia. Los algoritmos genéticos se implementan como la optimización y el procedimiento de búsqueda computarizado basado en el principio de la selección y la genética natural.

2.3.8.2. Programación genética GP

Esta es una variante de los algoritmos genéticos, que realizan una tarea específica en los programas de computadora. La programación genética o GP funciona con una “población inicial generada aleatoriamente, en este caso, los objetos matemáticos son programas de computadora compuestos de funciones y terminales apropiadas para el dominio del problema” (Mejia, P. 2019).

Por ende, en los algoritmos genéticos (AG) y en la programación genética (PG), cada individuo (programa de computadora) dentro de la población se evalúa mediante su valor de aptitud, es decir, la eficacia en el entorno del problema específico. La forma de medir la aptitud de un individuo, o función de aptitud. Esto varía según el problema en cuestión.

En conclusión, los GP resuelve los problemas en los siguientes 3 pasos:

- Crean una población inicial de programas de manera aleatoria.

- Llevan a cabo de manera iterativa los siguientes subpasos hasta que se cumpla el criterio de finalización, es decir que para cada programa se debe calcular el valor de fitness.
- Como resultado de la programación genética, se identifica el mejor programa que surgió en cualquier generación. Este resultado puede representar una solución, ya sea precisa o aproximada, al problema en cuestión.

2.3.9. Redes neuronales artificiales

Las redes neuronales artificiales (Artificial Neural Networks, ANN) según (Gharbi, R. Mansoori, G. 2005) son un conjunto interconectado de elementos, unidades o neuronas simples de procesamiento, cuya funcionalidad se basa libremente en la neurona del cerebro. La capacidad de procesamiento de la red se almacena en las fuerzas de conexión entre unidades, o pesos, obtenidos mediante un proceso de adaptación o aprendizaje de un conjunto de patrones de entrenamiento.

2.3.10. Agentes inteligentes y sistemas multiagente

Los agentes inteligentes y los sistemas multiagentes son conceptos de intercomunicación dentro de la inteligencia artificial. Es por ello, que un agente inteligente se define como una entidad autónoma con la capacidad de percibir su entorno y realizar acciones que permitan lograr objetivos específicos para lo que se ha programado. Mientras que, un sistema multiagente es un conjunto de dos o más agentes inteligentes que interactúan entre sí para lograr un objetivo común. Dicho lo anterior, los sistemas multiagentes son una forma de coordinación de múltiples agentes inteligentes, que permite trabajar en colaboración para así resolver problemas más complejos y alcanzar objetivos compartidos.

2.3.11. Lógica difusa

La lógica difusa (Fuzzy logic FL) “inventada en 1964, es un enfoque del razonamiento donde las reglas de inferencia son aproximadas en lugar de exactas. Es útil para manipular información incompleta, imprecisa o poco confiable” (Gharbi, R. Mansoori, G. 2005).

Los sistemas de lógica difusa también permiten combinar la información proveniente de los datos registrados con la del conocimiento humano experto, esto hace que la lógica difusa sea atractiva en la industria petrolera, siendo este sector muy complejo y con mucha incertidumbre en sus operaciones cuyo funcionamiento depende en gran medida de la habilidad del personal en cada operación.

2.4. Tipos de inteligencia artificial

La inteligencia artificial se puede organizar de distintas formas, ya sea, según su etapa de desarrollo o la acción que puede realizar la máquina. Es por eso que (Mueller y Massaron, 2018) categorizaron la inteligencia artificial en cuatro grupos que se pueden observar en la figura 2.4.

2.5. Técnicas de la ciencia de datos

La ciencia de datos es el estudio de los datos para adquirir conocimientos e información relevante sobre un sistema, utilizando métodos científicos avanzados, estadísticas, análisis y algoritmos para extraer conocimiento. Así mismo, genera un desarrollo en los procesos de gestión de base de datos, análisis estadísticos y analíticos. Los científicos de datos determinan las técnicas de administración de datos y bases de datos más apropiadas para almacenar, acceder, procesar y analizar los datos. Con base en lo anterior, se hizo una revisión bibliométrica que identificó las técnicas más usadas en la ingeniería de producción.

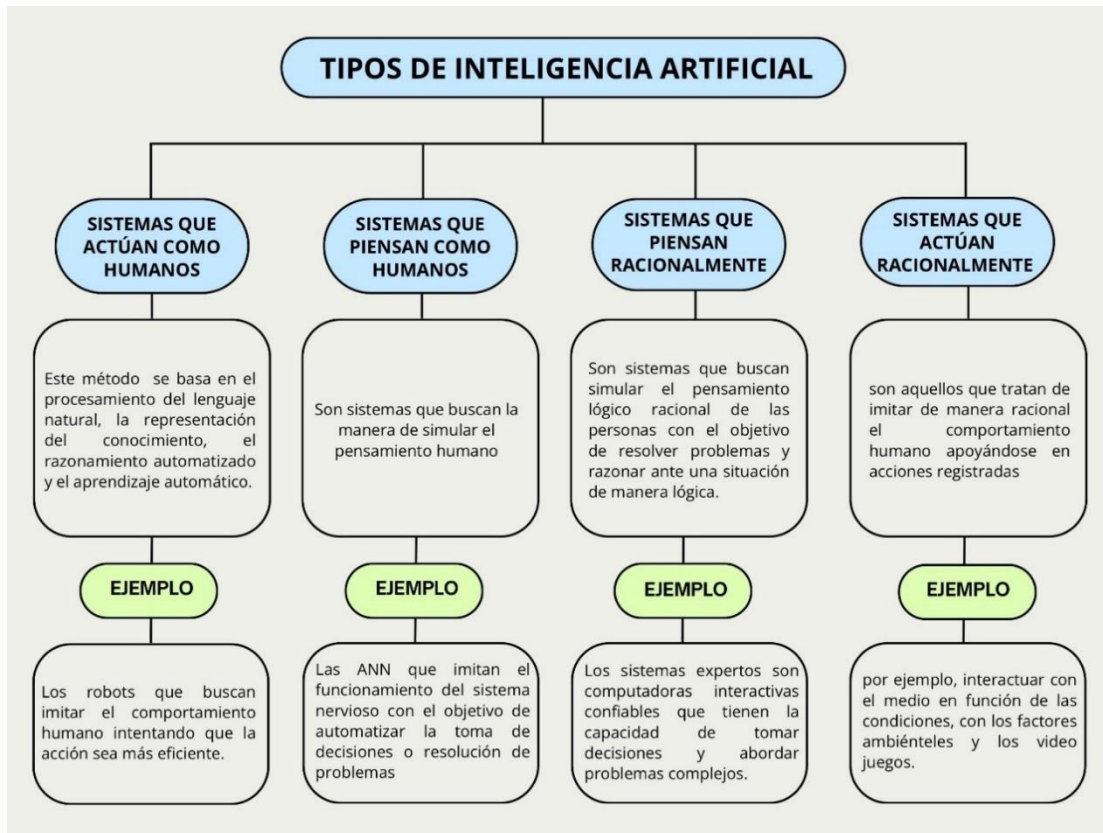


Figura 2.4 Tipos de inteligencia artificial

2.5.1. Big data

Big data proviene de dos fuentes, tradicional y alternativa, y se presenta en un formato estructurado o no estructurado. Los datos tradicionales consisten en datos que están fácilmente disponibles o que pueden generarse. Los datos no tradicionales son datos que se recopilan de fuentes como escáneres, sensores y satélites, mientras que los datos alternativos provienen de sitios web, redes sociales, datos personales, publicaciones en blogs, correos electrónicos, etc., estos datos consisten en datos que pueden derivarse de fuentes de datos tradicionales y no tradicionales. De hecho, los datos no tradicionales son casi ilimitados. Es por ello, que en la Big Data se debe tener en cuenta las siguientes características (ver figura 2-5)

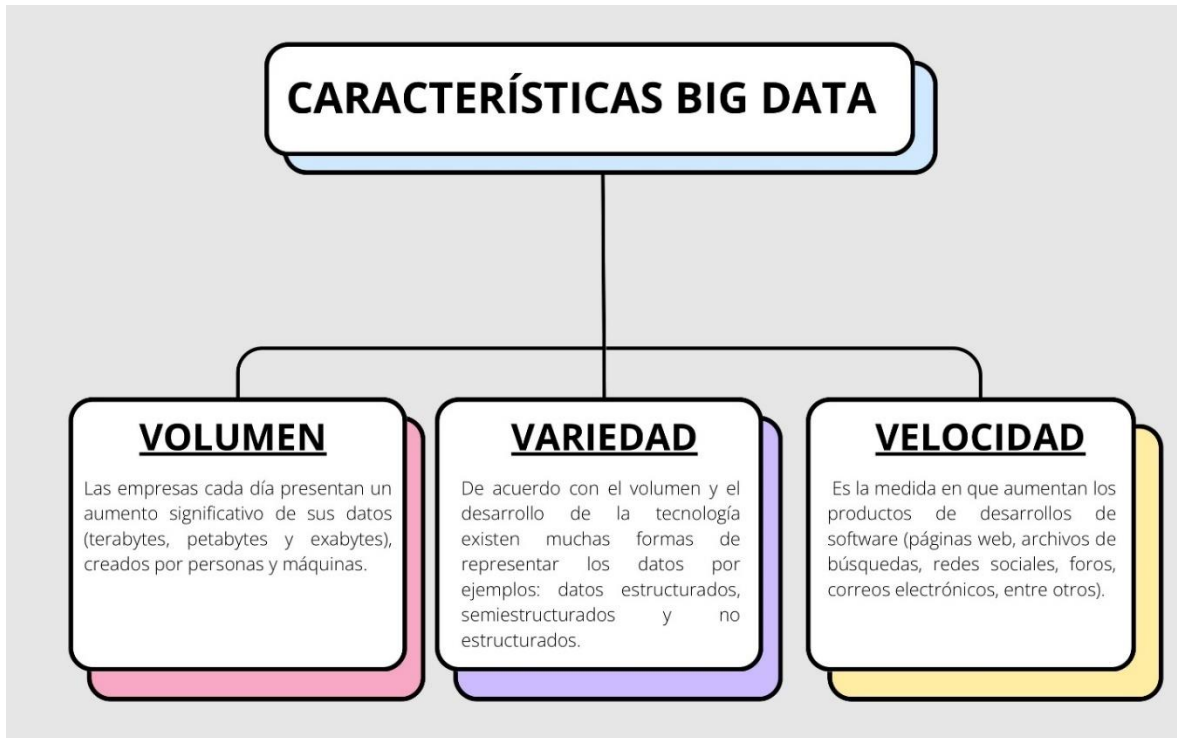


Figura 2.5. Características de la Big Data

En la Big Data el Modelo de datos es una representación conceptual de la información que se puede almacenar a partir de una base de datos, esta describe la estructura lógica de los datos y las relaciones entre ellos. Existen diferentes tipos de datos, como lo son:

- **Datos estructurados**

Este tipo de datos se dividen en estáticos (array, cadena de caracteres y registros) y dinámicos (listas, pilas, colas, árboles y archivos). Se puede definir que los datos estructurados son aquellos de mayor facilidad para acceder, pues tienen una estructura bien especificada.

- **Datos semiestructurados**

Estos datos no tienen un formato definido, lo que tienen son etiquetas que facilitan separar un dato de otro. Un dato de estos se lee con un conjunto de reglas de cierto nivel de complejidad.

- **Datos no estructurados**

No se pueden normalizar, no tienen tipos definidos ni organizados bajo algún patrón, ni almacenados de manera relacional, ni con base jerárquica de datos, ya que no son un tipo de dato predefinido. Sin embargo, los datos deben poder ser organizados, clasificados, almacenados, eliminados, buscados de alguna forma.

2.5.2. Minería de datos

La minería de datos (data mining) consiste en hallar patrones y correlaciones de grandes conjuntos de datos, hoy en día existe una necesidad inminente de convertir dichos datos en información útil. La minería de datos utiliza varios métodos, como estadísticas, inteligencia artificial, sistemas de bases de datos y métodos de aprendizaje automático (D, Binu, B, R, Rajakumar. 2021). Los métodos empleados para la minería de datos se clasifican en las siguientes categorías (ver figura 2-6).

2.5.3. Lógica difusa

La lógica difusa (Fuzzy logic FL) se usa comúnmente para modelar sistemas complejos e imprecisos usando una estrategia simple del principio si-entonces, donde “si” especifica el vector explicativo difuso, y “entonces” especifica la consecuencia del conjunto difuso. (D, Binu, B, R, Rajakumar. 2021. p 135),

Algunas de las características de un sistema difuso son:

- Utiliza variables lingüísticas: un sistema difuso utiliza términos lingüísticos para describir las entradas y salidas del sistema.
- Toma decisiones basadas en la incertidumbre: Permite tomar decisiones cuando la información es incierta o incompleta.

- Utiliza reglas difusas: Refuerzan la relación entre las variables lingüísticas de entrada y salida.
- Es adaptativo: Ajusta su comportamiento en función de la experiencia previa utilizando reglas difusas y así ajustarlos en función de los resultados obtenidos.
- Permite la combinación de múltiples entradas: Combina varias entradas para tomar decisiones más precisas y completas.

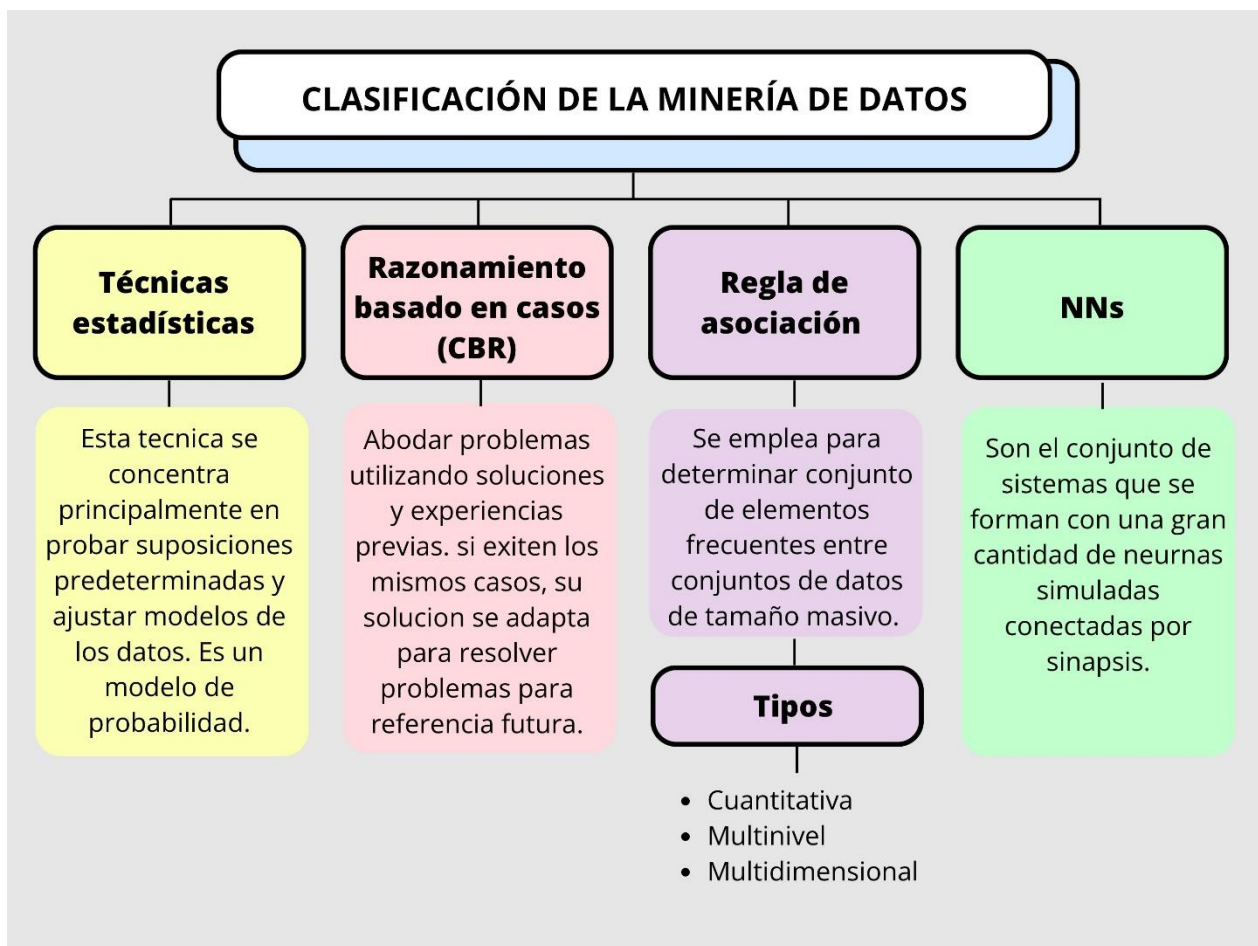


Figura 2.6. Clasificación de la minería de datos

2.5.4. Subconjuntos machine learning

Los principales algoritmos de Machine learning son aprendizaje supervisado y aprendizaje no supervisado. Sin embargo, el machine learning está conformado por más

conjuntos que se utilizan normalmente para el procesamiento y análisis de datos. Estos conjuntos se pueden observar en la siguiente figura 2-7.

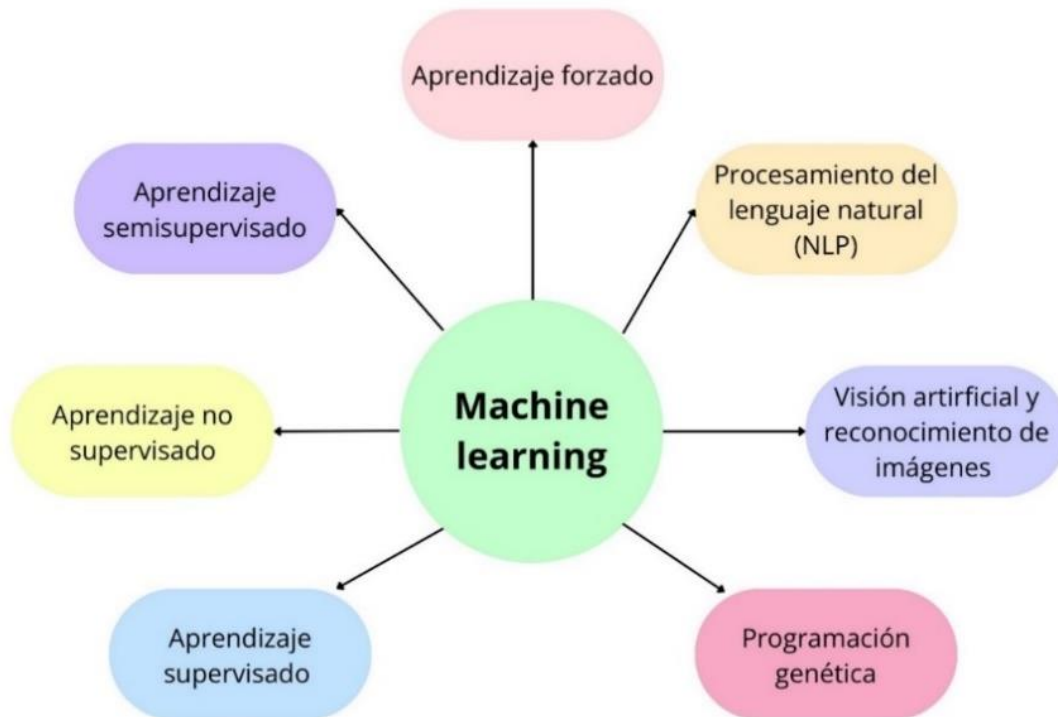


Figura 2.7. Subconjuntos de machine learning

2.5.4.1. Aprendizaje supervisado

El aprendizaje automático lee los datos de entrada y luego clasifica los datos en grupos preespecificados, el modelo aprende de muestras de datos cuyos resultados se conocen de antemano. Sin embargo, Géron, A. dice que los algoritmos de aprendizaje supervisado más importantes son: regresión lineal, regresión logística, máquinas de vectores de soporte, árboles de decisión y redes neuronales, en la figura 2-8 se pueden evidenciar varios tipos de algoritmos.

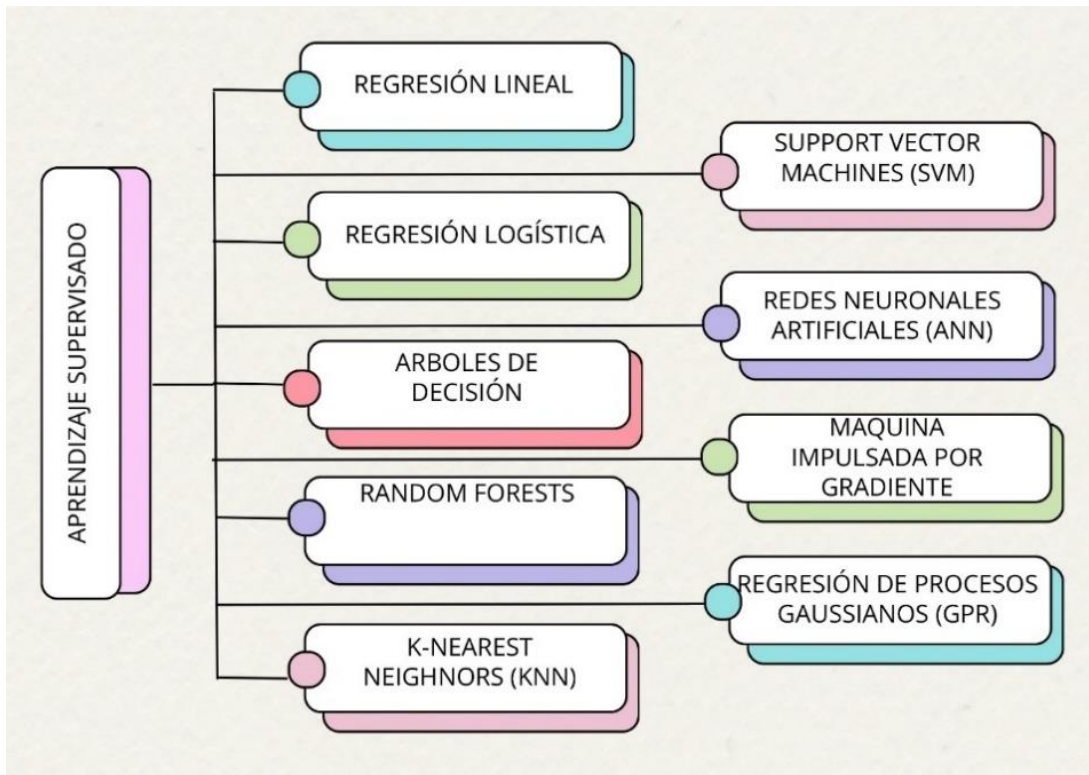


Figura 2.8. Tipos de algoritmos del aprendizaje supervisado

2.5.4.2. Aprendizaje no supervisado

El aprendizaje no supervisado consiste en clasificar los datos de entrada en grupos y subgrupos en función de atributos comunes, sin disponer de ninguna ayuda o supervisión. Ballesteros, B. Sánchez, W. 2021, mencionan que este método de aprendizaje se enfoca con extraer información o conocimientos útiles de los datos antes que intentar predecir los resultados basándose en datos de entrenamiento supervisado previamente disponibles”. Sin embargo, Géron, A. dice que los algoritmos de aprendizaje no supervisado más importantes son: agrupación (Clustering), visualización y reducción de dimensionalidad y aprendizaje de reglas de asociación. Sin embargo, en la figura 2-9 se pueden observar otros tipos de algoritmos.

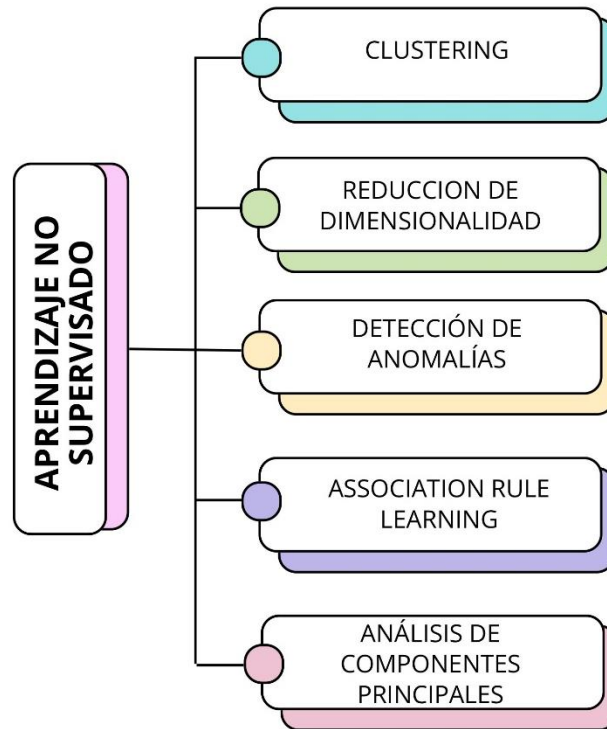


Figura 2.9. Tipos de algoritmos de Aprendizaje no supervisado

2.5.4.3. Aprendizaje semisupervisado

Es una combinación de algoritmos supervisados y no supervisados. Donde el aprendizaje semisupervisado puede manejar datos etiquetados (datos de entrada con su respectivo resultado) y datos no etiquetados (datos de entrada que no poseen un resultado conocido). Un ejemplo de este aprendizaje es el funcionamiento de Google fotos, una vez se cargan todas las fotos de su familia, este automáticamente reconoce el rostro de las personas.

2.5.4.4. Aprendizaje forzado

Se enfoca en observar el entorno, seleccionar, realizar acciones y obtiene recompensa. Es decir, busca estrategias y analiza las acciones para seleccionar cual es la mejor estrategia y así, obtener la máxima recompensa.

2.5.4.5. Procesamiento del lenguaje natural (NLP)

Es un campo multidisciplinario que combina conceptos de lingüística computacional, informática e inteligencia artificial. El objetivo principal es permitir interacciones entre las máquinas y entornos naturales que han venido evolucionando con el tiempo. Algunas de las principales aplicaciones del procesamiento de lenguaje son: traducción automática, reconocimiento de voz, sistemas de respuesta a preguntas, reconocimiento y resolución de contexto, resumen de texto, categorización de texto, extracción de información, análisis de sentimientos y emociones y segmentación de temas. (Sarkar, et al., 2018. p 40).

2.5.4.6. Visión artificial y reconocimiento de imágenes

Esta técnica se enfoca en el reconocimiento de imágenes y herramientas que ayudan a identificar y analizar grandes conjuntos de datos de imágenes, tales como fotografías o videos, y así extraer información útil a partir de ellas.

2.5.4.7. Programación genética

Es una técnica de programación basada en algoritmos evolutivos para crear automáticamente programas que resuelvan una tarea determinada. Esta técnica utiliza la teoría de la evolución de Charles Darwin buscando imitar el proceso de selección natural para así encontrar soluciones a problemas complejos. En otras palabras, "La Programación Genética es una técnica de programación evolutiva que permite diseñar programas de forma automática, estableciendo un proceso de evolución mediante selección natural, mutación y cruza." (Montes, M. 2007)

2.5.5. Algoritmos de machine learning utilizados en los sistemas de producción de la industria petrolera

2.5.5.1. Redes neuronales artificiales.

Las redes neuronales artificiales (artificial neural network, ANN) o una red neuronal (neural network, NNET) son un tipo de algoritmo de aprendizaje automático supervisado que se utiliza para la clasificación y la regresión. Son una de las ramas más destacadas del campo científico de la inteligencia artificial, las ANN “tratan de emular el comportamiento del cerebro humano, caracterizado por el aprendizaje a través de la experiencia y la extracción de conocimiento genérico a partir de un conjunto de datos” (Lopez, R. Fernández, J. 2008).

La neurona Biológica es un elemento fundamental en los sistemas neuronales biológicos de los seres humanos. Una neurona, consta de un cuerpo celular del que parte una rama principal o axón y un denso árbol de ramificaciones más cortas (árbol detríptico) como se muestra en la figura 10a. Desde el punto de vista funcional, las neuronas constituyen un canal de recepción de información que son las dendritas, estas reciben señales de entrada (inputs) que proceden de otras células. Un órgano de cómputo que hace referencia al cuerpo celular y su función es combinar e integrar los inputs recibidos, finalmente, la neurona está constituida por un canal de salida que se refiere al axón, este envía la salida generada por el cuerpo celular a otras neuronas.

En la figura 10b se muestra una neurona artificial con sus características más relevantes, por ejemplo, la conexión de entrada está formada por el axón de una neurona y las dendritas de otra que reciben el nombre de pesos sinápticos (W_{ij}), estos determinan la fuerza y tipo de relación entre las dos neuronas que están conectadas, la sinapsis es unidireccional, en cuanto a la unidad de salida, es la función que contiene el valor sumado de las entradas (z_i), en esta unidad se

aplica una función de activación para poder obtener una señal y esta sea transmitida a la siguiente neurona.

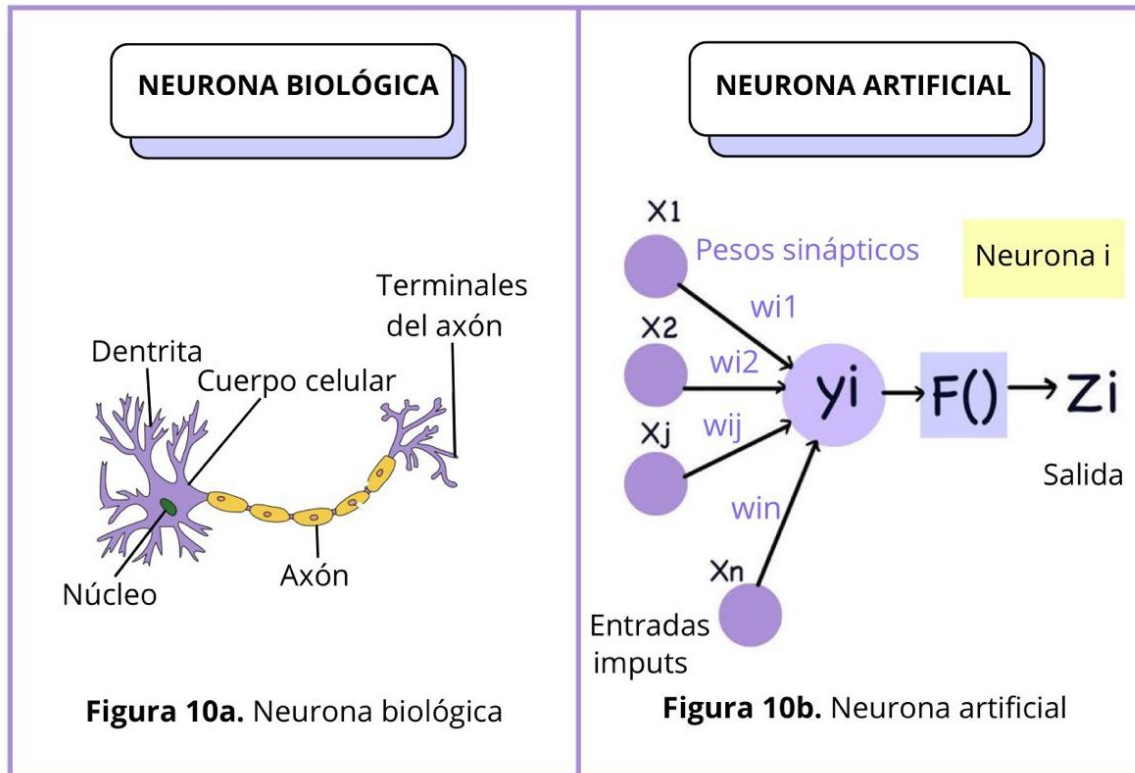


Figura 2.10. Neurona biológica y neurona artificial

Según lo mencionado anteriormente, la función de activación es de gran importancia en la red neuronal artificial puesto que se encarga de decidir si se activa o no una neurona. “Específicamente, la función de activación tiene el cometido de evaluar la salida de cada neurona y decidir si ese valor de salida debe ser considerado o conectado a las demás neuronas” (Ballesteros, B. Sánchez, W. 2021). Estas son algunas de las funciones de activación más empleadas (Nacelle, A. 2009):

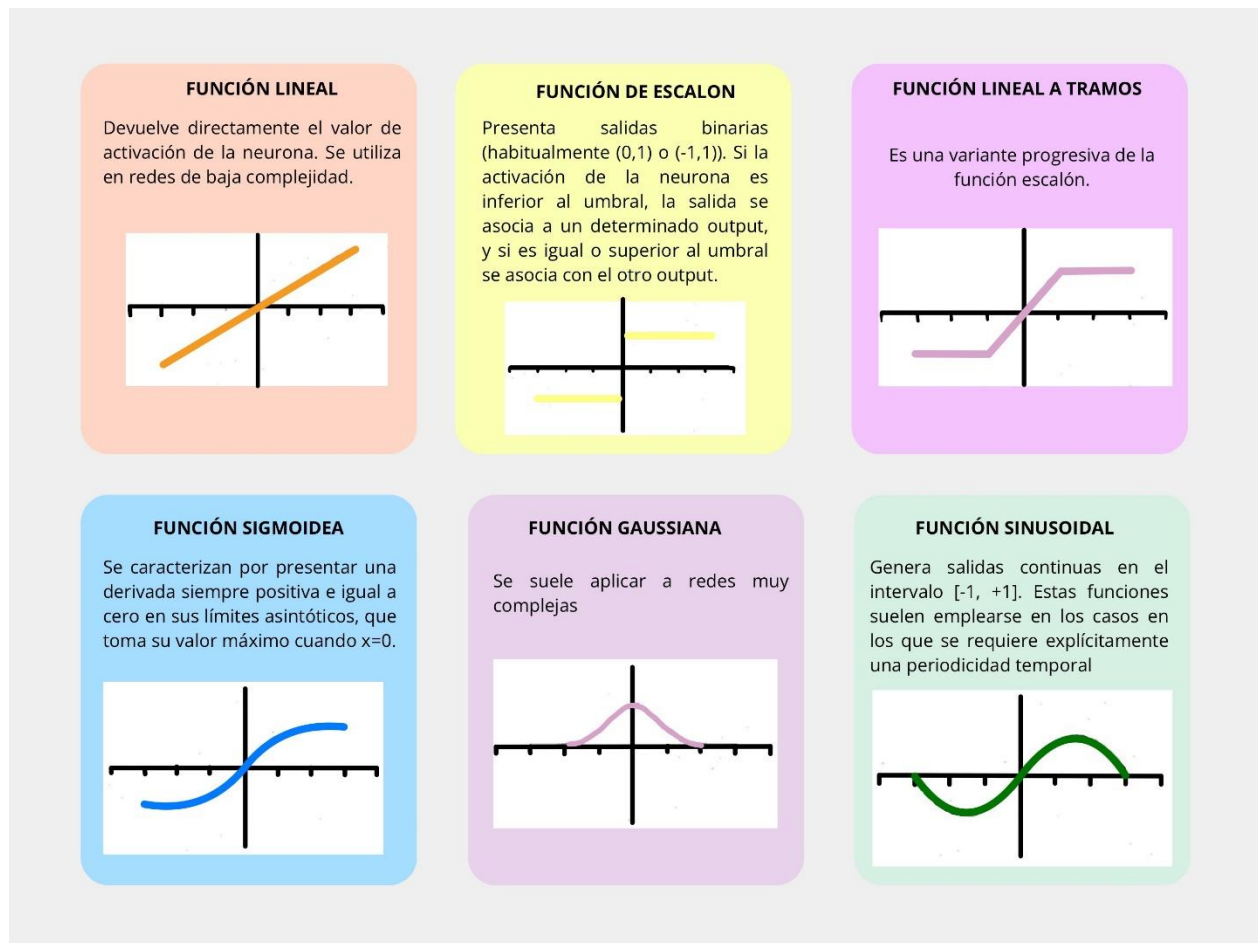


Figura 2.11. Funciones de activación

- **Estructura de una red neuronal artificial**

Por otro lado, una red neuronal artificial consta de una estructura que se basa en interconectar las neuronas de diferentes formas como se muestra en la figura 2-12, teniendo en cuenta una capa de entrada, una capa de salida y una o más capas ocultas.

- **Entrenamiento o aprendizaje**

Por otra parte, el aprendizaje de una red neuronal busca ajustar la arquitectura y los parámetros mediante algoritmos de aprendizaje supervisado o no supervisado para minimizar alguna función de error que indique el grado de ajuste de los datos.

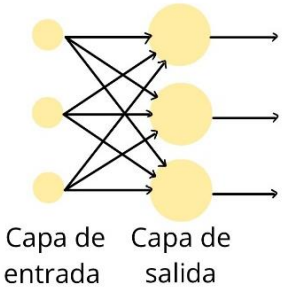
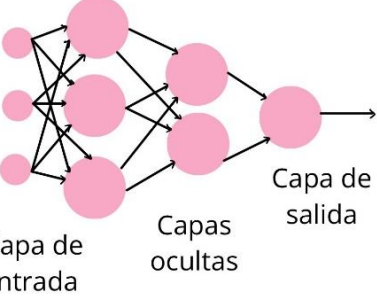
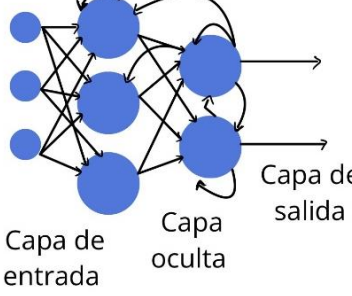
Figura 12a	Figura 12b	Figura 12c
 <p>Capa de entrada Capa de salida</p>	 <p>Capa de entrada Capas ocultas Capa de salida</p>	 <p>Capa de entrada Capa oculta Capa de salida</p>
Red monocapa	Red multicapa	Red recurrente
<p>Es la red neuronal mas sencilla ya que tiene una capa de neuronas que proyectan las entradas a una capa de neuronas de salida donde se realizan los diferentes cálculos.</p>	<p>Es similar a la red monocapa a diferencia que existe un conjunto de capas intermedias entre la capa de entrada y la de salida (llamadas capas ocultas) y esta red puede estar total o parcialmente conectada.</p>	<p>Es una red que es diferente a la monocapa y a la multicapa donde esta busca la existencia de lazos de realimentación en la red y estos lazos pueden ser entre neuronas de diferentes capas, neuronas de la misma capa o, entre una misma neurona.</p>

Figura 2.12. Arquitecturas de las redes neuronales artificiales. 12.a. Red monocapa. 12.b. Red multicapa. 12.c. Red recurrente.

Nota: Foto adaptada de Nacelle, A. 2009

El aprendizaje supervisado consiste en entrenar la red a partir de un conjunto de datos compuestos por patrones de entrada y salida. El objetivo del algoritmo de aprendizaje es lograr un ajuste de los pesos de manera tal que la salida generada por la ANN sea lo más cercano posible a la verdadera salida, dada una cierta entrada (Salas, R). Y en el aprendizaje no supervisado se presenta solo un conjunto de patrones a la ANN con el objetivo de ajustar el algoritmo de aprendizaje a los pesos de la red de tal manera que esta encuentre alguna estructura presente en los datos.

- **Arquitecturas neuronales**

Las redes neuronales constan de diversos tipos y de acuerdo con (Geron, 2019) algunas de las arquitecturas neuronales más usadas son las redes neuronales feedforward, las redes neuronales recurrentes y por ultimo las redes neuronales convolucionales. Así mismo, en la tabla 2.1 se evidencia un resumen de la topología, la estrategia de aprendizaje y la función de las ANN según (Yan, Y. 2018).

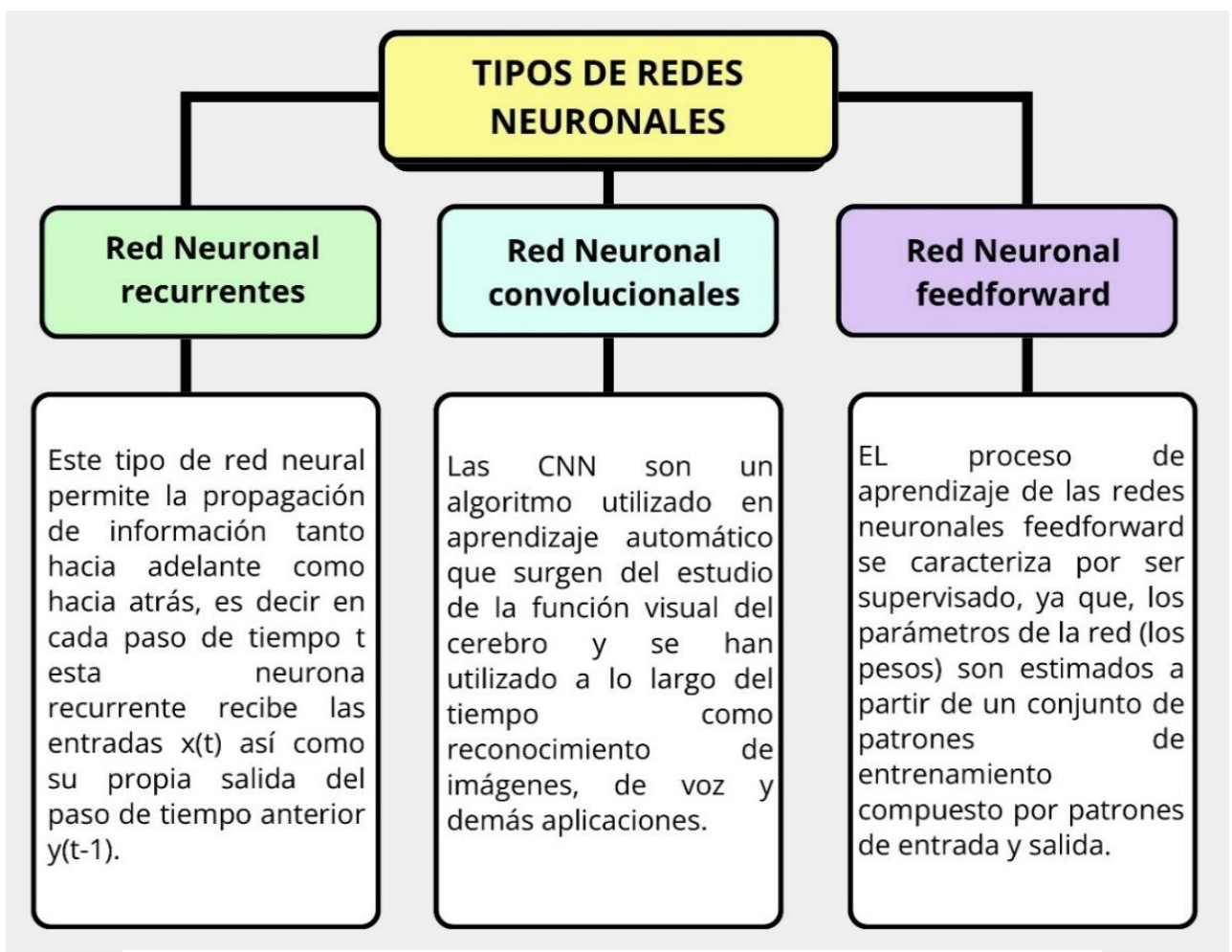


Figura 2.13. Tipos de redes neuronales

Funciones de la ANN:

ANN	Topología ingles	Topologia español	estrategia de aprendizaje	Funcion
MLP	Multilayer feedforward	avance multicapa	Supervisado	Reconocimiento de patrones, aproximación de funciones, modelado y control, clasificación
RBF	Three-layer feedforward	Avance de tres capas	Supervisado	Aproximación de funciones, clasificación
WNN	Three-layer feedforward	Avance de tres capas	Supervisado	Pronóstico, clasificación, aproximación de funciones
ELM	Three-layer feedforward	Avance de tres capas	Supervisado	Función Aproximación, clasificación
Elman	Recurrent	Recurrente	Supervisado	Previsión de series temporales, reconocimiento de patrones
Hopfield	Recurrent	Recurrente	Supervisado	Reconocimiento de patrones, memoria asociativa, optimización, procesamiento de imágenes
Kohonen	Single layer	Una sola capa	No supervisado	Reconocimiento de patrones, memoria asociativa, clasificación
PNN	Four-layered feedforward	Avance de cuatro capas	Supervisado	Reconocimiento de patrones, clasificación
CNN	Multilayer	Multicapa	Supervisado	optimización, clasificación
ART	Recurrent	Recurrente	No supervisado	optimización, clasificación
CMAC	Multilayer	Multicapa	Supervisado	aproximación, modelado y control de funciones
CM	Multiple NNs (experts)	NN múltiples (expertas)	Supervisado	reconocimiento de patrones, aproximación de funciones, modelado y control, clasificación

Tabla 2.1. Funciones de la ANN

2.5.5.2. Máquinas de vectores de soporte (SVM).

Las máquinas de vectores de soporte (o Support vector machines en inglés) son “una técnica útil implementada en la regresión y clasificación de datos al mapearlos en espacios futuros inconexos tanto en el dominio de entrada en problemas lineales como en dominios de características multidimensionales para problemas de clasificación y regresión no lineal (Bahaloo, S., et al. 2022). El enfoque de SVM es poder encontrar un hiperplano de un espacio de alta dimensión y este pueda desglosar de manera eficiente las dos clases de los diferentes datos.

Es decir, los hiperplanos que dividen diferentes clases de datos se calculan de tal forma que puedan maximizar la distancia entre los puntos de datos más cercanos de cada clase. Cuando el hiperplano de división no es lineal en el dominio de entrada inicial, es posible generar hiperplanos linealizados en el dominio de características multidimensionales mediante la aplicación de funciones específicas conocidas como núcleos al dominio de entrada.

2.5.5.3. Regresión de procesos de Gaussianos

Es una técnica de regresión o interpolación no lineal que modela los nuevos valores estimados (interpolados) derivados del proceso gaussiano determinado por una función de covarianza (Shadravan, A. 2015). Algunas de sus ventajas son la capacidad de realizar predicciones probabilísticas, ya que se pueden calcular intervalos de confianza empíricas y decidir en base a ellos si se debe reajustar la predicción en alguna región de interés, por otro lado, la regresión de procesos gaussianos es versátil ya que pueden especificar diferentes núcleos y proporcionarlos en núcleos comunes y en núcleos personalizados (Rasmussen, C., et al. 2006).

2.5.5.4. Análisis de componentes principales.

Análisis de componentes principales (Principal Components Analysis, PCA) Este tipo de algoritmo minimiza el error que ocurre al proyectar los datos en una dimensión. Así mismo, puede utilizarse para acelerar otros algoritmos al disminuir el tamaño de los vectores de muestra (Ballester, B. Sánchez, W. 2021).

2.5.5.5. Regresión de mínimos cuadrados parciales

La regresión de mínimos cuadrados parciales (PLSR, Partial least squares regression) son una técnica de modelado estadístico utilizada para construir relaciones entre un conjunto de variables predictoras (X) y una o varias variables de respuesta (Y), mediante la identificación de

combinaciones de variables predictoras que expliquen la mayor cantidad posible de la varianza en las variables respuesta. ‘El modelo de regresión PLS busca específicamente predecir variables dependientes, por lo que, en opinión de diversos autores, resulta de especial utilidad para propósitos de análisis causal predictivo en situaciones de alta complejidad con poco desarrollo teórico’ (Asensi, M., et al. 2014.).

2.5.6. Otros conceptos relacionados con la ciencia de datos

2.5.6.1. Internet de las cosas.

El internet de las cosas (IoT, internet of things) es un término utilizado para describir la interconexión de varios dispositivos y sensores a través de Internet, lo que permite el intercambio de datos y la comunicación entre ellos. Esta herramienta permite crear un sistema de almacenamiento, procesamiento y comunicación entre los procesos físicos de la empresa o sistemas electrónicos que están conectados entre sí; mediante sensores, el IoT permite recolectar datos de una manera más sencilla por ejemplo en un pozo con altas presiones se puede integrar sistemas de sensores que envíen alertas cuando estas presiones varíen. Por tal motivo, el IOT busca la forma de permitir que los objetos de la vida cotidiana se puedan conectar a internet para que creen un base de datos de información que ayuden a mejorar la interacción del mundo virtual con el mundo físico (Cirani et al., 2018).

2.5.6.2. Sistemas o computación inteligentes.

La utilización de técnicas de inteligencia artificial y aprendizaje automático en el ámbito de la ingeniería petrolera se conoce como computación inteligente, permitiendo a los profesionales de la industria analizar grandes volúmenes de datos y obtener información relevante en tiempo real. Por ejemplo, en los pozos de petróleo y en los equipos de producción se

recopilan una gran cantidad de datos que pueden ser procesados con la ayuda de la computación inteligente para monitorear la producción, pronosticar la producción futura, regular la presión del pozo y prevenir cualquier problema en la producción antes de que pueda agravarse (Sheremetov, L., et al. 2005).

2.5.6.3. Campos inteligentes

En la actualidad, durante casi dos décadas, la industria petrolera ha estado aplicando soluciones tecnológicas denominadas “pozos inteligentes” y “campos inteligentes” con el propósito de optimizar la gestión y supervisión de los procesos de extracción de petróleo, con el objetivo de mejorar la eficiencia operativa (Redutskiy, Y. 2017). Los beneficios de los campos inteligentes se originan en la capacidad de realizar mediciones detalladas de la presión, temperatura, caudal y composición del flujo, utilizando sensores especiales para obtener imágenes geofísicas y observar el comportamiento de los fluidos y los cambios relevantes de propiedad. Se logra un control completo de las operaciones a través del uso de válvulas de control de entrada y bombas de frecuencia variable, permitiendo realizar ajustes o cierres independientes de cualquier segmento del pozo si es necesario.

2.5.6.4. Gemelos digitales.

Es una técnica que utiliza análisis de datos, visualización y simulación de algoritmos para proporcionar información útil en la identificación de fallas y el diagnóstico de equipos. Baker Hughes ha presentado un modelo para el gemelo digital en la gestión del rendimiento de los activos, el cual incluye diversos elementos como la identificación y predicción del modo de falla, las recomendaciones de estrategia, la evaluación del estado de los activos, el costo del ciclo de vida y la medición del rendimiento. Esto demuestra cómo el gemelo digital es capaz de mejorar la eficiencia y el rendimiento en la gestión de activos a través de su aplicación en la industria. En

pocas palabras, es un proceso de simulación que utiliza modelos físicos y sensores para adquirir datos y completar el mapeo en el espacio virtual para reflejar el proceso del ciclo de vida de la entidad correspondiente. La figura 2-14 es un modelo gemelo digital del proceso de producción

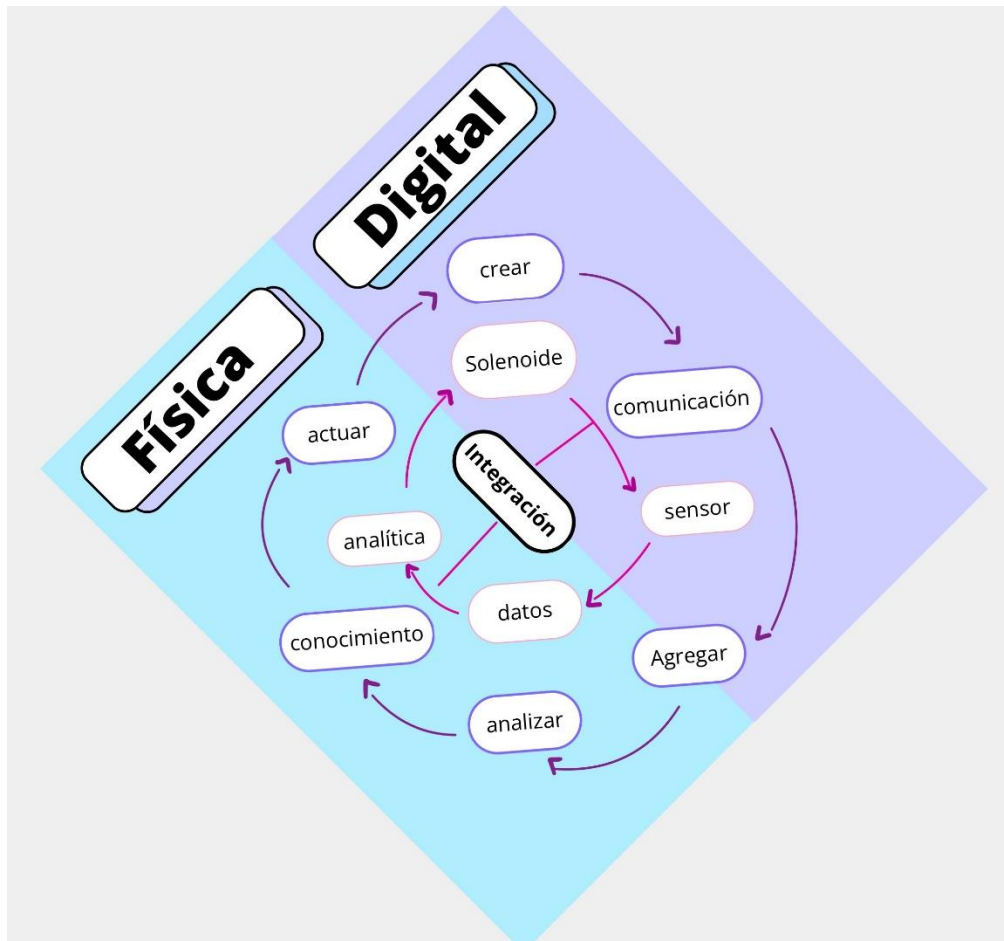


Figura 2.14. Modelo digital del proceso de producción, modelo cíclico

Nota: adaptado de (Parrot y Warshaw. 2017).

2.5.6.5. Digital oil and gas.

El Campo Digital Petrolero tiene como objetivo principal maximizar la recuperación de hidrocarburos, reducir el tiempo improductivo y aumentar la rentabilidad mediante el diseño e implementación de flujos de trabajo integrados que combinan la gestión de procesos empresariales con tecnología avanzada de información y experiencia en ingeniería. Este enfoque

permite agilizar y, en muchos casos, automatizar la ejecución de tareas que normalmente son realizadas por equipos multidisciplinarios. De esta manera, se busca mejorar la eficiencia y la productividad en la industria del petróleo y gas a través de la implementación de soluciones digitales y tecnológicas innovadoras (Parabavire, O. 2018).

2.5.6.6. Tecnología de comunicación inalámbrica.

Las tecnologías de comunicación inalámbrica son aquellas que permiten la transmisión de información entre dispositivos sin necesidad de cables físicos. Estas tecnologías se basan en ondas electromagnéticas para la transmisión de datos, voz, imágenes y vídeo a través del aire. En la industria petrolera esta se puede usar en el desarrollo de petróleo y gas, la tecnología inalámbrica se puede usar para monitorear el cabezal del pozo; en el transporte por tuberías, la red de tuberías se puede conectar de forma remota; En varias estaciones de petróleo y gas, la tecnología inalámbrica se puede utilizar para transmitir de manera eficiente la temperatura, la presión y otros datos recopilados a la sala de control. Además, con el desarrollo de IIoT en la industria del petróleo y el gas, aumenta la demanda de sensores inteligentes e inalámbricos (Hongfang Lu., et al. 2019)

- Algunas de las tecnologías de comunicación inalámbrica más comunes incluyen:
- Wi-Fi: Permite la conexión a Internet y la transmisión de datos entre dispositivos a través de redes inalámbricas.
- Bluetooth: Permite la comunicación inalámbrica entre dispositivos cercanos, como auriculares, altavoces, teclados y ratones.
- NFC: Permite la comunicación inalámbrica a corta distancia entre dispositivos, como la transferencia de datos entre teléfonos móviles.

- RFID: Se utiliza para la identificación y seguimiento de objetos mediante la comunicación inalámbrica entre un lector y una etiqueta RFID.
- Zigbee: Es una tecnología de comunicación inalámbrica utilizada en sistemas de automatización del hogar y de control industrial.
- 5G: Es la última generación de tecnología móvil que permite la transmisión de datos a velocidades muy altas y con una latencia muy baja.

2.5.6.7. Forecasting

El forecasting es una técnica que ayuda a predecir lo que sucederá en el futuro, es un elemento muy útil en el proceso de planificación (Martínez, R. 2001). Por lo tanto, es considerado una de las técnicas elementales de la ciencia de datos, que involucra la predicción de valores futuros o la identificación de tendencias basándose en información histórica y en patrones que se han observado en los datos.

2.5.6.8. Algoritmo Gray Wolf Optimizer (Algoritmo del lobo gris)

El algoritmo Grey Wolf Optimizer (GWA) es un algoritmo metaheurístico de optimización, basado en inteligencia colectiva y desarrollado por Seyed Mohammad Mirjalili en el año 2014 (Mirjalili. 2014).

Este algoritmo se basa en el comportamiento social de las mandas de los lobos grises, ya que estos viven y cazan de sus presas preferidas, es por ello, que este algoritmo busca la forma de simular las principales características de la caza buscando encontrar la solución más óptima. Para entender este algoritmo es necesario comprender como funciona una manada de lobo grises. Estos por instinto tienen una técnica especial de caza donde la colonia opera en manadas o grupos para cazar sus presas dividiendo los grupos en: Lobo Alpha(es el que ocupa la posición superior en la colonia teniendo el derecho de decidir sobre toda la colonia), Lobo Beta (Informan a los

lobos Alpha y ayudan a la toma de decisiones), Lobos Delta (Son subordinados al lobo beta y buscan actualizaciones constantes a los lobos Alpha y Beta) y los lobos Omega(son responsables de cazar y de cuidar a los lobos más jóvenes de la manada). Por tal motivo, se dio la creación de este algoritmo de lobo grises buscando la solución más cercana utilizando varias funcionalidades integradas, como se mencionó anteriormente.

Esta técnica se usa en problemas que requieren de mucho tiempo porque reduce el tiempo operativo para datos de mayor dimensión ya que el algoritmo divide los problemas complejos en subconjuntos de operación similar a los grupos de la colonia de lobos grises.

2.5.6.9. Sistema adaptativo de inferencia neuro-difusa (ANFIS)

ANFIS es una combinación de las redes neuronales y la teoría difusa por lo tanto es una red neuronal artificial adaptiva Takagi-Sugeno. Las ANN en este sistema se usan para configurar el sistema difuso y con base a los patrones de entrada y salida, la propia red busca ajustar automáticamente los parámetros de diseño del sistema difuso y lograr así la función de autoaprendizaje y adaptación del sistema difuso. Este tipo de redes son aplicables a complejos sistemas de mapeo lineal y no lineal. En la siguiente figura 2-15 se puede observar una red neuronal neuro-difusa (ANFIS).

Como se puede observar en la figura anterior la red ANFIS presenta cinco capas o niveles donde cada capa representa algo. (Hang, R. 1993)

- Capa 1: Se encarga de determinar los grados de pertenencia para la entrada x al nodo
- Capa 2: Los nodos en esta capa son no-adaptivos por lo que solo dependen de los datos de entrada.

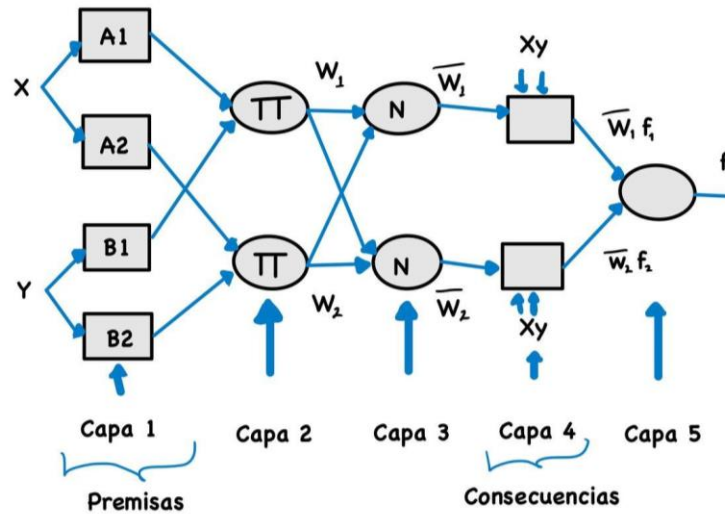


Figura 2.15. Red neuronal neuro-difusa (ANFIS).

Modificado de “Aplicación del modelo ANFIS para predicción de series de tiempo. Correa, G.J & Montoya, L. M”.

- Capa 3: Los nodos en esta capa son no-adaptativos y generan los pesos normalizados N.
 - Capa 4: Sus nodos son adaptativos, cuya salida es el producto del nivel de disparo normalizado y la combinación lineal de las entradas.
 - Capa 5: Con un único nodo no-adaptativo que calcula la salida total del sistema.

2.5.6.10. Redes de función de base radial (RBF)

La RBF es una red neuronal artificial que busca calcular la salida de la función según la distancia a un punto llamado centro, sirve para aproximaciones universales y se caracterizan por tener un entrenamiento o aprendizaje híbrido y hay tres capas diferentes en total, donde la capa 1 es de entrada, la capa 2 la oculta y la capa 3 de salida.

En la figura 2-16 se puede observar una arquitectura típica de RBF.

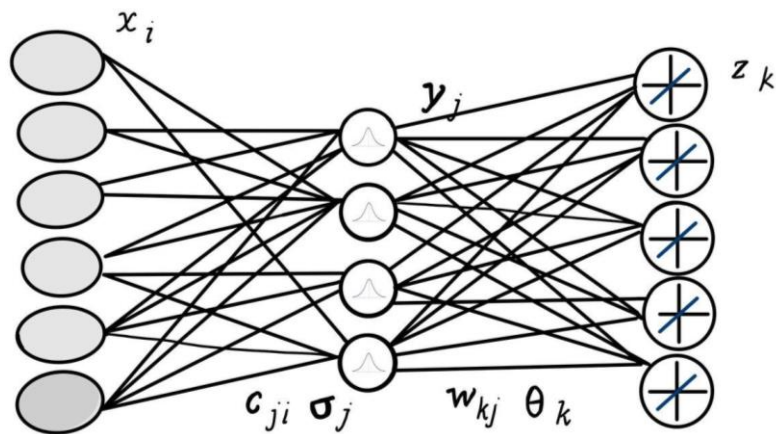


Figura 2.16. Arquitectura típica de Redes de función de base radial RBF

Aunque la arquitectura recuerda a la de un MLP, la diferencia es que las neuronas de la capa oculta en vez de calcular una suma ponderada de las entradas y aplicar una sigmoide, estas neuronas calculan la distancia euclídea entre el vector de pesos sinápticos (llamado en este tipo de redes de centro o centroide) y la entrada (de manera casi análoga a como se hacía con los mapas SOM) y, sobre esa distancia, se aplica una función radial con forma gaussiana. (Francisco. J. Palacios, 2003).

2.5.6.11. Naive Bayes (NBC)

El modelo de NBC es un tipo de algoritmo de clasificación de Aprendizaje Automático que es basada en el teorema de Bayes, por lo tanto, este es un clasificador probabilístico simple con fuerte suposición de independencia. A menudo proporciona una mejor precisión de clasificación en conjuntos de datos en tiempo real que cualquier otro clasificador. También requiere pocos datos de entrenamiento. El clasificador Naïve-Bayes aprende de los datos de entrenamiento y luego predice la clase de la instancia de prueba con la mayor probabilidad

posterior. También es útil para datos dimensionales altos ya que la probabilidad de cada atributo se estima independientemente (Chandra, 2007) (R. Mosquera, O. Castrillon, L. Parra 2018)

2.5.6.12. Algoritmo de aumento de gradiente (Gradient Boosting) (XGBoost).

Los modelos Gradient Boosting están formados por un conjunto de árboles de decisión individuales, entrenados de forma secuencial, de forma que cada nuevo árbol trata de mejorar los errores de los árboles anteriores. La predicción de una nueva observación se obtiene agregando las predicciones de todos los árboles individuales que forman el modelo. (J.Amat, 2007).

2.5.6.13. Redes neuronales de memoria de corto-largo plazo (LSTM).

La red neuronal de Memoria a Corto y Largo Plazo (LSTM, del inglés Long Short-Term Memory) es una extensión de la RNN, la cual ha logrado excelente rendimiento en diversas tareas, especialmente en problemas secuenciales (YiFei, 2018). En 1997, Sepp Hochreiter y Jürgen Schmidhuber introdujeron las LSTM, que han captado considerable atención por su capacidad para superar las limitaciones de las RNN convencionales. Las LSTM abordan estas limitaciones al incorporar una celda de memoria que retiene información a largo plazo mediante puertas de entrada que determina si se debe agregar nueva información a la celda, puerta de olvido que decide si se debe descartar la información existente y la puerta salida que controla el flujo de información desde la celda al resto de la red. En resumen, estas puertas controlan el flujo de información de manera selectiva, logrando que las LSTM mantengan un gradiente más estable durante el entrenamiento y resuelvan el problema del gradiente de fuga. (M. Frackiewichs, 2023).

2.5.6.14. Red funcional (FN – Functional network)

La red funcional es capaz de utilizar una combinación de conocimiento del dominio y de los datos en comparación con las redes neuronales en las que sólo se utiliza el conocimiento de

los datos. Las redes funcionales también pueden dar resultados óptimos solo con datos, pero son las principales candidatas para resolver problemas con datos y de conocimiento del dominio para problemas de clasificación y predicción. Además, son capaces de manejar una amplia gama de aproximaciones de funciones en problemas de ingeniería y estadística. (Kalam, S. et al. 2019)

2.5.6.15. Potenciación del gradiente (Gradient Boosting)

Gradient boosting es una técnica de aprendizaje automático utilizada para el análisis de la regresión y la resolución de problemas de clasificación estadística. Esta metodología genera un modelo predictivo en forma de un conjunto de modelos de predicción débiles, los cuales se combinan de manera iterativa para mejorar la precisión global del modelo final. Por ende, este consiste en un conjunto de árboles de decisión individuales, entrenados de manera secuencial, donde cada nuevo árbol se orienta a mejorar los errores de sus predecesores. La predicción de una nueva observación se logra combinando las predicciones de todos los árboles individuales que conforman el modelo. A diferencia de muchos métodos predictivos que generan modelos globales con una única ecuación para todo el conjunto de datos, el Gradient Boosting es una técnica que implica múltiples predictores basados en árboles de decisión. Esta aproximación contribuye a mejorar la precisión y adaptabilidad del modelo en comparación con enfoques más lineales. (J. Amat, 2007).

2.5.6.16. Máquina de aumento de gradiente ligero (lightGBM)

Light GBM o también llamado gradiente rápido es un algoritmo de alto rendimiento basado en un árbol de decisión. Principalmente utilizado para clasificación y diversas tareas de aprendizaje automático, este enfoque difiere de otros algoritmos de refuerzo al dividir las hojas del árbol utilizando el "mejor ajuste" en lugar de seguir una división por niveles de profundidad. Por lo tanto, cuando Light GBM trabaja en la misma hoja puede reducir más pérdidas que los

algoritmos por niveles, dando como resultado una mejor precisión que rara vez se puede lograr con los algoritmos de refuerzo existentes. Además, es sorprendentemente muy rápido, de ahí la palabra ‘Light’ (Barrios, J 2022).

3. Metodología para la selección en los procedimientos de diseño en el área de ingeniería de producción de hidrocarburo del ciclo profesional de ingeniería de petróleos de la Universidad Industrial de Santander

Como primer paso, se elabora una tabla con el contenido de las siguientes asignaturas: métodos producción de hidrocarburos, ingeniería del gas y facilidades de superficie de la carrera de ingeniería de petróleos de la Universidad Industrial de Santander. Cabe resaltar que dicho procedimiento se elaboró con la revisión del contenido del programa de la escuela ingeniería de petróleos referido a la universidad mencionada. Más adelante, se lista un conjunto de procedimientos (de diseño, de cálculo y de medición) que se realizan en la industria petrolera en el área de producción de hidrocarburos, cuyas fuentes son:

- Petroleum engineering handbook, Lary W. Lake.
- Petroleum production engineering. Boyun Guo. Xinghui Liu. Xuehao Tan. Second edition.
- Standard handbook of petroleum and natural gas engineering, Wiliam C Lyons. Gary J, Plisga. Michael D, Lorenz. Third edition.
- Volume I GENERAL ENGINEERING, John R. Fanchi, Editor.
- Volume III FACILITIES and CONSTRUCTION ENGINEERING Kenneth E. Arnold, Editor.
- Volume IV Petroleum engineering handbook, Lary W. Lake.

- Working guide to petroleum and natural gas production engineering, William C. Lyons.

A continuación, la tabla con el contenido de las asignaturas para extraer los temas correspondientes a procedimientos de diseño de ingeniería del área de producción de hidrocarburos.

3.1. Contenido de las asignaturas métodos de producción, facilidades de superficie e ingeniería del gas del ciclo profesional de ingeniería de petróleos.

- **Métodos de producción:** Esta asignatura busca que el estudiante adquiera competencias sobre los procesos y sistemas de producción existentes en los pozos de petróleo y gas. Los temas a tratar en esta unidad de estudio comprenden: La productividad de las formaciones que contienen hidrocarburos; El flujo de fluidos en el medio poroso y a través de tuberías; El análisis del sistema de producción para su posterior optimización; El diseño, el dimensionamiento, la especificación, la operación y la evaluación de los diferentes métodos de levantamiento artificial; Y finalmente las técnicas y estrategias de medición de los caudales provenientes de las formaciones productoras.
- **Ingeniera del Gas:** Esta asignatura busca que el estudiante desarrolle competencias en los procesos relacionados con la separación, tratamiento, procesamiento y transporte de gas natural, buscando el cumplimiento de los estándares de calidad, nacional e internacional.
- **Facilidades de superficie:** Esta asignatura busca que el estudiante desarrolle competencias sobre los procesos y sistemas existentes en las facilidades de

superficie para el manejo de la producción de los pozos de crudo y gas. Los temas para tratar en esta unidad de estudio comprenden: recolección, separación de dos fases, separación de tres fases, tratamiento de emulsiones, almacenamiento y fiscalización del crudo y el tratamiento de aguas aceitosas.

En la tabla 2 se muestra el contenido de las asignaturas mencionadas anteriormente, siendo parte del contenido del pensum académico de ingeniera de petróleos de la universidad industrial de Santander. La información está basada en el propósito de aprendizaje que se da a conocer por la escuela a los estudiantes que optan por el título de ingeniero de petróleos de la universidad industrial de Santander.

Tabla 3-1. Contenido de las asignaturas métodos de producción, facilidades de superficie e ingeniería del gas del ciclo profesional de ingeniería de petróleos:

Asignaturas	Contenido
Métodos de producción	<ul style="list-style-type: none"> - Flujo de fluidos en el medio poroso • Productividad de los yacimientos: <ul style="list-style-type: none"> - Índice de productividad (IP) - Método de Vogel - Método de Fetkovich - Comportamiento de entrada de los fluidos al pozo (IPR) (teniendo en cuenta los diferentes casos de IPR) - Flujo laminar (flujo Darcy) - Flujo turbulento (Flujo No-darcy) • Flujo de fluidos en el sistema de producción <ul style="list-style-type: none"> - Flujo en diferentes fases - Regímenes de flujo - Flujo multifásico en tuberías - Curvas de gradiente de presión • Análisis del sistema de producción (análisis nodal) <ul style="list-style-type: none"> - Nodos solución (en fondo del pozo, en cabeza de pozo, en el separador, en la presión de yacimiento). - Nodos funcionales (en choke de superficie, en la restricción de fondo, en las válvulas de seguridad, en el completamiento) - Análisis nodal de pozos de gas e inyección

Asignaturas	Contenido
	<ul style="list-style-type: none"> • Sistemas de levantamiento artificial <ul style="list-style-type: none"> - Bombeo neumático (gas Lift) - Bombeo mecánico (Rod pump) - Bombeo electro-sumergible (ESP) - Bombeo por cavidades progresivas (PCP) - Bombeo hidráulico (Hydraulic lift) - Bombeo por pistón (Pluger lift) - Sistemas de levantamiento artificial no-convencionales - Sistemas de levantamiento artificial combinados • Curvas de declinación (exponencial, armónica e hiperbólica) (en cada una se debe tener en cuenta: principio físico, modelamiento matemático, diseño, dimensionamiento y operación del sla). <ul style="list-style-type: none"> - Registros de producción - Pruebas y registros de producción
Facilidades de superficie	<ul style="list-style-type: none"> • Generalidades (química de hidrocarburos, propiedades físicas y químicas, comportamiento de fases, criterios de diseño) • Generalidades de las estaciones de producción (elementos, operación, localización y número de baterías de producción) • Sistemas de recolección <ul style="list-style-type: none"> - Líneas de flujo - Manómetros, válvulas, choques - Problemas con los sistemas de recolección (Abrasión, corrosión, formación de depósitos orgánicos e inorgánicos) • Separación de dos fases: petróleo y gas (separadores bifásicos) • Separación de gas – crudo – agua (separadores trifásicos) • Emulsiones • Sistema de tratamiento de crudo • Almacenamiento y fiscalización de crudo • Tratamiento de aguas aceitosas • Sistemas de transporte crudo
Ingeniería del gas	<ul style="list-style-type: none"> • Generalidades (Especificaciones del gas natural) • Equipos de proceso en la industria del gas. • Plantas de tratamiento de gas. • Plantas de procesamiento de gas. • Sistemas de transporte de gas.

Tabla 3.1. Contenido de las asignaturas métodos de producción, facilidades de superficie e ingeniería del gas del ciclo profesional de ingeniería de petróleos

Nota: Tomado de UIS - Escuela de Ingeniería de Petróleos - Posgrados y Pregrado - EIP -
INGENIERIA DE PETROLEOS - EIPWeb - Bucaramanga.

3.2. Procedimientos de diseño de ingeniería

El procedimiento de diseño de ingeniería en el área de producción de hidrocarburos implica el diseño de procesos y sistemas para garantizar que el sector Upstream y Midstream sea eficiente y seguro. Los ingenieros de petróleos del área de producción utilizan técnicas analíticas y tecnologías avanzadas para optimizar la producción, solucionar problemas y evaluar el rendimiento de los activos en el área de ingeniería de producción de hidrocarburos.

En cada etapa del proceso de diseño, se deben considerar diversos factores, como lo es la viabilidad técnica y económica del proyecto, los requisitos de seguridad y medio ambiente, la selección de los materiales y equipos adecuados, la capacitación del personal, entre otros. Cabe resaltar, que el procedimiento de diseño es un proceso dinámico, que puede requerir ajustes o modificaciones a medida que se van identificando nuevas necesidades o se van presentando nuevos desafíos durante la ejecución del proyecto.

Estos pasos de diseño son:

- **Identificación del problema:** Se identifica el problema o necesidad que se quiere resolver, con el fin de aumentar la producción de un campo de petróleo existente hasta el desarrollo de una nueva área de producción.
- **Evaluación de opciones:** En el momento en que se identifica un problema, se plantean las diferentes opciones para resolverlo, incluyendo la evaluación de la viabilidad técnica y económica.
- **Selección de la solución:** Después de evaluar las opciones, se selecciona la más eficiente para la solución del problema.

- **Diseño conceptual:** En esta fase, se realiza un diseño conceptual del proyecto, que incluye la definición de los objetivos, las especificaciones técnicas y los requisitos de seguridad y medio ambiente.
- **Diseño básico:** Después de completar el diseño conceptual, se desarrolla el diseño básico del proyecto, incluyendo el diseño de las facilidades de superficie tales como: separadores de producción, compresores, bombas y otros equipos de procesamiento necesarios para optimizar la producción de hidrocarburos.
- **Diseño detallado:** En esta fase se realiza un diseño detallado de cada componente del proyecto, incluyendo la selección de los materiales, la definición de los procedimientos de montaje, pruebas y la definición de los requisitos de calidad.
- **Construcción:** Después de completar el diseño detallado, se inicia la construcción del proyecto, incluyendo la construcción de los equipos, las tuberías y otros componentes del proyecto.
- **Pruebas:** Una vez completada la construcción, se hacen pruebas para verificar que el proyecto se complete adecuadamente, verificando la capacidad de producción, la comprobación de la seguridad y la calidad del proyecto.
- **Puesta en marcha:** Después de completar las pruebas, se inicia la puesta en marcha del proyecto. Una vez en funcionamiento, se realizan operaciones y mantenimiento periódicos para garantizar que el proyecto siga funcionando de manera adecuada y segura.

Es importante aclarar que en este proyecto de investigación se tendrá en cuenta aquellos procedimientos de cálculo, análisis y/o de medición que están fuertemente relacionados con los procedimientos de diseño, siempre y cuando su ejecución sea indispensable para realizar un determinado diseño.

3.3. Selección de procedimientos de diseño de ingeniería en las asignaturas métodos de producción, ingeniería del gas y facilidades de superficie del área de hidrocarburos.

Posteriormente, en la tabla 3-2 se evidencia la selección de los procedimientos de diseño, calculo y medición que se realizan habitualmente en la industria petrolera, específicamente en el área de producción de hidrocarburos. Dichos procedimientos se agrupan según el área mencionada, que es de tres materias: métodos de producción, ingeniería del gas y facilidades de superficie. El cuadro de selección se hizo con una lista no exhaustiva de los procedimientos, sino, solo tomando los más frecuentes en el área de producción de hidrocarburos. Esos procedimientos fueron extraídos de diferentes fuentes mencionadas al inicio del capítulo tres.

Áreas de producción	Si/No
Procedimiento	¿Presente en el plan de estudios?
Análisis de curvas de declinación (Exponencial, Armónica e hiperbólica).	Si
Análisis de curvas IPR.	Si
Análisis nodal.	Si
Diseño de bombas	Si
Diseño de estaciones de bombeo.	Si
Diseño de compresores.	Si
Diseño de deshidratadores de gas con desecantes sólidos.	Si
Diseño de deshidratadores de gas con glicol.	Si
Diseño de equipos de separación de sólidos.	No
Diseño de estabilizadores.	No
Diseño de estaciones de compresión.	Si
Diseño de intercambiadores de calor.	Si
Diseño de levantamiento artificial con bombeo hidráulico.	Si
Diseño de levantamiento artificial con gas lift.	Si
Diseño de sistemas de recolección (Líneas de flujo y tuberías)	Si
Diseño de oleoductos	Si
Diseño de plantas de endulzamientos con aminas.	Si
Diseño de separadores	Si
Diseño de tanques de almacenamiento	Si
Diseño de tratadores.	Si
Diseño del bombeo mecánico	Si

Áreas de producción	Si/No
Procedimiento	¿Presente en el plan de estudios?
Diseño del sistema de levantamiento artificial con bombas de cavidades progresivas (PCP).	Si
Diseño del sistema de levantamiento artificial con bombas.	Si
Diseño del sistema de levantamiento artificial con gas lift.	Si
Diseño del sistema de levantamiento artificial por pistón (pluger lift).	Si
Diseño del sistema de recolección de gas.	Si
Diseño del sistema eléctrico del campo	No
Diseño de instalación en alta mar y submarinas	No
Diseños e instalación de antorchas o teas.	No
Estimación del índice de productividad.	Si
Fracturamiento en pozos horizontales y desviados	No
Índice de productividad en pozos horizontales y desviados.	Si
Medición del contenido de agua en el gas.	Si
Medición del flujo del gas.	Si
Predicción de IPR (Vogel, Fetkovish, Standing (modificado), Fetkovish-Vogel y unificado).	Si
Predicción de la formación de hidratos.	Si
Prevención de la formación de hidratos.	Si
Sistemas de alivio de recipientes a presión.	No
Sistemas de seguridad.	Si
Sistema de transporte de crudo.	Si
Tratamiento de emulsiones.	Si

Tabla 3.2. Selección áreas de producción

3.4. Procedimientos de ingeniería presentes en el plan de estudios del área de producción de los hidrocarburos.

Después de revisar el contenido del área de producción de hidrocarburos presentes en el plan de estudios de la UIS y compararlo con los procedimientos de ingeniería presentados en las Tablas 2 y 3, se encuentran temas que corresponden a procedimientos de ingeniería, como cálculo, análisis, medición y diseño, los cuales se tratan en el desarrollo de las asignaturas mencionadas del área de producción de hidrocarburos mostradas en la tabla 3.

Por consiguiente, en la tabla 4 se muestran los temas que corresponden a dichos procesos de ingeniería del área mencionada anteriormente.

Asignaturas	Contenido
Métodos de producción	<ul style="list-style-type: none"> - Análisis de curvas de declinación de la producción. - Comportamiento de entrada de los fluidos al pozo (IPR): determinación del índice de productividad, predicción del IPR, análisis de curvas IPR. - Estudio de flujo multifásico en tuberías. - Análisis nodal. - Diseño del sistema de bombeo mecánico. - Diseño de instalaciones de gas lift. - Diseño del sistema de levantamiento artificial por bombas de cavidades progresivas (PCP). - Diseño del sistema de bombeo hidráulico
Facilidades de superficie	<ul style="list-style-type: none"> - Prevención y control de la depositación de asfaltenos y parafinas. - Diseño de estaciones de recolección de hidrocarburos. - Diseño y selección de separadores (Bifásicos y Trifásicos). - Diseño y selección de tratadores (térmicos y termo electrostáticos). - Tratamiento de emulsiones. - Diseño y selección de tanques de almacenamiento. - Diseño y selección de bombas. - Tratamiento del agua de producción. - Diseño de oleoductos.
Ingeniería del gas	<ul style="list-style-type: none"> - Diseño de separadores - Diseño de intercambiadores de calor. • Diseño de compresores. • Diseño de plantas de endulzamiento de gas (con aminas, con lecho, con membranas) • Diseño de plantas de deshidratación de gas (con glicol, con desecantes sólidos) • Fraccionamiento de gas. • Licuefacción de gas natural. • Diseño de gasoductos. • Medición de gas.

Tabla 3.3. Contenido de los procedimientos de ingeniería de producción de hidrocarburos

3.5 Selección de los procedimientos de diseño de ingeniería

La selección de los procedimientos es un aspecto importante en el diseño de ingeniería, ya que, este implica comprender las características de las diferentes áreas o temas, para así, seleccionar el que mejor se adapte al objetivo. Por lo tanto, se pueden aplicar diferentes técnicas de la ciencia de datos necesarias para ser presentadas en el área de producción de hidrocarburos del ciclo profesional de la carrera ingeniería de petróleos de la UIS.

Esto se realiza atendiendo las siguientes razones: que existan procedimientos y estudios previos sobre alguna técnica de la ciencia de datos al procedimiento que se va a ejecutar, en otras palabras, estudios que se hayan publicado en la literatura científica de la industria petrolera, en libros o revistas; específicamente en el área de producción; y, por último, que la técnica de la ciencia de datos aplicada signifique una ventaja en términos de calidad y resultados y/o reducción del tiempo de análisis o ejecución, con respecto a la manera convencional.

Los procedimientos que satisfacen los requisitos mencionados se pueden observar en las siguientes Tablas (Tabla 3.5, 3.6 y 3.7).

Titulo	Procedimiento	Técnicas(s) Aplicables(s)	Fuente
Un sistema basado en el aprendizaje automático para el flujo multifásico de autodiagnóstico	Flujo multifásico	ML, Artificial Neural Networks	Basbariol, T. et al. 2020
Un nuevo modelo para optimizar la asignación de gas en la operación de levantamiento de gas mediante algoritmos metaheurísticos	Diseño de implementación del gas lift	Algoritmo genético, algoritmo del lobo gris	Merzoug, A. et al. 2022
Una nueva correlación para predecir los caudales de gas utilizando inteligencia artificial: un enfoque industrial 4.0	Estimación de productividad del gas	ANN, Adaptive Neuro Fuzzy Inference System (ANFIS)	Kalam, A. et al. 2019
Un enfoque basado en redes neuronales de inteligencia artificial para pronosticar la producción en yacimientos no convencionales – Análisis comparativo con curva de declive	Análisis de curva de declinación para pronosticar la producción	ANN, forecasting, ML	Hassan, K. Clifford, L. 2021
Aplicación de técnicas de inteligencia artificial para predecir la productividad de pozos de espina de pescado	Estimación de la tasa de productividad	ANN, sistema de lógica difusa	Hassan, A. et al. 2019.
Aplicación de la Inteligencia Artificial para Estimar el Caudal de Petróleo en Pozos de Gas-Lift	Estimación de la tasa de productividad	ANN, Sistemas de interferencia neurodifusos artificiales (Artificial neuro-fuzzy inference systems), máquinas de vectores de soporte (support vector machines)	Mohammad, R. et al. 2020
Selección de levantamiento artificial mediante aprendizaje automático	selección de métodos de levantamiento artificial	ML, Naive Bayes, Red neuronal y árbol de decisión	Thanawit, O. et al. 2019
Pronóstico de producción autorregresivo y basado en aprendizaje automático: estudio de caso de Midland Basin	Análisis de curvas de declinación	ML	Gupta. et al. 2021
Análisis de la curva de declinación mediante inteligencia artificial	Análisis curvas de declinación	XGBoost, RNN, Random Forest	Shashipal, R. et al. 2019
Mejora de las predicciones en las operaciones petroleras mediante técnicas de inteligencia artificial.	Estimar la productividad del pozo	IA	Hassa. A. et al. 2019
Aplicación de aprendizaje automático para la predicción de la tasa de petróleo en pozos artificiales de extracción de gas	Estimar la productividad del pozo	Los sistemas artificiales de inferencia neurodifusa (ANFIS) y las máquinas de vectores de soporte (SVM) Artificial Neurofuzzy Inference Systems (ANFIS) and Support Vector Machines (SVM), Redes neuronales artificiales (ANN)	Mohammad, R. et al. 2019
La red neuronal como herramienta para predecir y controlar el régimen	Estimar la productividad del pozo	Memoria a largo y corto plazo (Long-Short Term	Vershinin, V. et al. 2020

Titulo	Procedimiento	Técnicas(s) Aplicables(s)	Fuente
tecnológico de los pozos de producción		Memory-LSTM), Red neuronal artificial	
Predicción de la producción de petróleo a través de modelos de regresión lineal y herramientas de Big Data	Predicción de la producción del petróleo	Big data, forecasting, ML	Alharbi, R. et al. 2023
La aplicación del análisis de Big Data en la optimización y selección Métodos de levantamiento artificial	Levantamiento Artificial	Big data, ANN, Aprendizaje no supervisado.	Junfeng, S. et al. 2018
El primer Levantamiento Artificial Inteligente Digital (DIAL) exitoso del mundo para Aplicación de doble cadena: un cambio de paradigma para el sistema de elevación de gas en alta mar Malasia	Levantamiento Artificial	ML	Nurfarah Izwana, S. et al 2022
Uso de la inteligencia artificial para predecir los derechos de propiedad intelectual de un pozo petrolero vertical en Solución Yacimientos derivados de gas: un nuevo enfoque	Predicción de la producción del petróleo (IPR)	IA, redes neuronales artificiales (ANN), lógica difusa (FL),	Salem, B. et al. 2018
Uso del gemelo digital de medidores Coriolis para la medición de flujo multifásico	Flujo Multifásico	Gemelo digital. (DT)	Sakethraman, M. et al. 2020
Uso del aprendizaje automático para la creación de modelos multivariantes de derechos de propiedad intelectual desde Datos de transmisión de frecuencia	Indice de productividad	ML, IA.	Mario A, P. et al. 2020
Diagnóstico de la condición de los pozos de bombeo de varillas de sucker basado en la Machine Learning de curvas de potencia eléctrica en el contexto de IoT	Diseño de bombeo mecanico	IoT	Cai, W. et al. 2020
Predicción del rendimiento de pozos en la formación Montney mediante enfoques de aprendizaje automático	Predicción de la producción del petróleo	ANN	Hamid, R. et al. 2020

Tabla 3.4. Procedimientos de diseño métodos de producción

Título	Procedimiento	Técnicas(s) Aplicables(s)	Fuente
Un sistema de ingeniería inteligente y computación predictiva basado en la nube para Diseño de tuberías y reducción de costos de operación	Diseño de Gasoductos	Gemelos digitales (digital Twins), IA	Chowdhury, K. et al. 2020
Una nueva correlación para predecir los caudales de gas utilizando Inteligencia: Un enfoque industrial 4.0	Estimación de la tasa de gas.	IA, Red neuronal (RNA), Red funcional (FN), Sistema adaptivo de interferencia neurodifusa (ANFIS)	Kalam, S. et al. 2019
Lograr el gemelo digital a través del soporte de analítica avanzada: una novedad Marco de detección para resaltar anomalías en tiempo real en series temporales	Diseño de planta de tratamiento de gas	Digital Twin	Cadei, L. et al. 2020
Gestión de la integridad de las tuberías de CO2: un enfoque de gemelo digital	Diseño de Gaseoductos.	Digital Twin.	Sun, T. 2023
Implementaciones de optimización de la producción de elevadores de gas DIAL inteligentes digitales en todos los activos		DIAL (Levantamiento artificial inteligente digital)	Ahmed, A. et al. 2021.
Enfoque inteligente para la mejora de la recuperación de petróleo GOSP	Diseño de separadores	ANN, CVM (máquina de vectores de soporte)	Dogali, A. et al. 2022
Novedosa aplicación de la IA: aprovechamiento del conocimiento tácito para la puesta en marcha de plantas de GNL	Planta de GLP	IA, big data, regression lineal, la regression de vectores de soporte (SVR Support Vector Regression) Y Máquina de refuerzo de gradiente de luz (Light Gradient Boosting Machine - LightGBM)	Lee, K. et al. 2022
Predicción de flujo de gas único a través de estranguladores utilizando inteligencia artificial Techniques	Predecir y estimar el caudal de gas a través de estranguladores.	IA, ANN, Lógica difusa (FL), máquina de vectores de soporte (SVM), Red funcional (FN) y árbol de decisión (DT).	Elhaj, M. et al. 2015
Automatización del diseño de tuberías submarinas mediante gemelo de campo digital	Diseño de tubería	Campo digital basado en la nube	Bhowmik, S. 2019

Titulo	Procedimiento	Técnicas(s) Aplicables(s)	Fuente
Automatización de la estructura submarina y el diseño de tuberías mediante gemelo de campo digital	Diseño de tubería	Digital Field Twin	Bhowmik, S. et al. 2020

Tabla 3.5. Procedimientos de diseño de ingeniería del gas

Titulo	Procedimiento	Técnicas(s) Aplicables(s)	Fuente
Aplicación de técnicas de soft computing a la medición de caudal multifásico: una revisión	Medición continua en líneas de flujo multifásico	red neuronal artificial (ANN), máquina de vectores de soporte (SVM), algoritmo genético (GA), programación genética (GP). y sistema neurodifuso adaptativo (ANFIS).	Yan, yong. Et al. 2018
Investigación de Aplicación en Tecnología de Gestión Inteligente basada en Oleoductos y Gasoductos	operación de oleoductos y gasoductos	Sistema SCADA inteligente	Liu, X. 2020
Desarrollo de un sistema experto para la selección de correlaciones flujo multifásico	a seleccionar la correlación de flujo multifásico más adecuada en ausencia de una presión de flujo medida	expert system (sistemas expertos)	Mohamed A. Abd El-Moniem l · Ahmed H. El-Banbi. 2017
Primeros Principios y Medición de Flujo Virtual de Aprendizaje Automático: Una Revisión de la Literatura	medir caudales multifásicos	Medición de flujo virtual - Virtual Flow Metering (VFM)	Timur, B. et al. 2019
Mejora de la modelización y el control de las operaciones de las instalaciones de transporte de petróleo y gas mediante inteligencia artificial	control de una estación de bombeo. Determina características importantes de la tubería	lógica difusa y las redes neuronales artificiales (RNA)	Neuroth, M. et al. 2000
Mejora de la predicción del rendimiento del estrangulador multifásico y la validación de las pruebas de producción de pozos mediante inteligencia artificial: nuevo milestone	Predicción del rendimiento multifásico	IA, Logica Difusa y Redes neuronales	Mohammed,A. 2015
Detección de fugas en gasoductos de gas natural mediante modelos inteligentes	Tratamiento de emulsiones	Gradient Boosting, Árboles de Decisión, Bosque Aleatorio (Random Forest), MVS	Akinsete, O. 2019

Título	Procedimiento	Técnicas(s) Aplicables(s)	Fuente
		y Red Neuronal Artificial	
Optimización del tamaño del estrangulador para flujo bifásico utilizando inteligencia artificial	Diseño de estaciones de recolección de hidrocarburos.	redes neuronales artificiales, lógica difusa (FL), máquinas de vectores de soporte (SVM) y redes funcionales.	Klamis, m. 2019
Hoja de ruta para desarrollar una red neuronal artificial para predecir el régimen de flujo bifásico en tuberías inclinadas	Diseño y selección de separadores Bifásicos.	Red Neuronal	Alsaif, A. 2022
Predicción de la relación de rendimiento de entrada de un yacimiento de gas mediante técnica de inteligencia artificial	IPR pozo de gas	IA, RNA, FL	Ketmalee, T. 2019

Tabla 3.6. Procedimientos de facilidades de superficie

4. Descripción de la metodología para la implementación de la ciencia de datos en el área de ingeniería de producción basados en ejemplos seleccionados.

4.1. Análisis bibliométrico

Se puede identificar el interés por el tema debido al aumento del petróleo, incentivando a las empresas a desarrollar nuevas tecnologías y así fortalecer las operaciones. Por tal motivo la aparición y evolución de la ciencia de datos ha permitido el desarrollo de soluciones para problemas complejos que anteriormente solo estaba al alcance de las soluciones analíticas o la simulación numérica.

Dado el notable interés en la investigación de estos temas, se observa una considerable cantidad de información generada en los últimos años. Por esta razón, ha captado la atención de la industria petrolera en investigar más estos temas, dando lugar a la formación de la disciplina denominada bibliometría. Esta disciplina hace uso de conocimientos estadísticos y matemáticos con el fin de analizar el funcionamiento científico, contando con mecanismos destinados a evaluar diversos aspectos de este fenómeno, conocidos como indicadores bibliométricos. Estos

indicadores abarcan elementos como la cantidad de documentos publicados por año, los autores, los países, las áreas temáticas, las palabras clave y las citas de autores. En base a lo anterior, la investigación se ha creado con base en Scopus que permite extraer y analizar la información adquirida, facilitando un buen criterio de investigación para la recolección de datos.

El criterio de la investigación se basó en un análisis tipo bola de nieve de citas de los artículos más relevantes en el tema de investigación, considerando que cada año se registra gran variedad de artículos es necesario acudir al análisis bibliométrico de tal forma que se puedan obtener los documentos relevantes para el tema de investigación en estudio.

Cabe resaltar, que los indicadores básicos se tendrán en cuenta para mostrar los documentos publicados por año, autores, países, área temática, para los indicadores de relación como palabras clave y citación de autores. Para ello, se utilizará la herramienta VOSviewer que sirve para construir y visualizar redes bibliométricas como palabras claves y citación de autores entre otras.

4.1.1. Scopus

Para realizar la búsqueda en Scopus se utilizó la ecuación, (("Data science") AND ("production") AND ("Oil and gas")) el día 02/02/2023. Se limitó la información teniendo en cuenta los documentos con enfoque al proyecto propuesto desde el año 2016.

4.1.1.1. Indicadores básicos: Al analizar los resultados de la búsqueda, se obtuvieron los siguientes gráficos estadísticos.

- **Documentos por año**

En la figura 4-1 se puede observar el crecimiento exponencial de las publicaciones y esto hace referencia al interés de los investigadores respecto al tema con el pasar de los años.

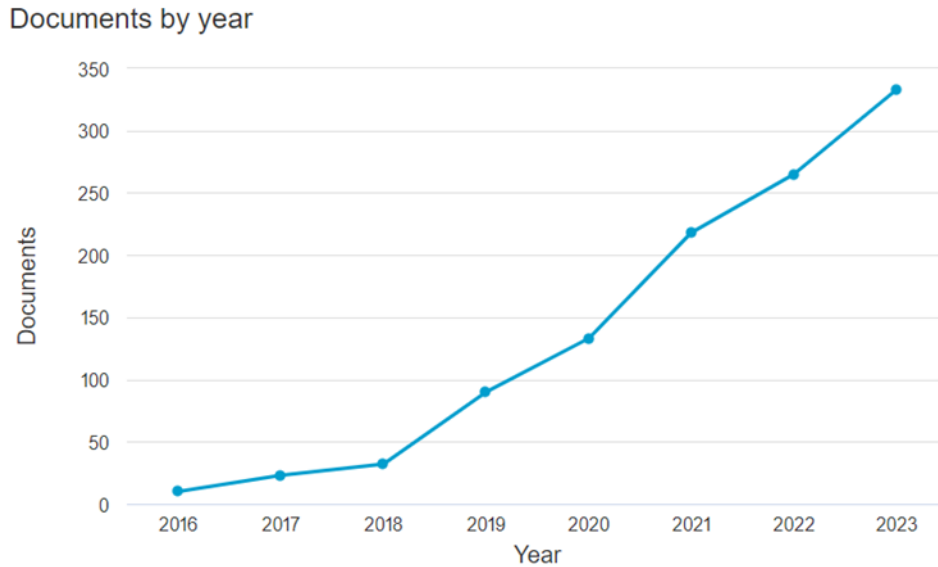
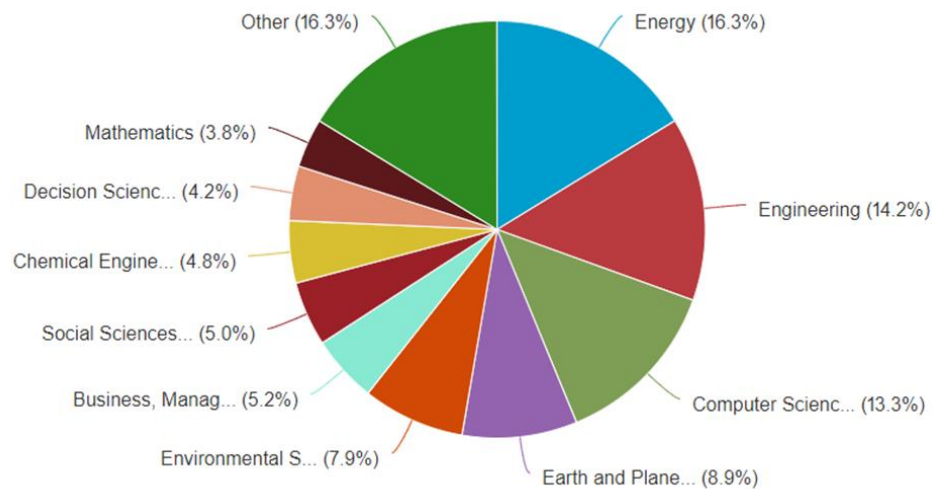


Figura 4.1. Desarrollo del tema con el pasar de los años en Scopus.

- **Área temática**

En la figura 4-2 se observa las áreas de investigación más relevantes en esta temática teniendo en cuenta los documentos publicados son energía e ingeniería con un 16,3% y 14,2% respectivamente. Seguido ciencias de decisiones y ambiente con un 13,3% y 8,9% respectivamente. Concluyendo que la energía e ingeniería predominan en los temas de interés de la investigación.

Documents by subject area

*Figura 4.2. Área temática en Scopus*

- **Indicadores de relación**

Para la realización de los indicadores de relación se utilizó el software de VOSviewer, el cual es una herramienta de software que nos ayuda a construir y visualizar redes bibliométricas, estas se pueden construir en base a citas, acoplamiento bibliográfico, co-citas o relaciones de autorías.

VOSviewer también ofrece funciones de extracción de texto que se pueden utilizar en redes de concurrencia de términos importantes extraídos de un cuerpo de literatura científica.

- **Palabras clave**

En la figura 4-3 se observa la concurrencia de palabras clave reunidas por colores en 5 clústeres los cuales representan algoritmos de agrupamiento por temáticas relevantes. En los colores que mejor se puede evidenciar es en el rojo y en el verde en donde la palabra más usada es big data, gas industry, petroleum industry, machine learning, fore casting respectivamente.

Además, es importante destacar que los círculos que tienen mayor tamaño representan más frecuencia de aparición de la palabra en los artículos.

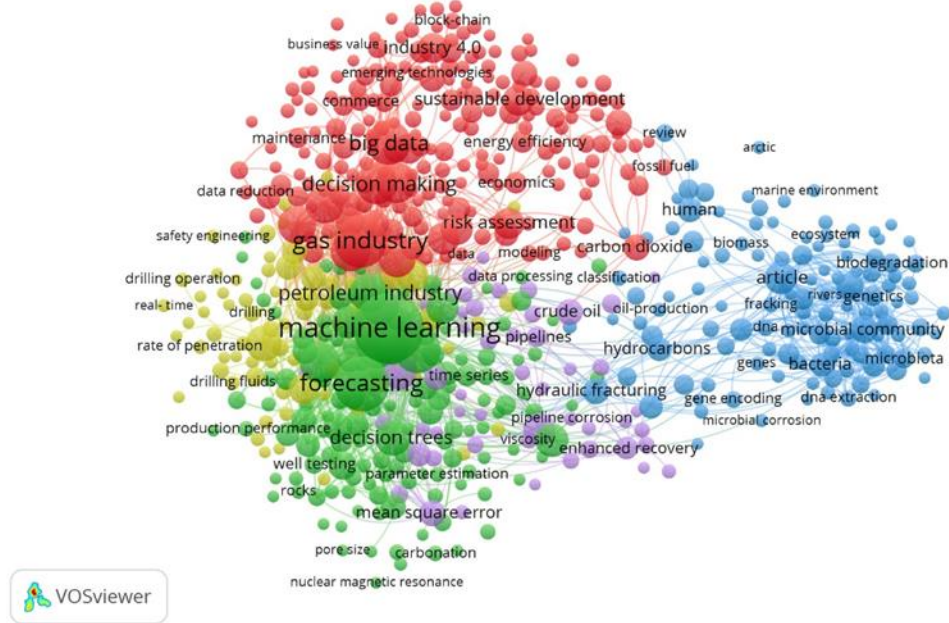


Figura 4.3. Palabras clave en Scopus

4.2. Indicadores relevantes para la selección de ejemplos

Para la selección de los ejemplos, se tuvo en cuenta los indicadores correspondientes al área de Ingeniería de Producción tales como: cálculos, análisis, medición y diseño de las asignaturas mencionadas del área de producción de hidrocarburos (producción de hidrocarburos, facilidades de superficie e ingeniería del gas), mencionadas anteriormente en el Item 3.4. Cabe resaltar que esta revisión se basó en fechas recientes de 2018 en adelante, así como en palabras claves que tuvieron una temática relevante con varias citas como se puede observar anteriormente en el ítem 4.1.2.5.

Por tal motivo, los ejemplos se explicarán a continuación de forma detallada y concisa, para así luego pasar a hacer la respectiva comparación con el método convencional que es lo visto en el pensum académico de Ingeniería de Petróleos de la Universidad Industrial De Santander.

4.2.1. Ejemplos métodos de producción

4.2.1.1. Un enfoque basado en redes neuronales de inteligencia artificial para predecir la producción en yacimientos no convencionales: análisis comparativo con la curva de declive¹

Objetivo

Predecir la producción en yacimientos no convencionales mediante redes neuronales.

Justificación

Frente al agotamiento de las reservas convencionales, se observa un incremento en la explotación de yacimientos no convencionales, este fenómeno es debido al aumento en la demanda de países en desarrollo. Los yacimientos no convencionales presentan características y propiedades distintas a los convencionales, lo cual plantea grandes desafíos en las técnicas tradicionales y destaca la necesidad de implementar enfoques novedosos a partir de conjuntos de datos para encontrar soluciones requeridas y tomar decisiones.

Datos

¹ Nota: Hassan, K. Clifford, L. 2021. An Artificial Intelligence Neural Networks Driven Approach to Forecast Production in Unconventional Reservoirs – Comparative Analysis with Decline Curve.

Se utilizaron los siguientes datos de entrada: presión en boca de pozo (WHP - wellhead pressure), presión en fondo de pozo (BHP - bottom-hole pressure) y parámetros del yacimiento, manteniendo la tasa de gas como parámetro objetivo.

Procedimiento

- 1. Entrada de datos brutos:** Se recopila la mayor cantidad de datos.
- 2. Aprendizaje automático:** Después de realizar la recolección de los datos, se selecciona un algoritmo apropiado para llevar a cabo los cálculos y obtener resultados con la menor cantidad de errores posibles.
- 3. Predicción:** Una vez entrenado el modelo con el menor error, se permite predecir el rendimiento de la producción y a medida que se adquieren más datos en el entrenamiento del modelo de aprendizaje automático mejor es el ajuste de este.
- 4. Optimización:** Con lo anterior mencionado, los ingenieros de producción planifican estrategias de optimización para mitigar el tiempo no productivo. Gracias a este método basado en datos, se pueden identificar las zonas calientes en yacimientos no convencionales, que tienen tendencia a una mayor producción. Además, también se pueden realizar análisis de sensibilidad para visualizar los resultados.
- 5. Arquitectura de redes neuronales:** Las ANN aprenden de la función matemática y pueden seguir utilizándose con fines de entrenamiento (Más información de este tema en el ítem 2.5.5.1).
- 6. Modelización basada en datos:** Para el entrenamiento, la red neuronal realiza ciclos de propagación hacia delante y hacia atrás. Con el fin de facilitar la comprensión del modelado, como se evidencia en la Figura 4-4.

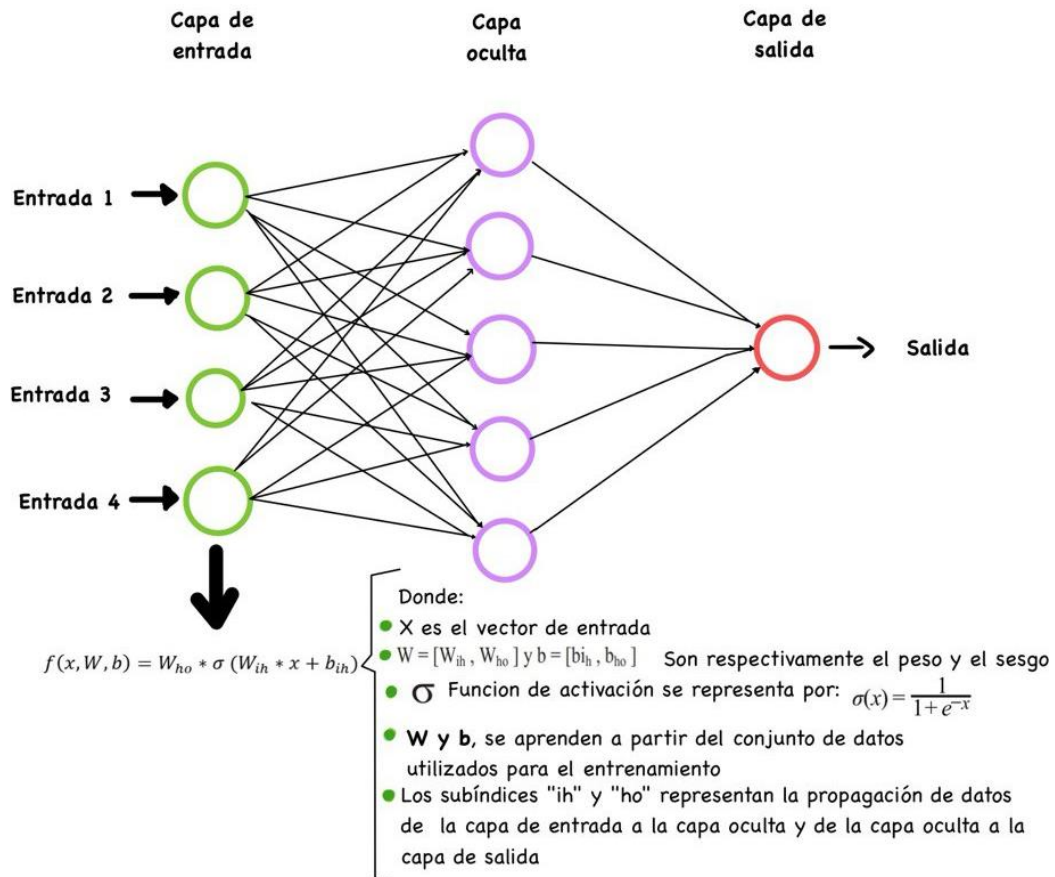


Figura 4.4. Modelo ANN, según ejemplo

7. Selección del número óptimo de capas ocultas: La precisión de los resultados es función de las numerosas capas ocultas y neuronas de la red. El número de capas ocultas se realiza mediante la siguiente ecuación 1:

$$N_h = \frac{N_s}{(\infty * (N_i + N_o))}$$

Ecuación 4-1. Número de capas ocultas

Donde: N_h es el número de capas ocultas, ∞ es el factor de escala constante que oscila entre el valor de 2-10. N_s es el número de muestras del conjunto de datos de

entrenamiento, mientras que N_i y N_o son el número de capas de entrada y salida que depende del número de parámetros en las capas de entrada y salida.

- 8. Pronóstico de series temporales:** En este paso se estudia una variable durante un periodo determinado, los datos se recogen en tiempo real y se analizan para generar resultados gráficos y numéricos. El análisis del pronóstico proporciona información sobre el comportamiento del parámetro en el futuro.
- 9. Validación del modelo:** El modelo de redes neuronales entrenado se valida y comprueba mediante técnicas estadísticas calculando el coeficiente de determinación, error cuadrático (R^2). Un valor r-cuadrado de 1 significa que el resultado predicho es preciso al 100% e igual al resultado objetivo en tiempo real. Al entrenar el modelo de redes neuronales con distintos parámetros de pozo y yacimiento, se debe asegurar que el error r-cuadrado se aproxima a 1.

Resultados

Se llevó a cabo una previsión estadística utilizando el conjunto de datos de entrenamiento para yacimientos de gas no convencionales (tight y shale gas), se procedió a validar y probar dicha previsión mediante el modelo de Red Neuronal Artificial. El resultado obtenido fue un coeficiente de determinación (r-cuadrado) de 0,996. Es decir, en este caso se demuestra que la técnica ANN es una herramienta con mucho potencial ya que predice con exactitud la producción en tiempo real.

4.2.1.2. Aplicación de técnicas de inteligencia artificial para predecir la productividad de los pozos de espina de pescado²

Objetivo

Determinar la productividad de los pozos de espina de pescado o también llamados “Fishbone Wells” mediante técnicas de inteligencia artificial

Justificación

Los pozos de espina de pescado son aquellos que se someten a fracturamiento hidráulico, pero se distinguen por realizar esta operación en múltiples direcciones. Esta característica le otorga la capacidad de lograr tasas de producción elevadas en comparación con los otros pozos, además, aumenta el área de contacto entre el fondo del pozo y la región del yacimiento. Por ende, son pozos que se destacan por tener un tiempo operativo reducido y un costo competitivo en comparación con la fracturación hidráulica convencional. Sin embargo, estos pozos reportan modelos limitados para determinar la productividad de los pozos siendo así, de gran importancia para el diseño y la finalización de un pozo de producción como en la selección de las facilidades de superficie.

Datos

Los modelos se desarrollaron y validaron utilizando 250 conjuntos de datos, siendo las entradas el índice de permeabilidad (Kh/Kv), presión de fondo de pozo que fluye y longitud lateral para determinar la productividad del pozo. Se utilizaron datos de campo reales.

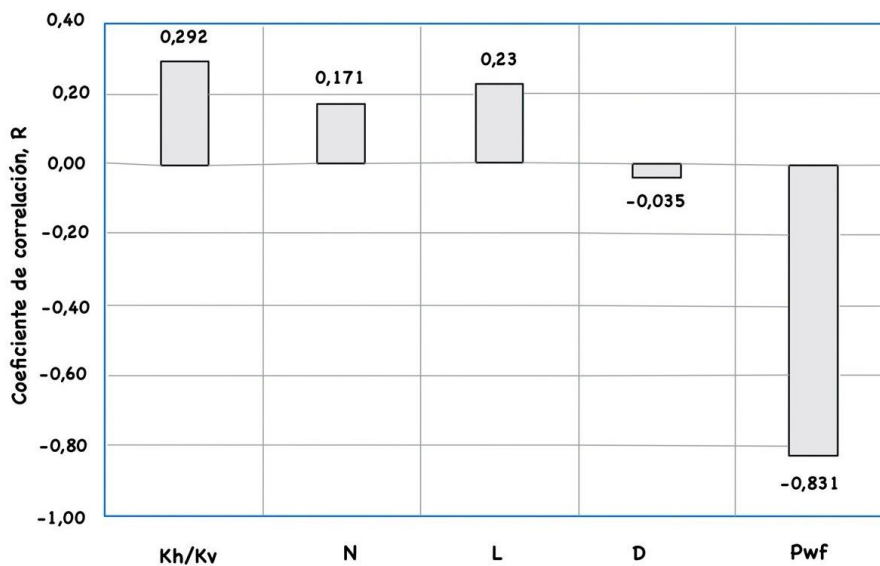
Metodología

² Nota: Hassan, A. et al. 2019. Application of Artificial Intelligence Techniques to Predict the Well Productivity of Fishbone Wells

- 1. Recolección y análisis de datos:** los datos utilizados fueron generados por un software comercial, con más de 250 corridas realizadas. Los datos se dividen en dos conjuntos: los datos de entrenamiento que son el 70% de los datos totales y los de prueba que se usaron para evaluar la confiabilidad de los modelos, estos son un 30% de los datos totales.
- 2. Análisis estadístico:** Se describe la probabilidad de ocurrencia de diferentes valores en un conjunto de datos por medio de funciones matemáticas, luego se obtiene una estimación aproximada de cómo se distribuyen esos datos. Todo se resume en la tabla 4.2.

Parámetro	Kh/Kv	No. de laterales	longitud (ft)	distancia (ft)	Pwf (psia)	caudal (scf/D)
Mínimo	1	2	700	1300	14,7	0
máximo	1000	14	3100	5200	4800	197903,226
significar	61	6,667	2.759,523	2.723,809	2359,558	81860,474
modo	10	6	3100	2600	14,7	0
rango	999	12	2400	3900	4785,3	197903,226
Desviación Estándar	211,275	2,499	693,099	685,121	1551,738	48712,516
oblicuidad	4,192	1,412	-1,916	2,0689	0,09535	-0,118
curtosis	18,73	5,358	5,308	9,503	1,7184	2,216
coeficiente de variación	346,352	37,491	25,116	25,153	65,763	59,507

Tabla 4.1. Análisis estadístico del ejemplo



Longitud lateral, k_h/k_v – relación de permeabilidad, r_{wf} – presión de fondo de pozo que fluye

3. Determinación de coeficiente de correlación: Medir el impacto de los parámetros de entrada en el caudal del pozo, como se muestra en la Figura 4-5

4. Análisis de sensibilidad para cada modelo de inteligencia artificial

En el artículo se realizó una investigación de tres modelos de inteligencia artificial con el objetivo de determinar cuál de los tres exhibe el menor error porcentual promedio (AAPE – average absolute percentage error) y un valor máximo del coeficiente de correlación (CC – correlation coefficient).

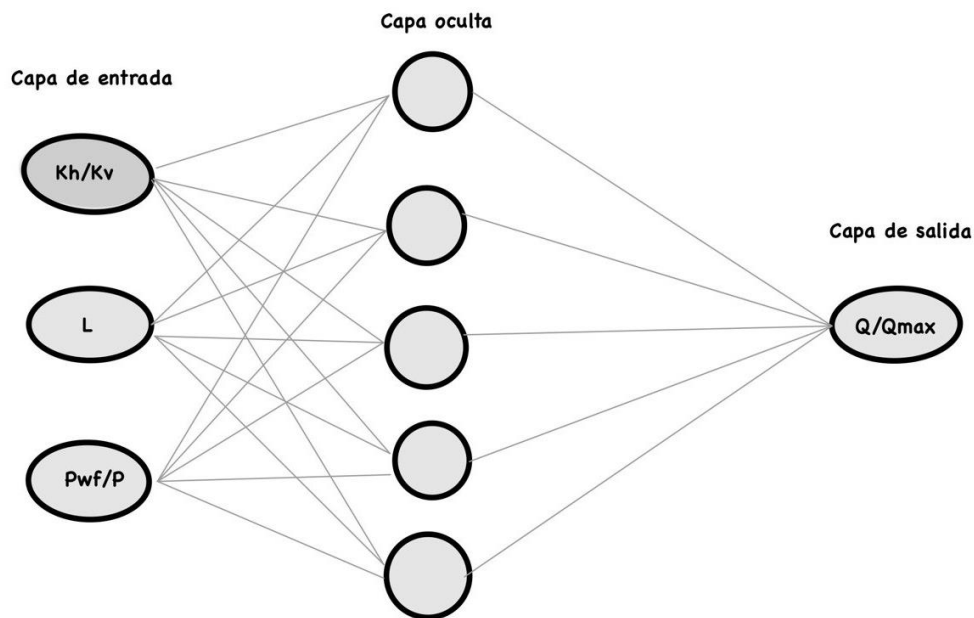


Figura 4.6. Modelo ANN ejemplo de producción

- **Red neuronal artificial**

Los datos de entrada del modelo fueron: relación de permeabilidad, longitud lateral y la presión de fondo del pozo. El modelo ANN desarrollado consta de tres capas: entrada, oculta y salida como se observa en la Figura 22. Se encontró que el número óptimo de neuronas era 5, el modelo ANN se entrenó utilizando el 70% de los datos.

Se investigaron diversos escenarios con el fin de mejorar los parámetros del modelo, cuyos resultados se muestran en la siguiente tabla 4.3:

Escenario No.	No. De capas ocultas	No. De neuronas en la capa	CorrCoef_Test	AAPE_test
1	1	10	0,9849	15,706
2	1	20	0,9796	7,2327
3	2	20	0,9547	14,4647
4	3	20	0,9868	11,6491

Tabla 4.3. Resultados de los diversos escenarios del ejemplo

Donde:

- **CorrCoef_Test** (Correlation coefficient for testing data set): Coeficiente de correlación para probar el conjunto de datos.
- **AAPE_test** (Average absolute percentage error for testing data set): Error porcentual absoluto promedio para el conjunto de datos de prueba
- **Sistema de lógica difusa**

En la tabla 4.4 se resumen los resultados de ANFIS:

Escenario No.	Cluster radius	No. De interacciones	CorrCoef_Test	AAPE_test
1	0,1	200	0,9822	14,4962
2	0,3	200	0,9838	14,1589
3	0,8	200	0,9845	13,9187
4	0,7	200	0,9845	13,9242
5	1	200	0,9845	13,9208
6	0,6	100	0,9845	14,0791

Tabla 4.4. Resultados del sistema de lógica difusa, según el ejemplo

- **Red de función de base radial (RBF)**

En este modelo se examinaron diferentes valores de objetivo, extensión y número máximo de neuronas para optimizar los parámetros y los resultados de este modelo se resumen en la Tabla 4.5.

Escenario No.	Objetivo	Extensión	MN, máximo Número de neuronas	CorrCoef_Test	AAPE_test
1	0	100	10	0,8786	19,3701
2	0	100	15	0,983	11,467
3	0	100	20	0,9851	11,1697
4	0	50	20	0,9851	11,1464
5	0,5	10	20	0,8614	32,8188

Tabla 4.6. Resumen de parámetros y resultados del modelo de red de función de base radial

Resultados

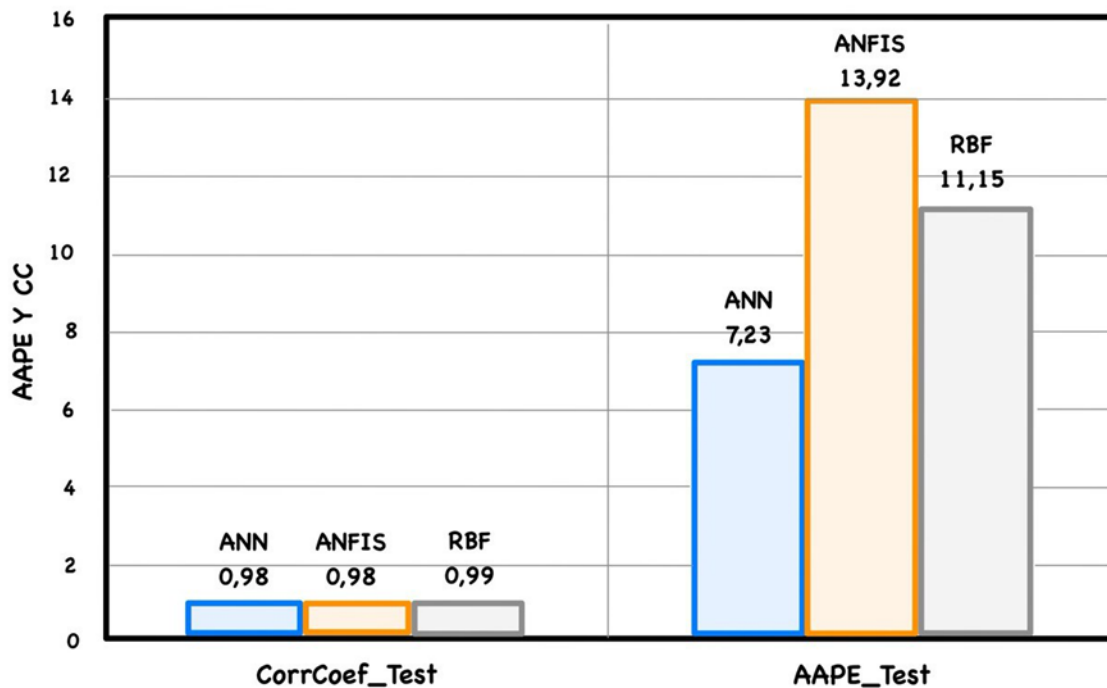


Figura 4.7. Comparación de las diferentes técnicas de inteligencia artificial.

El mejor escenario obtenido del modelo de redes neuronales artificiales fue el segundo escenario con un error absoluto promedio mínimo del 7,23% y un coeficiente de correlación de 0,979 utilizando 1 capa oculta con 20 neuronas.

El modelo de lógica difusa presenta el mejor escenario mediante el uso de cinco funciones de membresía, una función de membresía de salida lineal, un radio de grupo de 0,8 y un número de iteración de 200. para lo cual la AAPE fue del 13,92 % y el coeficiente de correlación fue de 0,985

Por último, para el sistema de lógica difusa el caso óptimo fue el escenario cuatro, donde se utilizó un valor objetivo de 0,0, un valor de dispersión de 50 y un MN de 20. Para este caso, el coeficiente de correlación fue de 0,985 y el error absoluto fue de 11,14%

En la figura 4-7, se observa que al comparar la productividad de espina de pescado obtenida utilizando la red neuronal artificial (ANN), el sistema de lógica difusa (ANFIS) y la función de base radial (RBF), el mejor modelo desarrollado es una red neuronal artificial la cual mostró el mejor rendimiento de predicción, con un error absoluto del 7,23% que indica que el modelo ANN supera a todos los métodos de IA al momento de predecir la productividad de espina de pescado.

4.2.1.3. Predicción de la optimización del sistema de producción de petróleo a través de modelos de regresión lineal y herramientas Big Data³

³ Alharbi, R. et al. 2023. Prediction of Oil Production Through Linear Regression Model and Big Data Tools.

Objetivo

Determinar la tasa de producción de petróleo y gas mediante modelos de aprendizaje automático.

Justificación

La dirección futura de la industria petrolera se sustenta en la optimización de la producción actual y las proyecciones a futuro del petróleo y gas. Con el transcurso del tiempo, se observa un creciente interés por parte de científicos e investigadores en abordar este tema mediante la aplicación de la ciencia de datos. Este interés se justifica por la relevancia del petróleo en la generación de energía, desempeñando así un papel crucial en la mejora de la economía. La influencia positiva se extiende a los ámbitos políticos, tanto en países en desarrollo como en aquellos ya establecidos.

Datos

La estimación de la producción está influenciada por parámetros como la presión en la tubería y en el fondo del pozo, y puede realizarse eficazmente mediante correlaciones empíricas o análisis nodal. El proceso condujo a la selección de seis variables como entradas para predecir la producción de petróleo, incluyendo la temperatura promedio en la boca del pozo, el tamaño promedio del estrangulador y la tasa promedio de elevación de gas, entre otras.

Metodología.

Los investigadores emplearon modelos de regresión lineal y polinomial para anticipar la producción diaria del campo Volve. Para ello, se comenzó con la adquisición de datos de sensores, que fueron posteriormente procesados donde se aplicó el modelo CRISP (Proceso Estándar Interindustrial Cruzado) a los datos procesados, lo que implicó entender, preparar, modelar y evaluar los datos.

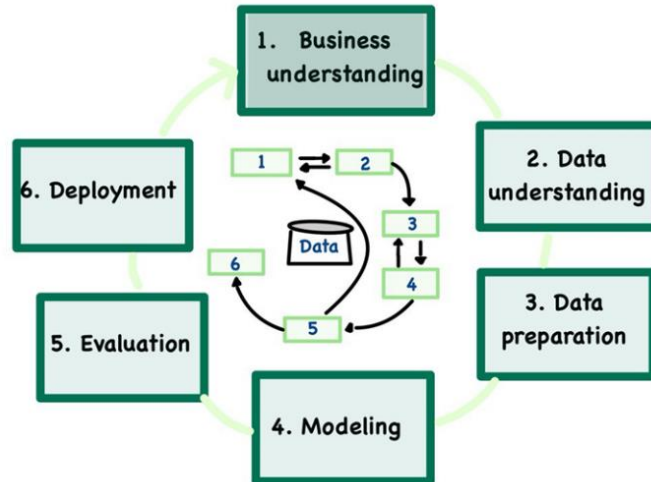


Figura 4.8. Modelo CRISP (Proceso Estándar Interindustrial Cruzada)

Después, los datos se trasladaron a Rapid Miner, donde se realizó un análisis de correlación para investigar las relaciones entre ellos y entre las variables. Luego, se utilizó la regresión lineal con las variables identificadas en el análisis de correlación.

Paso a paso:

- 1. Importación de datos:** Durante la carga de los datos, se consideran diversos parámetros que abarcan información tales como: El tipo de producto (inyección de agua o producción de petróleo), la presión promedio en el fondo del pozo, la longitud promedio de la tubería de perforación, la fecha de producción, el volumen de petróleo, la temperatura promedio en el fondo del pozo, el volumen de gas, el tamaño promedio del estrangulador, la temperatura promedio en la boca del pozo, la presión promedio del anillo, la presión promedio en la cabeza del pozo, el volumen de agua y el tipo de flujo (producción o inyección). Además, los datos de producción de petróleo se registran mediante la recolección diaria.

2. **Análisis de datos exploratorios:** En el análisis de datos se llevan a cabo varios procesos, que incluyen la identificación de patrones, la validación de suposiciones y la detección de anomalías mediante el uso de visualizaciones y resúmenes estadísticos. Por ende, en esta investigación se importaron datos de 7 pozos del campo Volve. En consideración a los siete pozos analizados, se concluye que los datos de algunos pozos deben ser descartados, ya que no son aptos para su uso en el proceso de entrenamiento ya que se detecta la presencia de valores vacíos.
3. **Procesamiento de datos:** El procesamiento de los datos revela información notable según el pozo, lo que demanda atención para abordar los datos faltantes. Por ende, los investigadores emplean la técnica de llenado directo para completar los valores ausentes, ya que el uso del valor medio no es apropiado debido a información presente en el estudio. Por lo tanto, se opta por completar los valores nulos con una cantidad ligeramente mayor.
4. **Escala de datos:** Este proceso ajusta los valores numéricos de las columnas a una escala uniforme, manteniendo las diferencias de rango y preservando la información. Siendo así, los datos se transforman en un rango de [0,1]. Ec 2

$$X' = (X - X_{\min}) / (X_{\max} - X_{\min})$$

Ecuación 4-2. Ecuación numérica para el escalado o normalización de datos

5. **Selección de funciones:** Implica evaluar su relevancia e impacto en los resultados. Se eligen aquellas que contribuyen a una predicción precisa y están correlacionadas con el resultado esperado, ya sea de forma automática o manual.
6. **Aplicación de algoritmo de aprendizaje automático:** En un análisis de regresión, se estudian observaciones que involucran al menos dos características o

variables para comprender la relación entre ellas. La regresión examina cómo estas variables se relacionan entre sí y con el fenómeno elegido, utilizando suposiciones, pruebas y análisis adicionales. Los resultados se representan con la letra "y" para el fenómeno, y las variables de entrada se denotan como "X", formando un vector " $X = (X_1, \dots, X_r)$ " que contiene las características independientes. El objetivo es establecer una relación lineal entre la variable dependiente "Producción de petróleo" y las variables independientes de la formación. Se asigna un coeficiente a cada característica de entrenamiento y se incluye una intersección en la ecuación lineal. La instancia de entrada se representa como un vector de características " $X = (X_0, X_1, \dots, X_n)$ ", y la salida predicha como " $Y = (W_0X_0 + W_1X_1 + \dots + W_nX_{n+b})$ ".

Resultados

Los investigadores usaron las bibliotecas Scatterplot y Matplotlib para representar gráficamente cómo el modelo predice la producción de petróleo en comparación con los valores reales, como se evidencia a continuación en la figura 4-9.

El error cuadrático medio (ECM) refleja el promedio de las diferencias al cuadrado entre los valores originales y los predichos en el conjunto de datos, lo que cuantifica la dispersión de los residuos. Por otro lado, la raíz cuadrada del error cuadrático medio (RMSE) EC 3. representa la desviación estándar de los residuos, proporcionando una medida del error promedio en la predicción del modelo.

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

Ecuación 4-3. Valor de error cuadrático medio RMSE

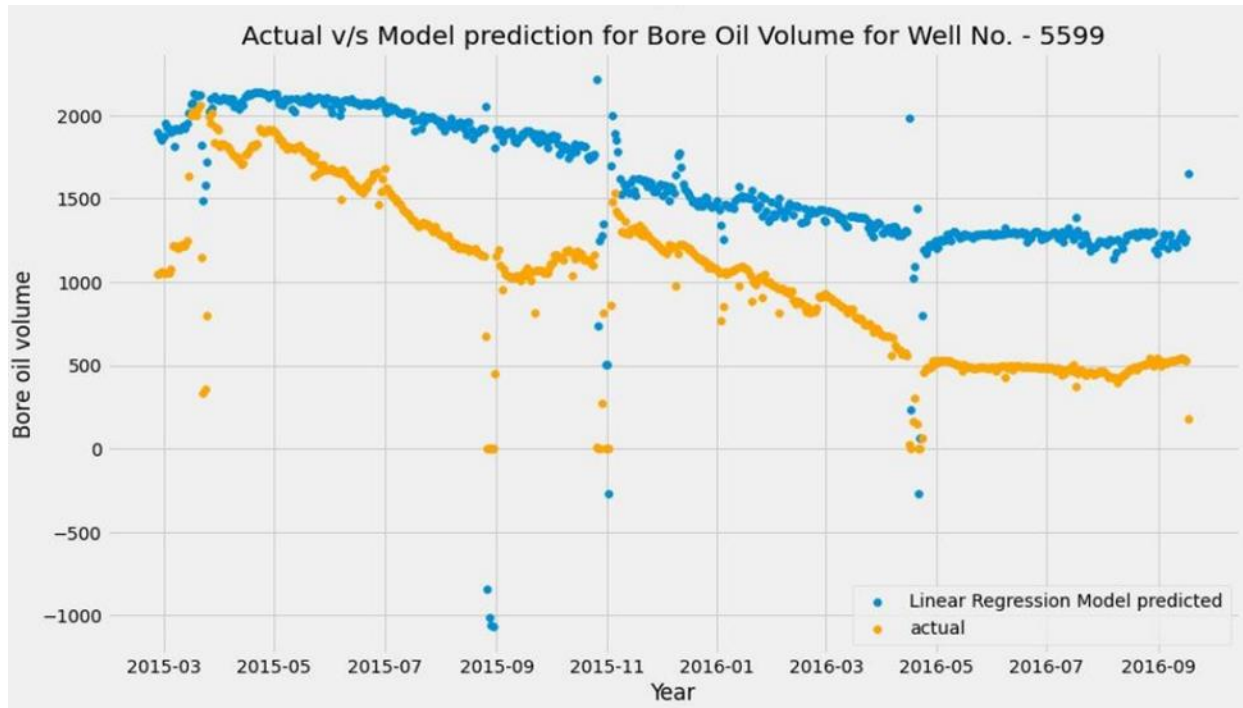


Figura 4.9. Modelo de predicción actual vs modelo por progresión lineal

Tomado de: Alharbi, R. et al. 2023. *Prediction of Oil Production Through Linear Regression Model and Big Data Tools.*

Los resultados confirman la precisión del modelo utilizado en el estudio, ya que el error absoluto medio es cero (MSE: 0.2, RMSE: 0.3). Esto refuerza las ventajas del método empleado por los investigadores.

Por lo que el modelo de regresión lineal utilizado en este estudio incorpora una amplia gama de características, incluyendo volúmenes, temperatura, presión y tamaño del estrangulador, lo que mejora la comprensión del sistema. La elección de utilizar únicamente este modelo se debe a su simplicidad y precisión, lo que ha resultado en una predicción exitosa. A diferencia de otros enfoques, los investigadores no han priorizado la importancia relativa de cada característica

individualmente, sino que han preferido un enfoque más integral para desarrollar un modelo general que englobe todas las características del sistema.

4.2.1.4. Inteligencia artificial para predecir el IPR de pozos verticales en yacimientos derivados de gas en solución: un nuevo enfoque⁴

Objetivo

Predecir la IPR de pozos petroleros impulsados por gas en solución mediante el estudio de las técnicas de red de propagación (BPN) y lógica difusa (FL).

Justificación

La anticipación del Índice de Rendimiento de Pozos (IPR) desempeña un papel crucial en la identificación de un plan de producción óptimo, en la concepción del equipamiento de producción y en la selección adecuada del método de levantamiento artificial. Por consiguiente, se requiere la implementación de un método veloz y confiable que posibilite la predicción precisa del IPR en pozos petroleros impulsados por gas en solución.

Datos

Los datos provienen de pruebas de tasa real en pozos verticales productores de petróleo operando por debajo de la presión del punto de burbuja. Se recopilan datos de diversas estructuras de pozos y propiedades de yacimientos e incluyen caudales de petróleo, agua y gas, presión del yacimiento, densidad API del petróleo, presión de fondo de pozo y caudales de petróleo correspondientes.

Metodología.

El artículo presenta el uso de dos modelos de inteligencia artificial.

⁴ Salem, B. et al. 2018. Using Artificial Intelligence to Predict IPR for Vertical Oil Well in Solution Gas Derive reservoirs: A New Approach.

- **Primer modelo: Red neuronal artificial (ANN)**
 1. **Estructura del primer modelo (ANN):** Se construye el modelo IPR con dos capas ocultas. La primera capa consta de tres neuronas para las entradas donde las entradas son la presión de flujo de fondo del pozo (P_{wf}), el caudal de gas (Q_g) y la densidad del aceite (ρ_o) y la segunda capa contiene la salida, que es el caudal de aceite (Q_o).
 2. **División de los datos:** Los datos se dividen en tres grupos: 70% para entrenamiento, 15% para validación y 15% para probar el modelo.
 3. **Comparación del modelo:** Se compara el modelo de la curva IPR con el método de Fetkovich para evidenciar cual genera mejores resultados.
 4. **Resultados:** Se gráfica y se observa el modelo.
- **Segundo modelo: Lógica Difusa (Sistema de interferencia Neuro-Fuzzy Adaptivo)**
 1. **Construcción del segundo modelo (ANFIS):** Se crea la lógica difusa bajo el nombre de Sistema de Inferencia Neuro-Fuzzy Adaptativo (ANFIS).
 2. **Comparación de resultados:** Se comparan los resultados con el modelo ANN y el Método de Fetkovich
 3. **Proceso de entrenamiento y análisis de resultados:** Se contrasta la fecha prevista con los datos reales en el proceso de evaluación del modelo y se traza el entrenamiento y las pruebas.

Resultados

En el estudio, se compararon dos modelos de inteligencia artificial, la red neuronal artificial Back Propagation Network (BPN) y el sistema de inferencia difusa adaptativa Fuzzy Logic (FL), junto con correlaciones empíricas, para mostrar el rendimiento del flujo de entrada de

pozos petroleros verticales con flujo bifásico. Se utilizaron datos de 207 puntos bajo diferentes condiciones de presión en el fondo del pozo, dividiendo 145 puntos para el entrenamiento y el resto para la validación y las pruebas. Los resultados mostraron que el modelo FL tuvo un menor error en datos no observados en comparación con el otro modelo de inteligencia artificial y el modelo Fetkovich, como se puede observar en la Figura 4-10.

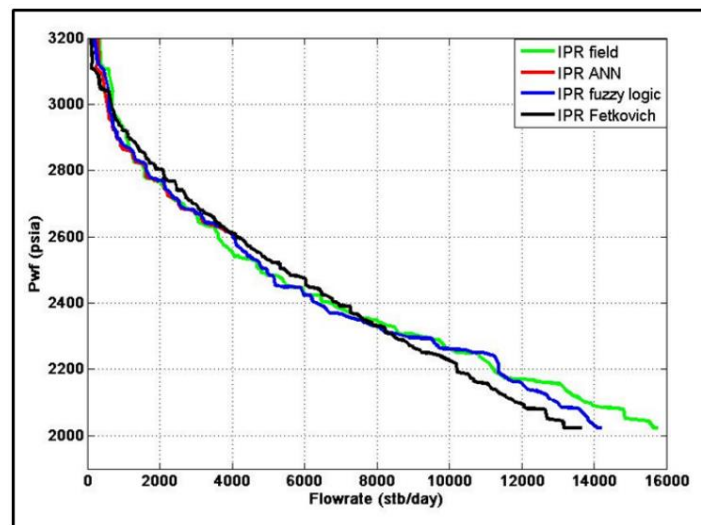


Figura 4.10. IPR de campo comparados con el método de Fetkovich y AI

4.2.2. Ejemplos ingeniería del gas

4.2.2.1. Gestión de la integridad de las tuberías de CO₂: Un enfoque gemelo digital⁵

Objetivo

Mitigar los posibles riesgos de los gasoductos existentes para el transporte de CO₂ mediante un enfoque de gemelo digital

⁵ Sun, T. 2023. CO₂ Pipeline Integrity Management: A Digital Twin Approach

Justificación

Hoy en día, la industria petrolera tiene como uno de los principales retos, buscar la forma de reducir las emisiones de gases de efecto invernadero. La EPA (protección del Medio Ambiente de Estados Unidos) confirmó que el CO₂ es el principal gas de efecto invernadero y en el segundo lugar se encuentra el CH₄ (Metano). La industria petrolera ve la necesidad de generar estrategias para reducir estas emisiones en el sector de los hidrocarburos y así mejorar la eficiencia de este. En este sentido, se está explorando la posibilidad de crear o renovar gaseoductos, ya sea en entornos marítimos o terrestres. Sin embargo, la viabilidad de estos proyectos se ve obstaculizada por los elevados costos asociados y la huella de carbono resultante. Por lo tanto, se está buscando construir, a partir de gemelos digitales, una solución que optimice la eficiencia y seguridad en los procesos industriales de CO₂.

Datos

El gemelo digital basado en la nube para la gestión de integridad de las tuberías de CO₂ recibe datos de diversas fuentes, como sensores (Aceleración del sensor de movimiento, tensión, tensión y compresión absoluta a partir de galgas extensométricas submarinas, perfil de corriente de un perfilador acústico de corrientes de caída (ADCP), datos de ondas del radar de ondas, sensor de presión, sensor de temperatura, caudalímetro), sistemas de control y datos históricos.

Metodología

- **Diseño de tuberías para el transporte de CO₂**

Los factores a considerar en los gasoductos de CO₂ son: las propiedades termodinámicas, presencia de impurezas en el CO₂, fatigas por corrosión y fracturas en la tubería. Estas requieren una atención especial en comparación con los de gas natural. Para asegurar la eficiencia del transporte de

crudo a través de la tubería, es crucial monitorear y controlar de manera precisa las condiciones de temperatura y presión durante la transición de fase.

- **Concepto de Digital twin – Gemelos digitales**

El sistema de gemelos digitales consta de tres componentes principales, que se muestran en la figura 4-11:

El primer componente es un modelo computacional: Este se actualiza en función de los diferentes datos de los sensores de campo

El segundo componente es el sistema IoT/sensor: Este sistema IoT integra la computación en la nube y análisis de grandes datos en sistemas de tuberías submarinas donde la robustez, la fiabilidad y la seguridad son muy deseadas.

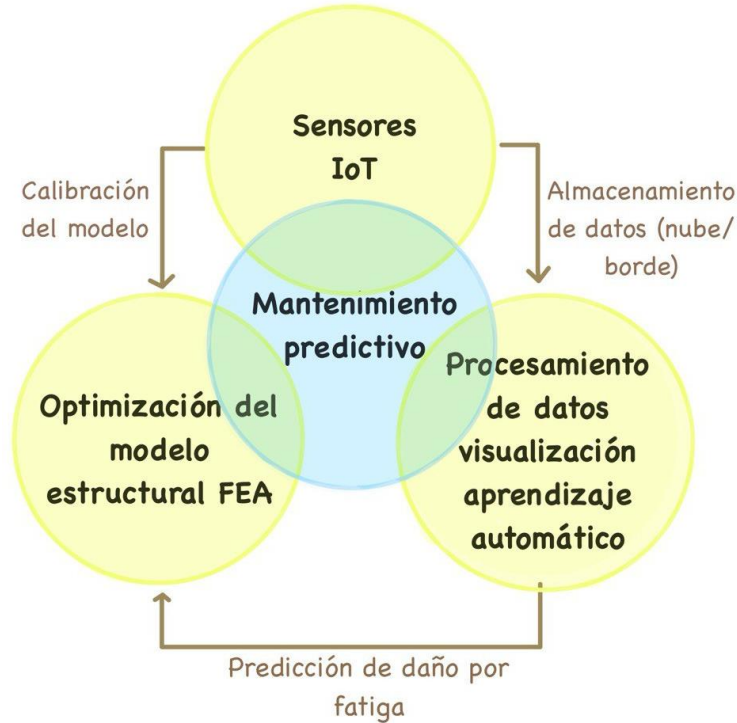


Figura 4.11. Bloques de construcción de gemelo digital

Tercer componente es el análisis de datos: Utilizado para descubrir información dentro de las mediciones y aplicar el algoritmo de aprendizaje automático con el objetivo de determinar la vida útil restante. Es decir, transformara los datos en decisiones procesables para el mantenimiento predictivo mediante el uso del aprendizaje automático.

- **Gemelo digital para tuberías de CO2**

El modelo digital basado en la nube recopila información de diversas fuentes, que incluyen sensores, sistemas de control y datos históricos. centraliza estos datos en una plataforma basada en la nube para llevar a cabo tanto el método de elementos finitos (FEM) como los modelos basados en datos. Las variables de entrada a considerar para el modelo de elementos finitos (FEM- Finite Element Model) en alta mar es la frecuencia de oleaje, la corriente, la presión y la temperatura. El resultado de ambos modelos puede predecir la vida útil restante a la fatiga y evitar que el sistema de tuberías de CO2 sufra graves daños y tambien optimiza los costes para lograr un enfoque rentable

El concepto de gemelo digital basado en la nube para la gestión de la integridad de las tuberías de CO2 aplica la metodología de:

- **Gestión de la integridad de las tuberías de CO2**

Las frecuencias naturales medidas se comparan con las frecuencias naturales calculadas a partir del modelo FE y se ajusta el modelo FE, así como las tensiones. El modelo de FE se ajusta y se emplea para verificar el modelo basado en datos utilizando diversas lecturas de sensores como datos de entrada y salida. Los datos medidos de tensiones y aceleraciones se comparan con los resultados simulados del modelo FE de elementos finitos y se utilizan como entrada para el modelo basado en datos ML. Posteriormente, se utiliza un algoritmo de ML para desarrollar un programa de mantenimiento predictivo con el objetivo de prevenir posibles fallos en el futuro.

Resultados

El mantenimiento predictivo mediante el uso de gemelos digitales reducirá significativamente el coste operativo, al implementar una estrategia de mantenimiento optimizada que se sustenta en inspecciones en tiempo real fundamentadas en la evaluación de riesgos.

4.2.2.2. Una nueva correlación para predecir los caudales de gas utilizando inteligencia artificial: un enfoque industrial 4.0⁶

Objetivo

El objetivo del trabajo es desarrollar un modelo de inteligencia artificial que proporcione estimaciones precisas de las tasas de gas en un yacimiento de gas, con el fin de optimizar las actividades de los pozos y yacimientos.

Justificación

En la actualidad, los ingenieros especializados en la gestión y producción de yacimientos confían en las tasas de producción de los pozos como parte fundamental de la optimización de diversas actividades, incluido el monitoreo de los yacimientos. Sin embargo, las tasas individuales de los pozos de gas no suelen estar fácilmente disponibles, lo que requiere su estimación mediante métodos como el uso de medidores de flujo multifásico (MPFM) y análisis de pruebas de pozos. Estos métodos, aunque valiosos, pueden ser costosos, estar sujetos a incertidumbre y ser técnicamente complejos, lo que a su vez puede dar lugar a estimaciones imprecisas de las tasas de gas. Por consiguiente, se busca implementar diversas herramientas de aprendizaje automático, tales como redes neuronales artificiales (RNA), redes funcionales (FN) y sistemas adaptativos de inferencia neurodifusa (ANFIS), con el objetivo de desarrollar modelos de estimación de tasas de gas altamente precisos.

⁶ Kalam, A. et al. 2019. A Novel Correlation to Predict Gas Flow Rates Utilizing Artificial Intelligence: An Industrial 4.0 Approach

Datos

Se recurre a datos reales de producción de un yacimiento de gas seco para elaborar y validar el modelo propuesto. Se ejecutan procedimientos de limpieza y reducción de datos para garantizar la pertinencia y precisión de los parámetros de entrada del modelo, evitando así la formación de correlaciones incorrectas. Este proceso se realiza utilizando una variedad de herramientas de aprendizaje automático con el objetivo de desarrollar modelos exactos para estimar las tasas de gas.

Metodología

Aplicación de RNA: Para construir un nuevo modelo empírico que evalúa el caudal de gas (Q_{gas}) dictado por la presión en boca de pozo, la temperatura aguas arriba, el tamaño del estrangulador y el caudal de condensado y agua puede desglosarse en los siguientes pasos:

- 1. Definición de variables de entrada y capas ocultas:** Las variables mencionadas (presión en boca de pozo, temperatura aguas arriba, tamaño del estrangulador y caudal de condensado y agua) se consideran como neuronas de entrada. Se determina que una capa oculta compuesta por 19 neuronas es óptima después de un análisis de sensibilidad sobre el número de neuronas.
- 2. Selección de funciones de transferencia:** Se elige la función Tan-sigmoide para la capa oculta y la función Lineal para la capa de salida.
- 3. Elección del algoritmo de entrenamiento:** Después de analizar varios algoritmos, se opta por la retropropagación de Levenberg-Marquardt para entrenar el modelo de RNA.
- 4. Conjunto de datos de entrenamiento y prueba:** Se utilizan 1.797 puntos de datos de 7 pozos para entrenar el modelo de red neuronal directa (FFNN), mientras que para probar la correlación se emplean 4.846 puntos de datos de 3 pozos.

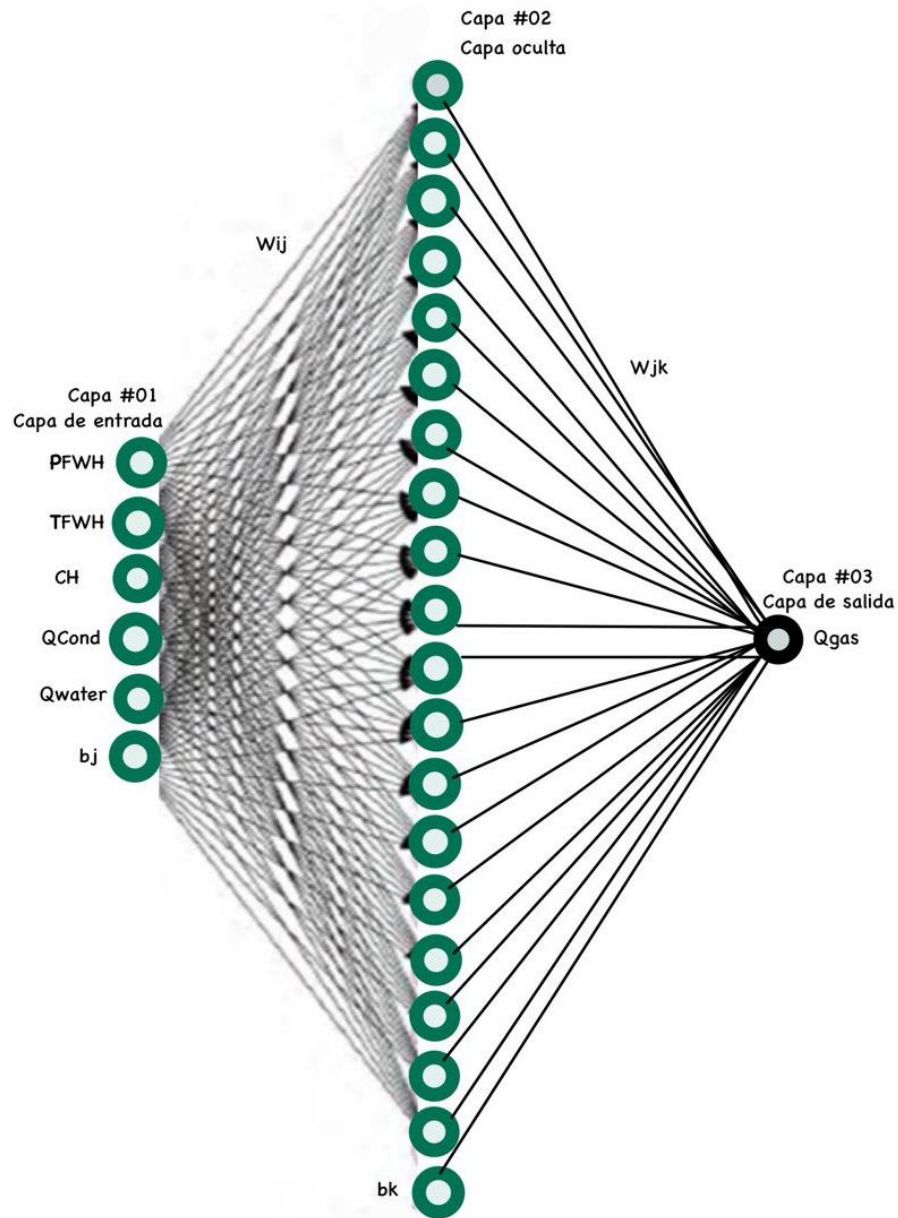


Figura 4.12. Modelo de ANN para predecir los caudales de gas

- 5. Optimización de pesos y sesgos:** Se prueban varias combinaciones de pesos y sesgos mediante múltiples realizaciones para evitar mínimos locales y asegurar una convergencia adecuada del modelo.
- 6. Evaluación del modelo:** El resultado de estas realizaciones se evalúa en función del menor error medio porcentual absoluto, lo que permite juzgar la eficacia y precisión del modelo.

Aplicación del modelo FN: Para predecir el caudal de gas utilizando el algoritmo de aprendizaje de eliminación hacia atrás (BE) se divide en los siguientes pasos:

- 1. Selección de algoritmos de aprendizaje:** Se prueban cuatro algoritmos de aprendizaje diferentes: eliminación hacia atrás (BE), adelante-atrás (FB), búsqueda exhaustiva (ES) y atrás-adelante (BF). Después de la evaluación, se determina que el algoritmo de eliminación hacia atrás (BE) produce el mejor modelo.
- 2. Definición de variables de entrada y objetivo:** Se fija el caudal de gas como parámetro objetivo del modelo, mientras que la presión en boca de pozo, la temperatura aguas arriba, el tamaño del estrangulador y el caudal de condensado y agua se establecen como variables de entrada.
- 3. Conjunto de datos de entrenamiento y prueba:** Se utiliza el mismo conjunto de datos tanto para entrenar como para probar el modelo FN BE.
- 4. Evaluación del modelo:** El resultado de las realizaciones se evalúa utilizando el menor error porcentual absoluto medio y el mayor coeficiente de determinación (R^2). permitiendo desarrollar y evaluar el modelo FN BE para predecir el caudal de gas con una precisión aceptable, utilizando un enfoque de aprendizaje de eliminación hacia atrás.

Desarrollo de modelo Sistema de inferencia neurofuzzy adaptivo (ANFIS): Para predecir el caudal de gas utilizando el algoritmo de aprendizaje Subtractive Clustering (SC) puede desglosarse en los siguientes pasos:

- 1. Descripción de ANFIS:** ANFIS es una combinación de redes neuronales artificiales (RNA) y lógica difusa, desarrollada a finales de la década de 1990. Esta técnica tiene una amplia aplicación en diversos campos de la ingeniería.

2. **Selección del algoritmo de aprendizaje:** Se prueban dos algoritmos de aprendizaje diferentes: Subtractive Clustering (SC) y Grid Partitioning (GP). Después de la evaluación, se determina que el algoritmo SC produce un modelo superior.
3. **Definición de variables de entrada y objetivo:** El caudal de gas (Qes) se predice en función de la presión y la temperatura del cabezal del pozo, el tamaño del estrangulador y el caudal de condensado y agua
4. **Evaluación del modelo:** El resultado de múltiples realizaciones se evalúa utilizando el menor error medio porcentual absoluto y el mayor coeficiente de determinación (R2). Permitiendo desarrollar y evaluar el modelo ANFIS-SC para predecir el caudal de gas con una precisión aceptable, utilizando un enfoque de aprendizaje de Subtractive Clustering.

Resultados

Los investigadores realizaron una comparación rigurosa entre los diferentes modelos de inteligencia artificial utilizando métricas como el error porcentual absoluto medio (AAPE) y el coeficiente de determinación. Los resultados muestran que el modelo de RNA es capaz de predecir las tasas de gas con una precisión superior al 90%, lo que representa un avance significativo en comparación con los métodos tradicionales.

4.2.3. Ejemplos facilidades de superficie

4.2.3.1. Aplicación de técnicas de computación blanda a la medición de flujo multifásico: una revisión⁷

⁷ Yan, yong. Et al. 2018. Application of soft computing techniques to multiphase flow measurement: A review

Objetivo

Revisión de técnicas de computación blanda para la medición de flujo multifásico, centrándose particularmente en la medición de tasas de flujo.

Justificación

Las mezclas de petróleo/gas/agua representan flujos de dos o tres fases, ya sea gas-liquido o liquido-liquido, más comunes durante los procesos de producción, transporte y transferencia de custodia en la industria del petróleo y el gas. La importancia de medir el flujo multifásico en esta industria radica en que esta medición suministra información esencial para evaluar y mejorar el rendimiento de los pozos de extracción de hidrocarburos. El éxito al medir los flujos volumétricos de las fases depende de la eficiencia en su separación. La precisa cuantificación del flujo de cada fase de producción permite a los operadores tomar decisiones más fundamentadas sobre el desempeño de los pozos, ya que los ingenieros pueden identificar, comprender y abordar problemas asociados con el flujo de pozos múltiples, optimizar las operaciones de levantamiento artificial y desarrollar modelos dinámicos más precisos de los yacimientos. Con el avance reciente de la inteligencia artificial y el aprendizaje automático, las técnicas de computación blanda ofrecen enfoques alternativos a los métodos estadísticos convencionales, ampliando así las capacidades de los modelos empíricos.

Metodología

La computación blanda comprende un conjunto de enfoques que busca utilizar la tolerancia a la imprecisión y la incertidumbre con el fin de lograr manejabilidad, robustez y soluciones de bajo costo. En ocasiones, se emplea el termino de ‘’computación blanda’’ de manera intercambiable con ‘‘sensores blandos’’ o ‘‘sensores virtuales’’. Sensor suave es un nombre común para una pieza de software que se utiliza para derivar información deseable a partir de mediciones disponibles.

- **Aplicaciones de técnicas de computación blanda**

La aplicación de técnicas de computación blanda en la medición de flujo multifásico se enfoca principalmente en la estimación de tasas de flujo, fracciones de fase y la identificación del régimen de flujo. A continuación, se describen sistemas de medición multifásica indirecta que integran sensores convencionales junto con enfoques de computación blanda:

- **Sensores ultrasónicos**

Cuatro sensores ultrasónicos han sido implementados con una Red Neuronal Artificial (ANN) para reconocer el patrón de flujo y calcular la fracción de volumen de gas en flujos de dos fases, respectivamente. La entrada de la ANN se compone por las proporciones de energía por cuatro sensores acústicos, la capa de salida produce el patrón de flujo identificado o la estimación de la fracción de volumen de gas.

- **Dispositivos de presión diferencial**

Se presenta una metodología innovadora para la medición del flujo de gas húmedo utilizando un medidor Venturi con garganta extendida y técnicas de aproximación de computación blanda. Se ha creado un modelo de ANN de retropropagación (BP – ANN) y un modelo SVM para estimar el caudal de gas y el caudal de líquido a través de las características estáticas y dinámicas de las presiones diferenciales. En la figura # se observa el sistema de medición.

- **Sensores eléctricos**

El enfoque propuesto en el artículo, basado en una Red Neuronal Artificial (ANN), está diseñado para calcular los flujos de gas y líquido en un sistema bifásico de aire-agua en una tubería horizontal. Se identificaron cinco parámetros característicos del modelo mecanicista de flujo de slug, incluida la

velocidad de traslación, la retención de slug, la retención de película, la longitud de slug y la longitud de la película y se tomaron como entradas a la red neuronal. Se implementó una red neuronal de retroalimentación con diez nodos ocultos para predecir los caudales de gas y líquido.

- **Caudalímetros Coriolis**

Las redes neuronales, como la de función de base radial y el perceptrón multicapa, reciben cuatro variables de entrada que incluyen temperatura, amortiguación, caída de densidad y caudal para estimar el error del flujo másico. Luego, las lecturas de flujo másico del medidor de flujo Coriolis se corrigen con el error de flujo másico estimado.

Resultados

Las técnicas presentadas anteriormente son superiores a los métodos convencionales para resolver problemas desafiantes que son difíciles de describir mediante modelos analíticos o matemáticos. Al incorporar estas técnicas de computación blanda, los sensores ultrasónicos, los caudalímetros Venturi, los sensores de conductancia y los caudalímetros Coriolis son capaces de medir los caudales individuales o las fracciones de fase en un flujo bifásico o trifásico.

4.2.3.2. Detección de fugas en gasoductos de gas natural mediante modelos inteligentes⁸

Objetivo

Estudiar la capacidad de los modelos inteligentes para detectar pequeñas fugas en un gasoducto de gas natural.

⁸ Akinsete, O. 2019. Leak Detection in Natural Gas Pipelines Using Intelligent Models

Justificación

Los gasoductos son la forma más eficiente y rentable de transportar gases inflamables, especialmente el gas natural, a distancias cortas y medias. Sin embargo, los métodos tradicionales de detección de fugas en estas tuberías no cumplen con los estándares de la industria en cuanto a sensibilidad, confiabilidad, robustez y precisión. Por lo tanto, se requiere la implementación de herramientas avanzadas como el análisis de Big Data, modelos de inteligencia artificial y un enfoque transitorio en tiempo real para mejorar la detección de fugas en los conductos de gas y así evitar incidentes en la industria.

Datos

Los investigadores recolectaron datos históricos utilizando mediciones SCADA de variables como el tipo de funcionamiento de la tubería, medidores de presión, temperatura, longitud, diámetro de tubería y caudal, tomadas tanto en la entrada como en la salida, durante un lapso de cuatro meses. Estas mediciones se realizaron a intervalos de dos minutos, generando un conjunto total de alrededor de 100,000 puntos de datos.

Metodología

El problema se abordó como un problema jerárquico que combina regresión y clasificación, donde un modelo inteligente funciona como regresor y un algoritmo de detección de fugas actúa como clasificador.

En base a lo anterior se debe tener en cuenta:

simulación de fugas: Aquí se empleó la ecuación desarrollada por Scott. Et al. (1999) para representar una fuga de gas monofásica en el gasoducto, donde los investigadores validaron esta

ecuación utilizando datos simulados y confirmaron su validez una vez que pasó el período transitorio inicial de la fuga. Por tal motivo, desarrollaron una función para simular la fuga en la tubería en función del tiempo t , basándose en esta relación.

$$q = CF_L(P_{in}^2 - P_{out}^2)^n$$

Donde:

- Q : es el caudal de gas de salida en metros cúbicos/segundos
- C : es una constante de la tubería
- F_L : es la reducción de la eficiencia debida a una fuga (es 1 cuando no hay fuga) y n suele ser 0,5. Esta se puede expresar de la siguiente forma:

$$F_L = [1 + L_{LD}(Q_{LD}^2 + 2Q_{LD}^2)]^{-n}$$

Detección de fugas: Se emplea un análisis residual donde un modelo inteligente actúa como observador de la tubería, anticipando el flujo de salida con parámetros operativos clave. Un algoritmo de detección de fugas colabora con el modelo para identificar fugas cuando un indicador específico supera un umbral, evitando falsas alarmas mediante una señal constante durante varios conjuntos de datos previos.

Índice de fuga: Se realiza el monitoreo de los excesos de flujo durante la hora previa, contando la cantidad de excesos (a) que exceden un umbral determinado. Esta cantidad se utiliza para calcular un índice de fugas mediante una fórmula exponencial, que es una variante de la función sigmoidea comúnmente empleada en modelos de regresión logística y redes neuronales. Esta fórmula se expresa como:

$$f(a) = \exp\left(\frac{-1}{1+a^2}\right)$$

Elección del umbral: La elección del umbral viene dada por: $threshold=MAE+0.0$. Donde MAE, conocido como el error absoluto medio del modelo (Mean Absolute Error), se calcula como parte del proceso de entrenamiento y ajuste de los modelos mediante la búsqueda en cuadrícula. En este proceso, se utilizan métricas como el error cuadrático medio (RMSE), el error absoluto medio (MAE) y el coeficiente de determinación (puntuación R2). Los modelos empleados en este estudio incluyen:

- Refuerzo por gradiente (GB)
- Bosque aleatorio (RF)
- Árbol de decisión (DT)
- Red neuronal artificial (RNA)
- Máquina de vectores de soporte (SVM)

Métricas de rendimientos de la detección de fugas: Los investigadores seleccionaron las métricas de rendimiento de detección de fugas basándose en las métricas aplicables definidas por API, 1995b. Estas métricas son:

- Sensibilidad, que se refiere a la mínima fuga detectable y al tiempo requerido para que el sistema emita una alarma en caso de que ocurra una fuga.
- Fiabilidad, que se relaciona con la capacidad del sistema para ofrecer resultados precisos sobre la posible existencia de una fuga en una tubería.
- Precisión, que indica la capacidad del sistema para estimar con exactitud el caudal de la fuga.
- Robustez, que se refiere a la capacidad del sistema para continuar funcionando y proporcionar información útil incluso en condiciones cambiantes de operación del oleoducto.

Entrenamiento: El entrenamiento inicial se realizó en lotes, seguido de la incorporación de datos adicionales validados de SCADA para mejorar y robustecer el modelo. La elección del entrenamiento por lotes se debió a limitaciones de hardware, aunque el entrenamiento en línea sería preferible en su ausencia. Se asignó el 60% de los datos para el entrenamiento, reservando el 20% para conjuntos de validación y prueba, respectivamente.

Resultados

Los resultados obtenidos indicaron un alto potencial, demostrando un desempeño notablemente satisfactorio del modelo inteligente en comparación con un modelo transitorio en tiempo real, que actualmente es la opción preferida en la industria para detectar fugas en gasoductos de gas natural.

5. Comparación del procedimiento obtenido en la aplicación de las técnicas de la ciencia de datos respecto a la forma tradicional de los ejemplos seleccionados

La integración de la Ciencia de Datos en la Ingeniería de Petróleos tiene como objetivo primordial disminuir el tiempo no productivo de los ingenieros, permitiéndoles dedicarse a la identificación de conocimientos prácticos y, por ende, descubrir oportunidades subutilizadas para aumentar o mejorar la producción en los campos que supervisan. Esto se logra mediante un proceso de trabajo sistemático que aprovecha los datos disponibles para eliminar la recopilación manual de información, generar resultados visualmente atractivos y facilitar la colaboración con un equipo interdisciplinario. Se fomenta la minería y el análisis de datos con precisión, seleccionando los modelos matemáticos adecuados según las demandas del negocio y los datos disponibles, respaldado todo por una configuración tecnológica óptima. Esta integración facilita la adopción de las prácticas más efectivas y provoca un cambio de perspectiva que mejora la competencia de los ingenieros de campo y consultores, enriqueciendo así su

conjunto de habilidades en una industria en constante evolución y demanda. Esta aplicación conlleva a unas ventajas y desventajas tales como se observan a continuación:

Por ende, en este capítulo, se llevará a cabo una comparación entre el método convencional y la técnica de la ciencia de datos empleada en los ejemplos explicados en el capítulo 4, conforme al

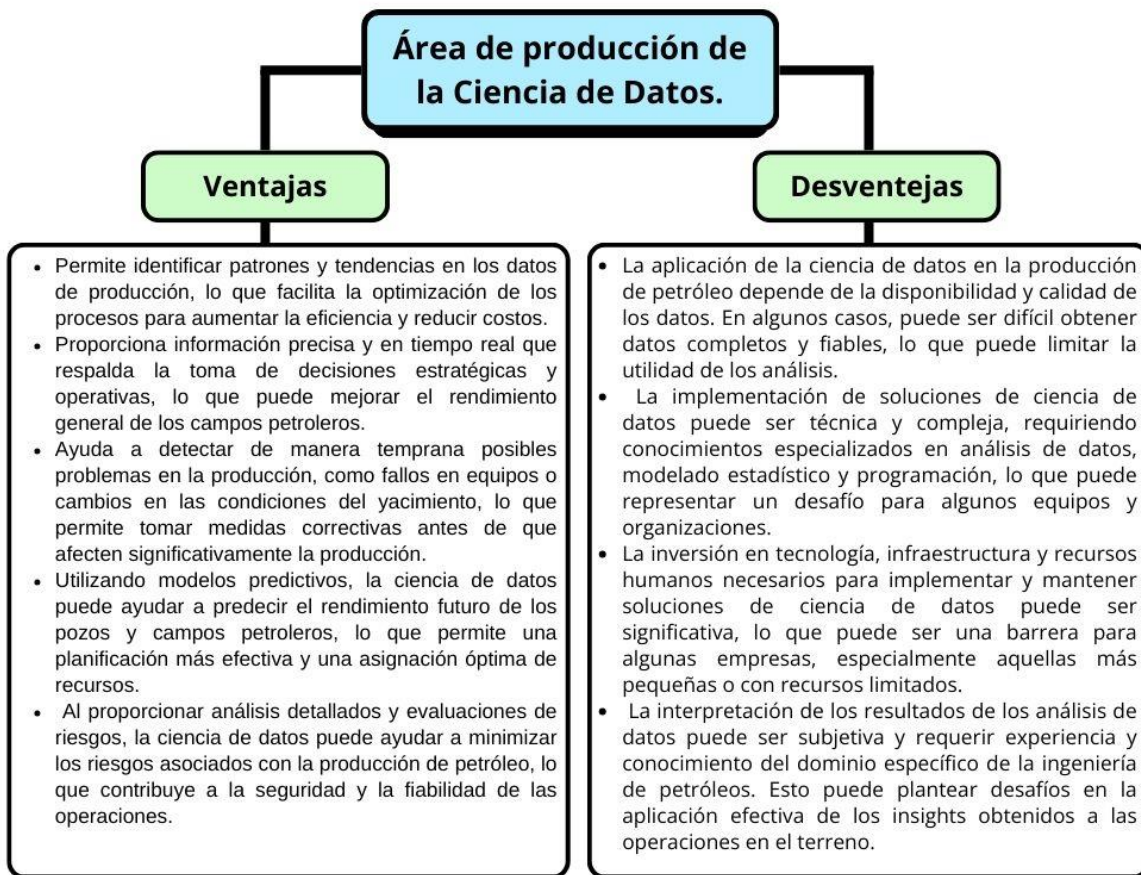


Figura 5.1. Ventajas y desventajas del área de producción de la ciencia de datos.

currículo del programa de Ingeniería de Petróleos de la Universidad Industrial de Santander. Como se mencionó previamente, es importante destacar que, según la revisión bibliométrica realizada, las técnicas más aplicadas son aquellas que se explicaron en la metodología de los ejemplos anteriores. Por consiguiente, se tomará apoyo de las siguientes técnicas para llevar a cabo la comparación correspondiente.

5.1. Un enfoque basado en redes neuronales de inteligencia artificial para predecir la producción en yacimientos no convencionales: análisis comparativo con la curva de declive

- Método de ciencia de datos

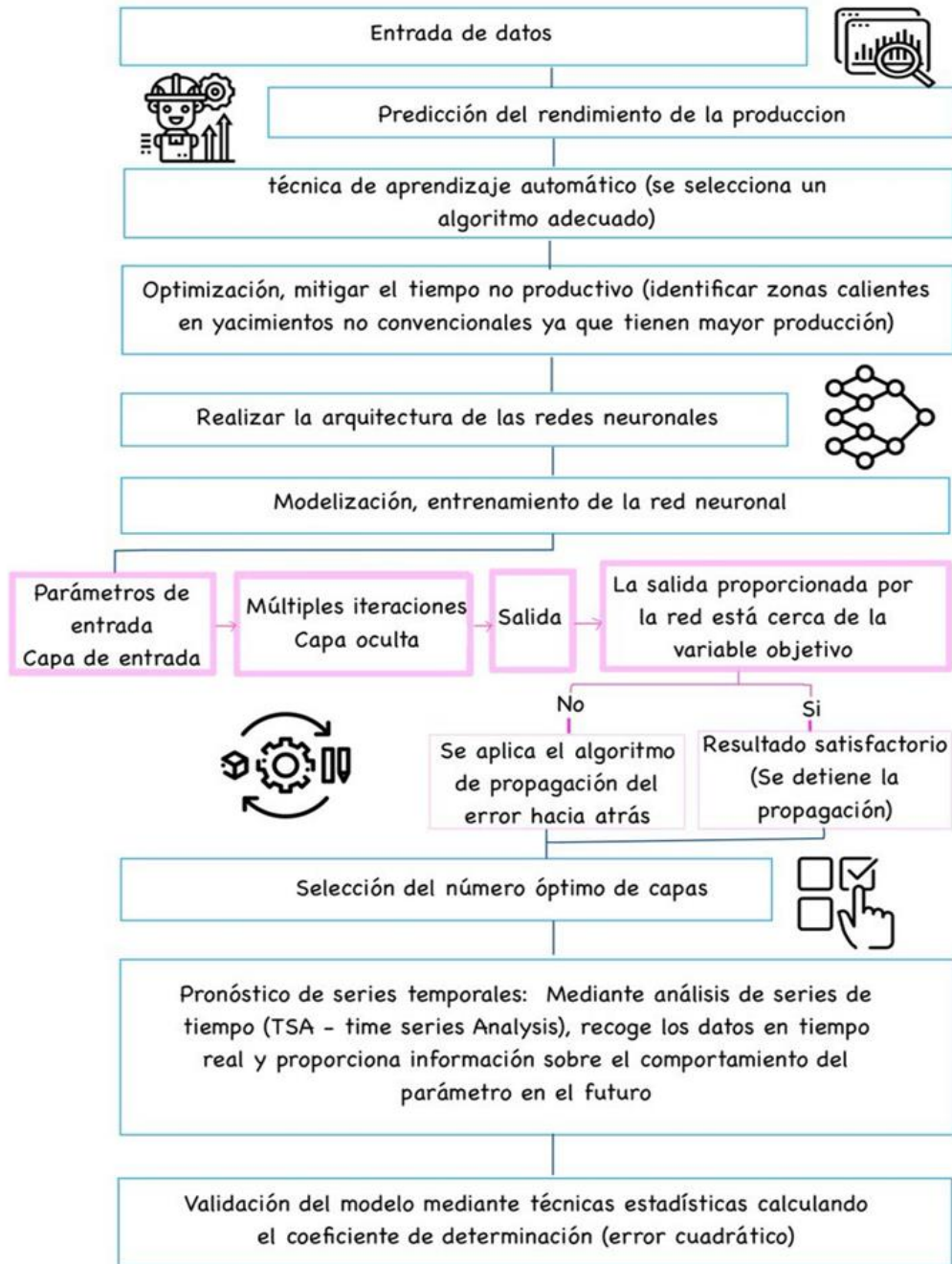


Figura 5.2. Método de ciencia de datos para predecir la producción en yacimientos convencionales

- Método convencional

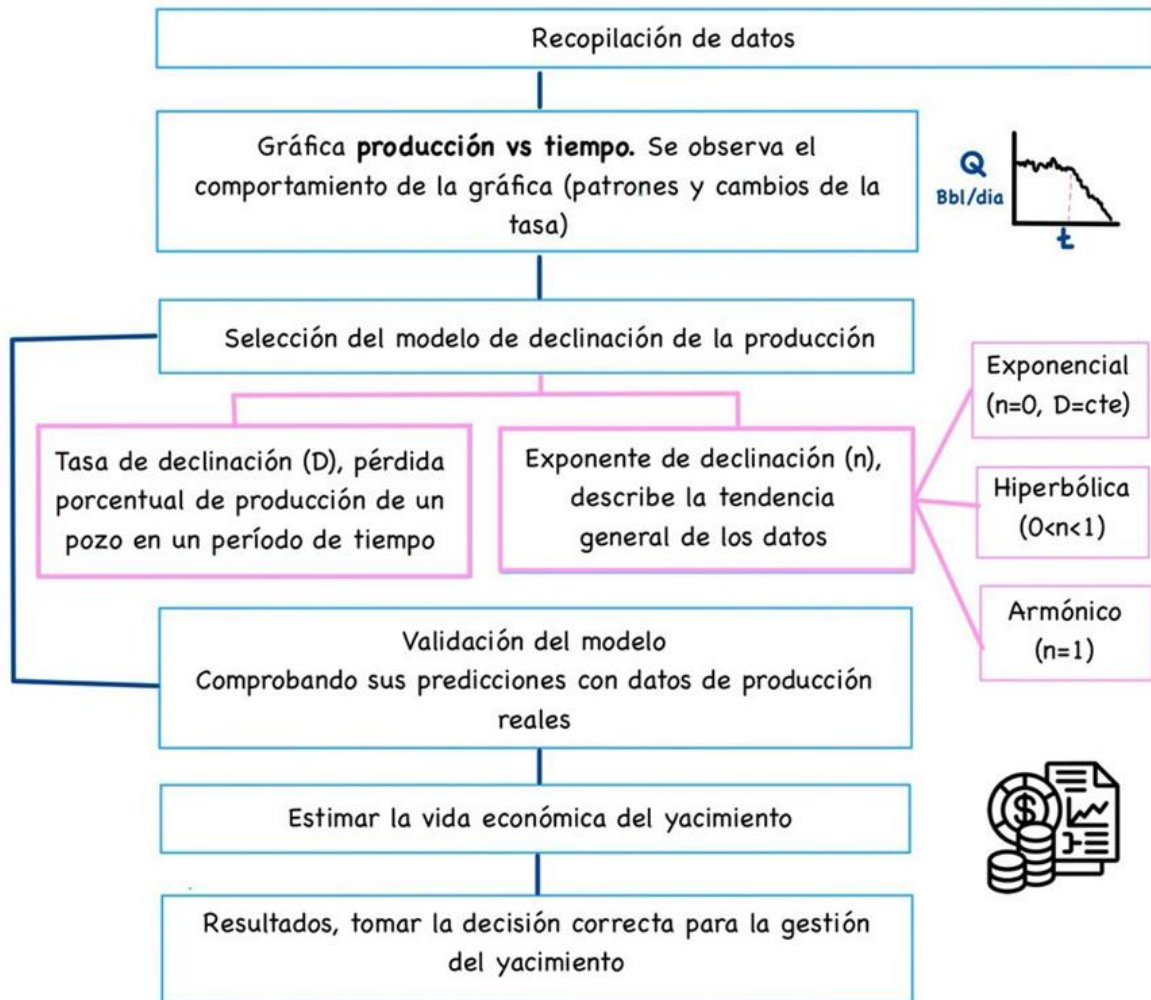


Figura 5.3. Método convencional para predecir la producción en yacimientos convencionales

5.2. Aplicación de técnicas de inteligencia artificial para predecir la productividad de los pozos de espina de pescado

- **método de ciencia de datos**

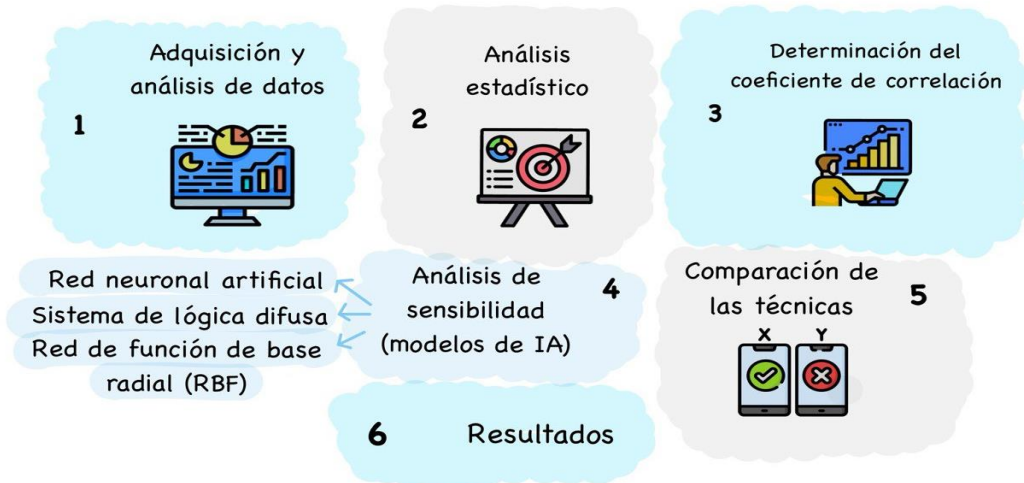


Figura 5.4. Método de ciencia de datos para predecir la productividad de los pozos de espina de pescado

- **Método convencional**

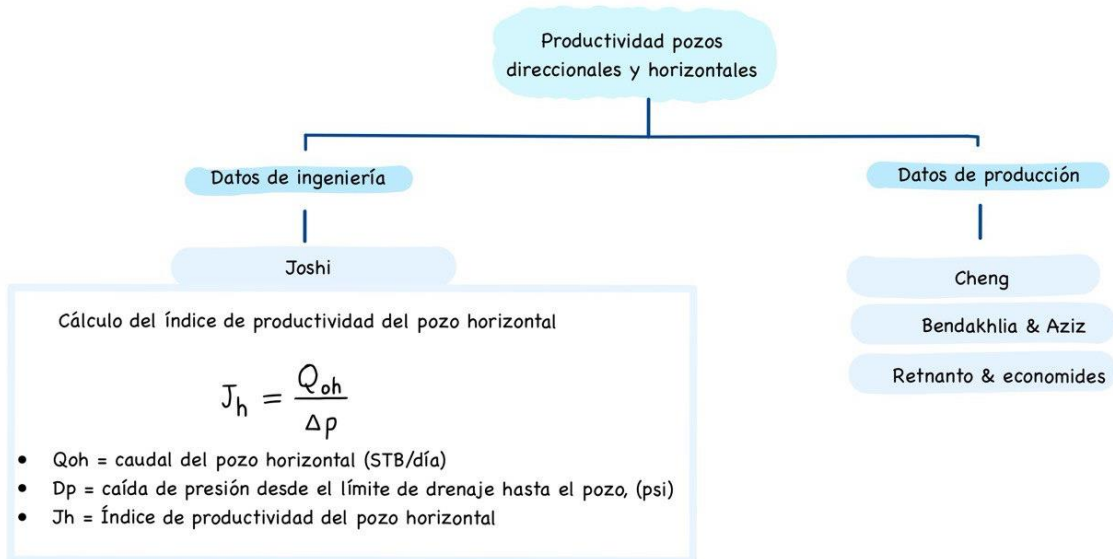


Figura 5.5. Método convencional para predecir la productividad de los pozos de espina de pescado

5.3. Predicción de la optimización del sistema de producción de petróleo a través de modelos de regresión lineal y herramientas Big Data

- método de ciencia de datos y método convencional



Figura 5.6. Método de la ciencia de datos vs método convencional para la predicción

5.4. Inteligencia artificial para predecir el IPR de pozos verticales en yacimientos derivados de gas en solución: un nuevo enfoque

- **Método de ciencia de datos**

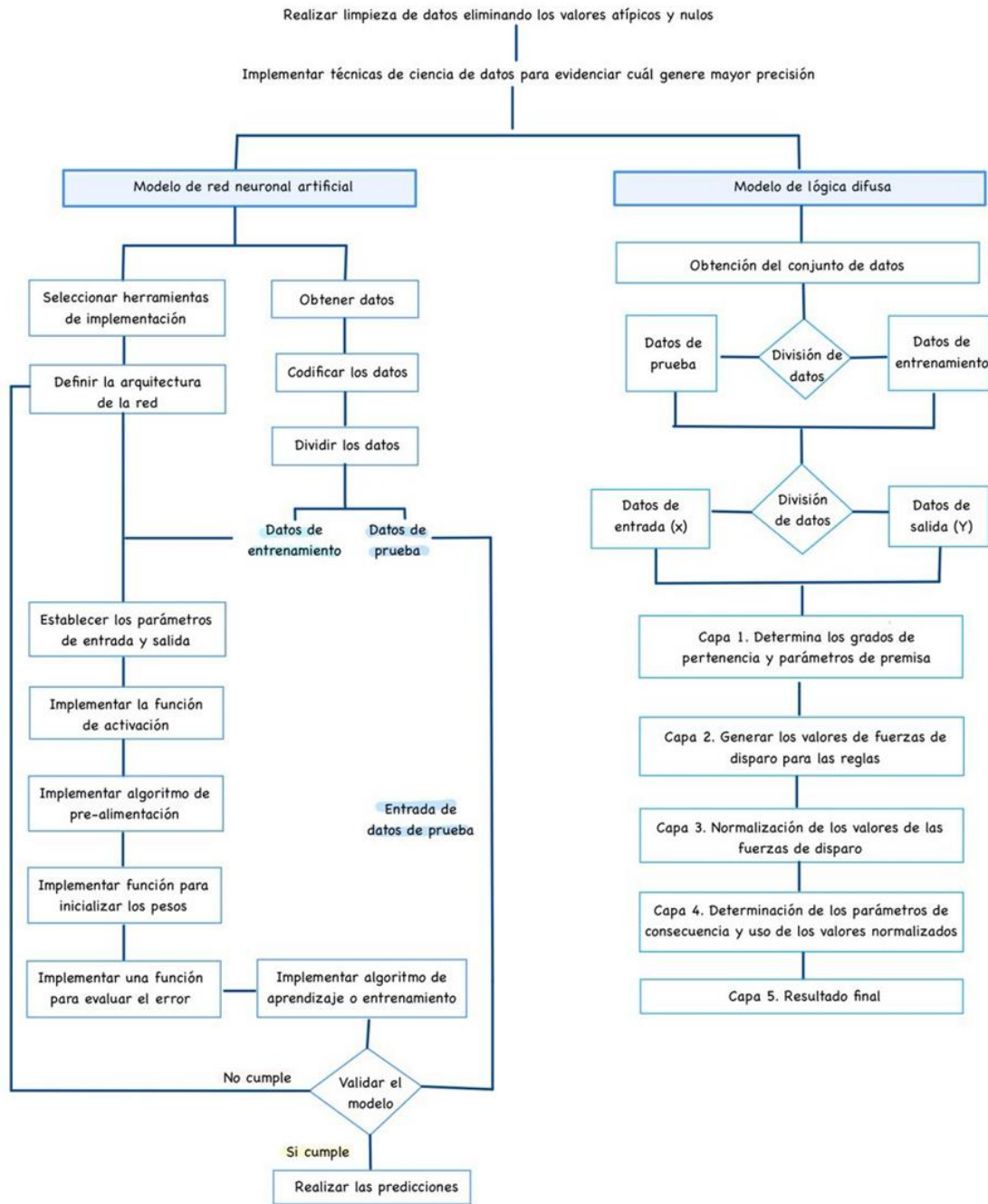


Figura 5.7. Método de ciencia de datos para predecir el IPR de pozos verticales

- método convencional

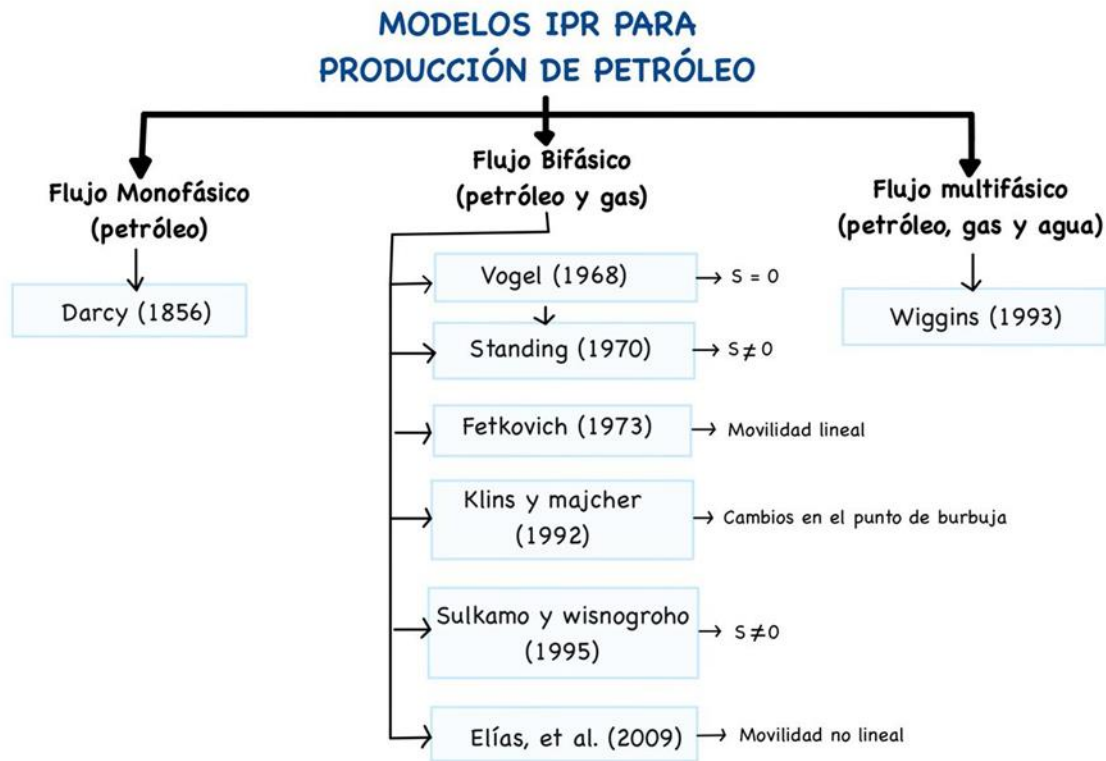


Figura 5.8. Método convencional para predecir el IPR de pozos verticales

5.5. Gestión de la integridad de las tuberías de CO2: Un enfoque gemelo digital

- Método de ciencia de datos

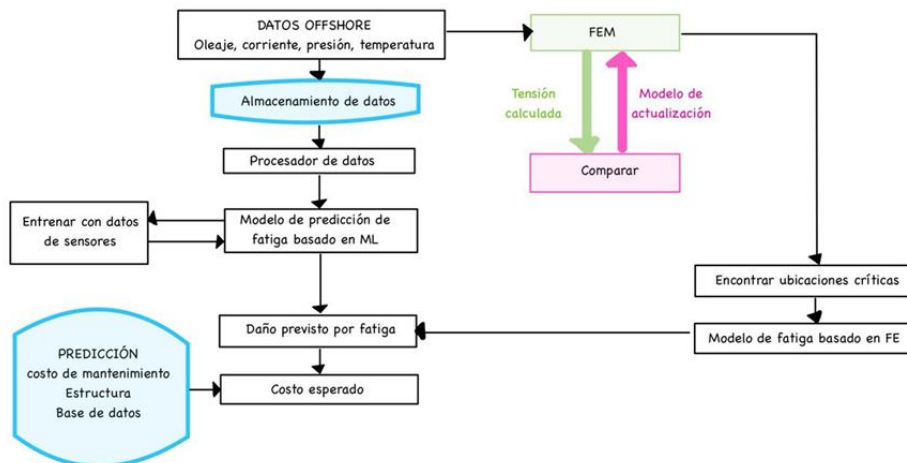


Figura 5.9. Método de ciencia de datos para la gestión de la integridad de las tuberías de CO2

• **Método convencional**

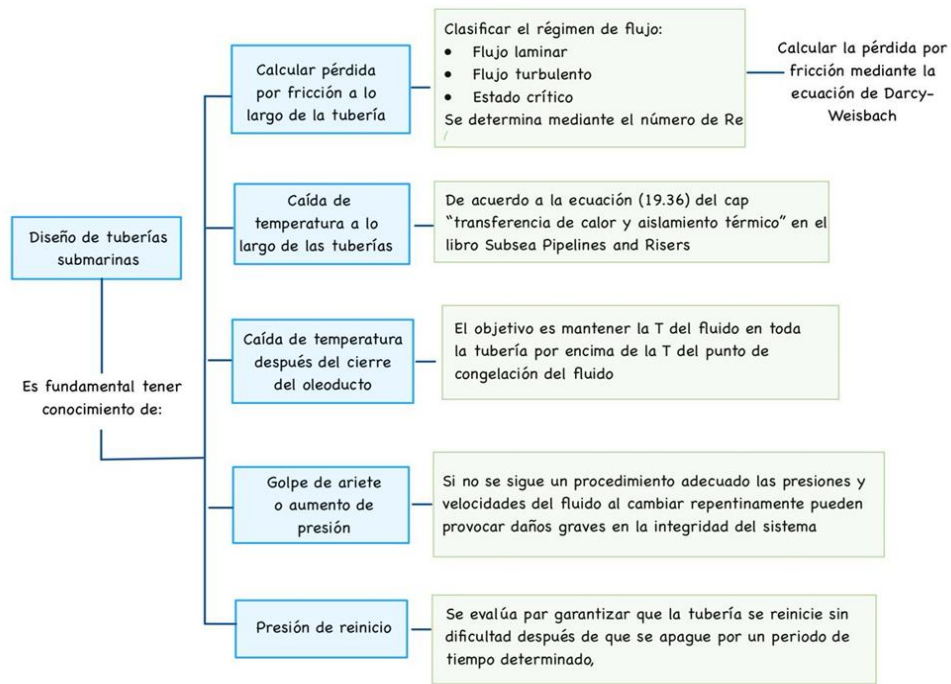


Figura 5.10. IPR de campo comparados con el método de Fetkovich y AI

5.6. Una nueva correlación para predecir los caudales de gas utilizando inteligencia artificial: un enfoque industrial 4.0

• **Método de ciencia de datos**

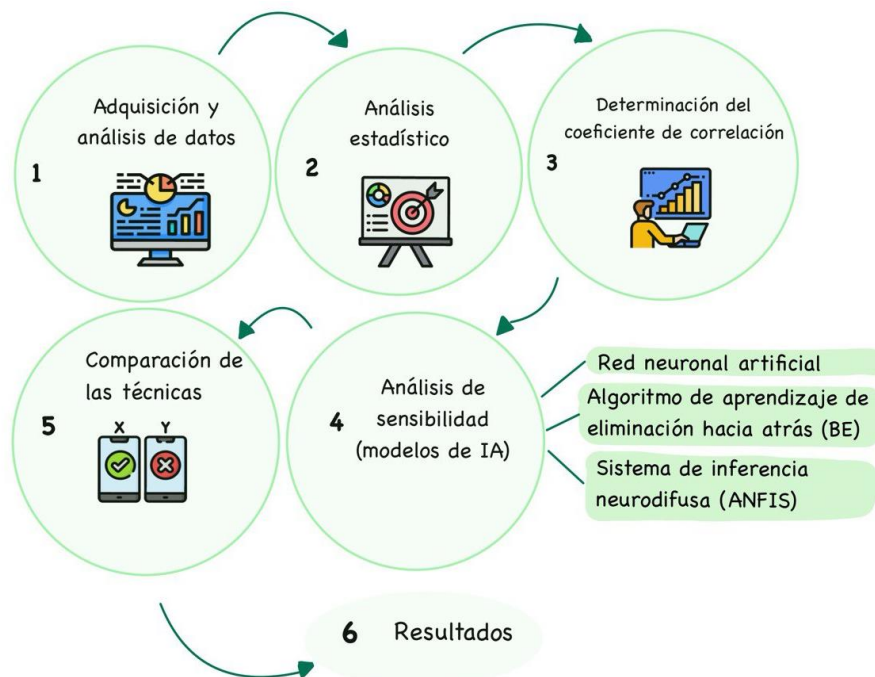


Figura 5.11. Método de ciencia de datos para predecir los caudales de gas

- Método convencional

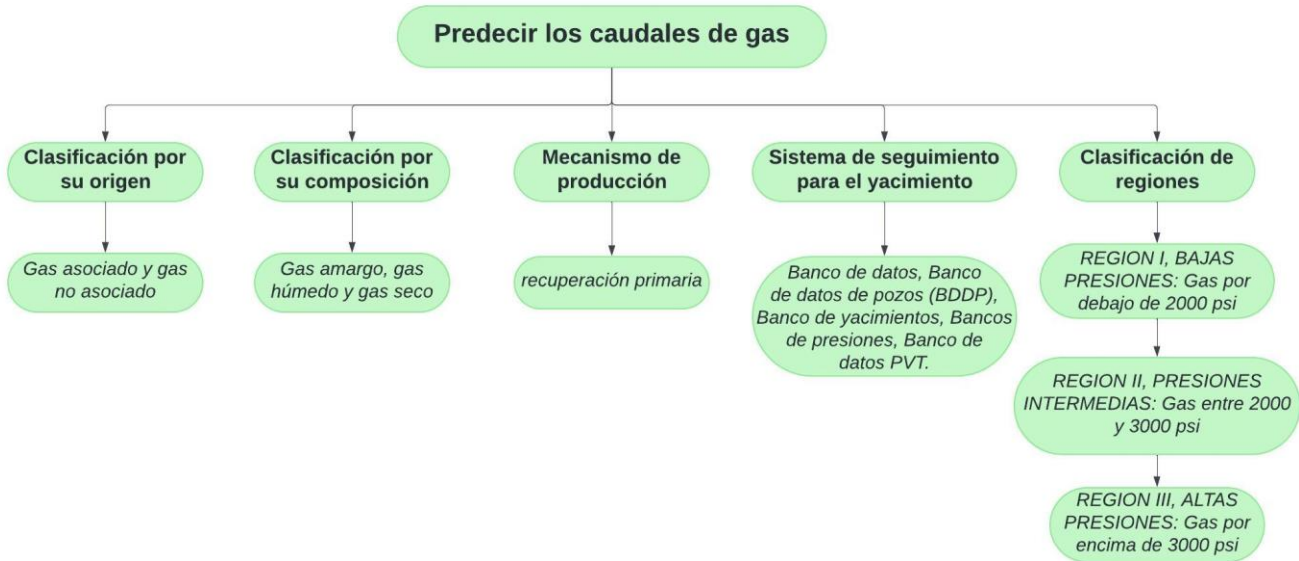


Figura 5.12. Método convencional para predecir los caudales de gas

5.7. Aplicación de técnicas de computación blanda a la medición de flujo multifásico: una revisión

- Método de ciencia de datos

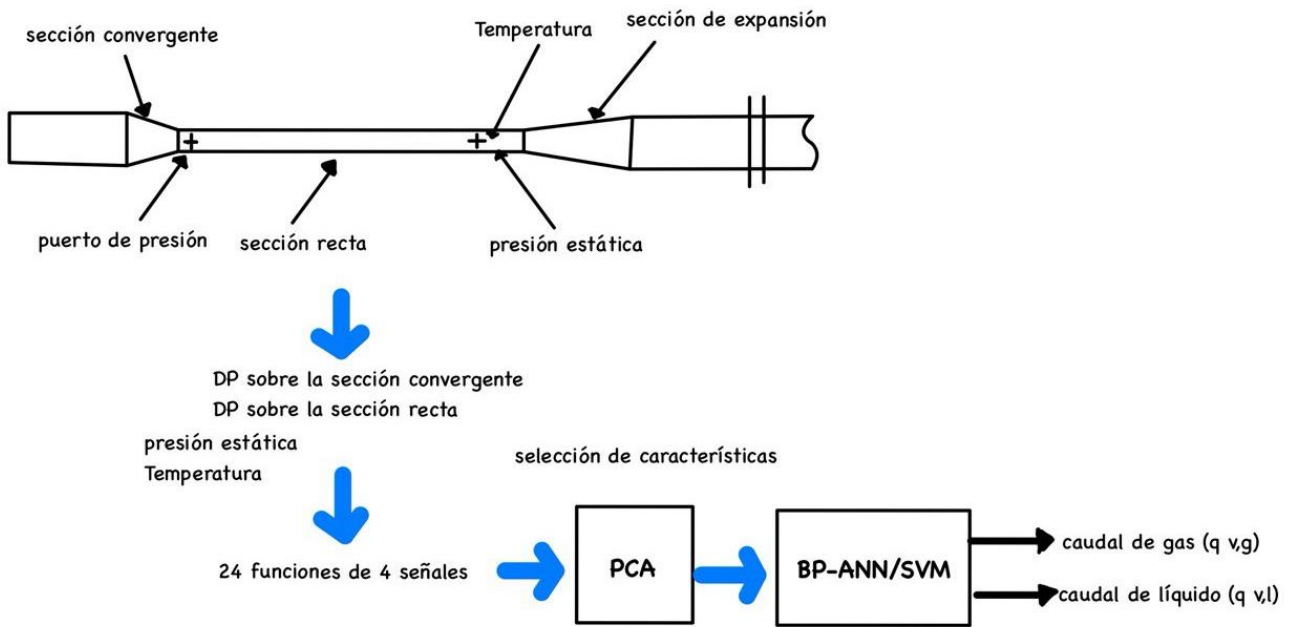


Figura 5.13. Método de ciencia de datos para la medición de flujo multifásico

• **Método convencional**

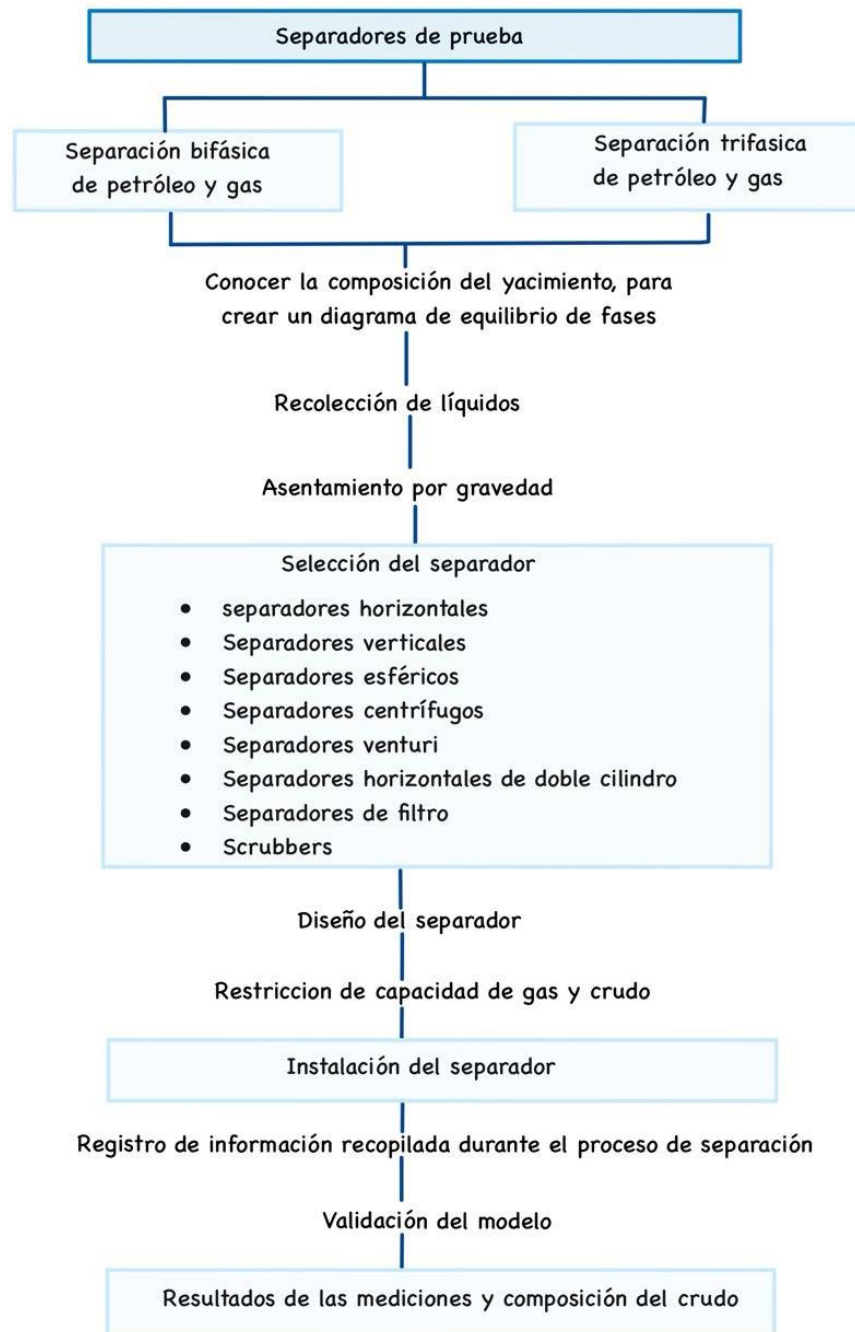


Figura 5.14. Método convencional para la medición de flujo multifásico

5.8. Detección de fugas en gasoductos de gas natural mediante modelos inteligentes

- Método de ciencia de datos

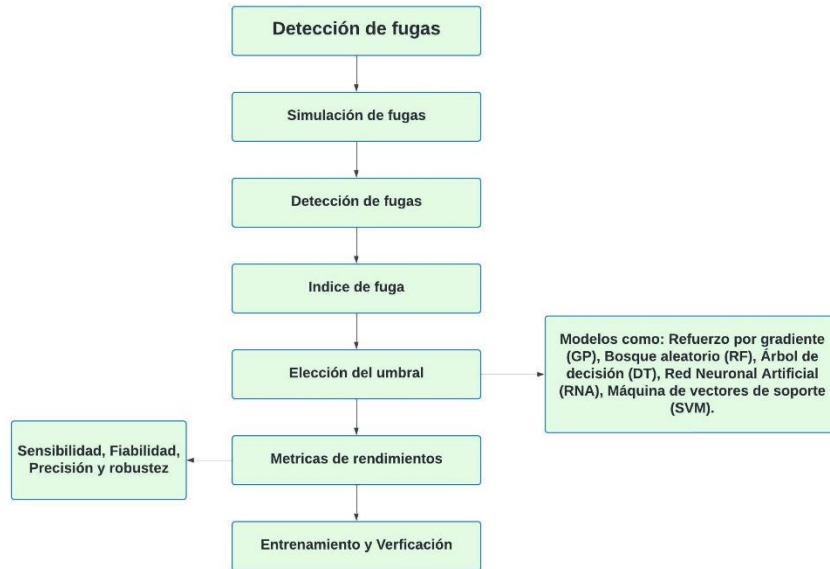


Figura 5.15. Método convencional para detectar fugas en gasoductos

- Método convencional

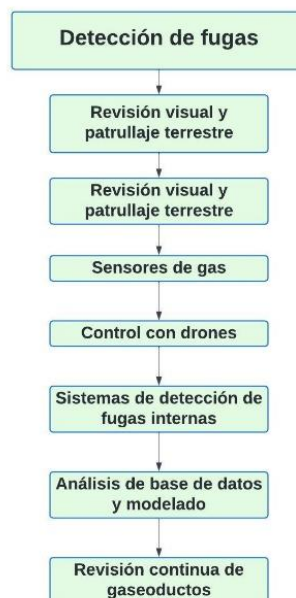


Figura 5.16. Método convencional para detectar fugas en gasoductos

6. Conclusiones

El estudio de la ciencia de datos en la industria petrolera está experimentando un crecimiento significativo, impulsado por la necesidad de soluciones innovadoras para mejorar las operaciones. La bibliometría ofrece un enfoque sistemático para analizar la investigación en este campo, proporcionando información valiosa sobre las tendencias y los avances en el área. Este análisis ayuda a orientar futuras investigaciones y a impulsar el desarrollo de nuevas tecnologías y prácticas en la industria petrolera.

Se concluyó que la integración de la ciencia de datos en la ingeniería de producción de petróleo representa una oportunidad invaluable para mejorar la eficiencia, reducir costos y aumentar la seguridad en las operaciones petroleras. Al permitir la identificación de patrones y tendencias en los datos de producción, la ciencia de datos facilita la optimización de los procesos, lo que conduce a una mayor eficiencia y ahorros significativos.

Se analizó que las técnicas más usadas en la industria petrolera son las técnicas de aprendizaje automático y redes neuronales, ya que, desempeñan un papel fundamental en la optimización de los procesos, la anticipación de posibles fallos, la mejora de la eficiencia y la toma de decisiones en el área de producción de petróleo y gas, debido a su capacidad para procesar grandes volúmenes de datos y extraer información valiosa que permite a las empresas petroleras mejorar su competitividad, reducir costos y maximizar la producción de manera sostenible.

Los resultados de la implementación tanto de la red neuronal artificial como del modelo de lógica difusa fueron evaluados comparativamente con el IPR calculado mediante la correlación de Fetkovich. Se encontró que el error porcentual promedio asociado a cada modelo de IPR fue de 13,5% para Fetkovich, 12,6% para la red neuronal y 1,22% para la lógica difusa. Estos hallazgos sugieren que la técnica de lógica difusa es la mejor opción para abordar este problema, dado que exhibe la menor desviación en relación con los datos reales y el menor error porcentual promedio.

7. Recomendaciones

Se sugiere a la escuela de ingeniería de petróleos la inclusión de una materia o contexto centrado en la aplicación de la ciencia de datos en el área de producción. Esto se debe a que medida que avanza el tiempo se ha evidenciado un crecimiento significativo de estas técnicas dejando atrás los métodos convencionales. El propósito es mantener actualizado a los estudiantes con las últimas tecnologías y tendencias en ciencia de datos específicamente en el sector petrolero.

Dado que el presente estudio constituye una investigación exploratoria y, por consiguiente, general, se recomienda llevar a cabo un estudio socioeconómico que demuestre el ahorro obtenido al emplear la ciencia de datos en el área de ingeniería de petróleos en comparación con los métodos convencionales. Para ello, podría realizarse un estudio sobre la aplicación de una técnica de ciencia de datos en particular en el cual se evidencie el potencial, por ejemplo, las redes neuronales artificiales o árbol de decisiones.

Se recomienda utilizar algoritmos de machine learning para realizar pronósticos de producción en campos petroleros colombianos. Asimismo, se aconseja a las investigaciones futuras llevar a cabo una exploración de los beneficios obtenidos al aplicar la ciencia de datos en los sectores colombianos.

Bibliografía

- Ahmed, A. et al. 2021. *Digital Intelligent Artificial Lift DIAL Gas Lift Production Optimization Deployments Across Assets*.
- Ahmed, Tarek. (2001). *Reservoir Engineering Handbook (2nd Edition)*. Elsevier. Retrieved from <https://app.knovel.com/hotlink/toc/id:kpREHE0001/reservoir-engineering/reservoir-engineering>
- Akinsete, O. 2019. *Leak Detection in Natural Gas Pipelines Using Intelligent Models*.
- Alharbi, R. et al. 2023. *Prediction of Oil Production Through Linear Regression Model and Big Data Tools*.
- Alsaif, A. 2022. *Road map to develop an artificial neural network to predict two-phase flow regime in inclined pipes*.
- Asensi, M., et al. 2014. *Métodos y técnicas cualitativas y cuantitativas aplicables a la investigación en ciencias sociales*. ISBN: 978-84-16062-32-4.
- Bahaloo, S. Mehrizadeh, M. Adel Najafi-Marghmaleki., 2022. *Review of application of artificial intelligence techniques in petroleum operations*, *Petroleum Research*, Volume 8, Issue 2. <https://doi.org/10.1016/j.ptlrs.2022.07.002>.
- Ballesteros, B. Sánchez, W. 2021. *Estudio exploratorio de la aplicación de la ciencia de datos en las materias del ciclo profesional del pénsum de la carrera de ingeniería de petróleos de la Universidad Industrial de Santander. Bucaramanga*.
- Barrios, J. 2022. *Light GBM vs XGBoost. ¿Cuál es mejor el algoritmo?* *Health Big Data*. <https://www.juanbarrios.com/light-gbm-vs-xgboost-cual-es-mejor-algoritmo/>
- Basbariol, T. et al. 2020. *A Machine Learning – Based System for Self – Diagnosis Multiphase Flow*.
- Beggs, H.D. 2003. *Production Optimization Using NODAL Analysis, Tulsa, Oklahoma*. p. 5-35

- Bhowmik, S. 2019. *Subsea Pipeline Design Automation Using Digital Field Twin*.
- Bhowmik, S. et al. 2020. *Subsea Structure and Pipeline Design Automation Using Digital Field Twin*.
- Binu, D. Rajakumar, B. 2021. *Artificial Intelligence in Data Mining: Theories and Applications*.
- Buchanan, B. Duda, R., 1983. *Principles of Rule-Based Expert Systems*. ELSEVIER. Volume 22.
[https://doi.org/10.1016/S0065-2458\(08\)60129-1](https://doi.org/10.1016/S0065-2458(08)60129-1)
- Cadei, L. et al. 2020. *Achieving Digital-Twin Through Advanced Analytics Support: A Novelty*
- Cai, W. et al. 2020. *Well Condition Diagnosis of Sucker-Rod Pumping Wells Based on the Machine Learning of Electrical Power Curves in the Context of IoT*.
- Chandel, A., & Sood, M. (2014). *Searching and Optimization Techniques in Artificial Intelligence: A Comparative Study and Complexity Analysis*. IJAR CET, 866-871
- Chandra, B. 2007. *Enfoque sólido para estimar probabilidades en el clasificador Naive-Bayes*.
- Chowdhury, K. et al. 2020. *A Cloud-Based Smart Engineering and Predictive Computation System for Pipeline Design and Operation Cost Reduction*.
- Cirani, S., Ferrari, G., Picone, M., & Veltri, L. (2018). *INTERNET OF THINGS Architectures, Protocols and Standards*. John Wiley & Sons, Inc. doi:10.1002/9781119359715
- Constantine D Spyropoulos, *AI planning and scheduling in the medical hospital environment, Artificial Intelligence in Medicine, Volume 20, Issue 2, 2000. Pages 101-111*. <https://acortar.link/GBrye5>
- Correa, G.J & Montoya, L. M. 2013. *Aplicación del modelo ANFIS para predicción de series de tiempo*.
- D, Binu. B, R, Rajakumar. 2021. *Artificial intelligence in Data Mining theories and applicacions*. Elsevier Inc. p 136.
- Dogali, A. et al. 2022. *Intelligent Approach for GOSP Oil Recovery Enhancement*.
- Elgendy, M. 2020. *Deep Learning for Vision System*. ISBN: 9781617296192.

- Elhaj, M. et al. 2015. *Single Gas Flow Prediction through Chokes using Artificial Intelligence Techniques*.
- Elias et al. 2009. *New Inflow Performance Relationship for Solution-Gas Drive Oil Reservoir, SPE 124041, El Cairo University*.
- Frackiewicz, M. (n.d.) 2023. *Cómo ChatGPT está cambiando el juego en la comunicación por satélite - TS2 SPACE. Cómo ChatGPT Está Cambiando El Juego En La Comunicación Por Satélite - TS2 SPACE*.
<https://ts2.space/es/como-chatgpt-esta-cambiando-el-juego-en-la-comunicacion-por-satelite/>
- Franceschetti, D. R. (Ed.). (2018). *Principles of Robotics & Artificial Intelligence*. Grey House Publishing.
- Francisco. J. Palacios, 2003. *ELEMENTOS BÁSICOS DE UNA RED NEURONAL ARTIFICIAL*.
<https://advancedtech.wordpress.com/2007/08/31/elementos-basicos-de-una-red-neuronal-artificial/>
- Frankish, K., & Ramsey, W. M. (2014). *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139046855>.
- Géron, A. 2017. *Hands-On Machine Learning with Scikit-Learn and TensorFlow*.
<http://oreilly.com/catalog/errata.csp?isbn=9781491962299>.
- Géron, D. (2019). *Machine Learning for Beginners: Step-by-Step Guide to Learning and Mastering Machine Learning for Absolute Beginners with Real Examples*. (Publicación independiente)
- Gharbi, R. Mansoori, A. 2005. *An introduction to artificial intelligence applications in petroleum exploration and production*. ELSEVIER. doi:10.1016/j.petrol.2005.09.001
- Gupta. et al. 2021. *Autoregressive and Machine Learning Driven Production Forecasting – Midland Basin Case Study*.
- Hamid, R. et al. 2020. *Well Performance Prediction in Montney Formation Using Machine Learning Approaches*.

- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques (Tercera ed.)*. Elsevier
- Hang, R. J. 1993 “ANFIS: adaptive-network-based fuzzy inference system,” *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 23, pp. 665-685.
- Hassa. A. et al. 2019. *Improved Predictions in Oil Operations Using Artificial Intelligent Techniques*.
- Hassan, A. et al. 2019. *Application of Artificial Intelligence Techniques to Predict the Well Productivity of Fishbone Wells*.
- Hassan, K. Cliffor, L. 2021. *An Artificial Intelligence Neural Networks Driven Approach to Forecast Production in Unconventional Reservoirs – Comparative Analysis with Decline Curve*.
- Hongfang Lu., et al. 2019. *Oil and Gas 4.0 era: A systematic review and Outlook*.
<https://doi.org/10.1016/j.compind.2019.06.007>
- Joaquin. A. 2007. *Gradient Boosting con Python*.
- Junfeng, S. et al. 2018. *The Application of Big Data Analysis in the Optimizing and Selecting Artificial Lift Methods*.
- Kalam, A. et al. 2019. *A Novel Correlation to Predict Gas Flow Rates Utilizing Artificial Intelligence: An Industrial 4.0 Approach*
- Kalam, S. et al. 2019. *Recent Understanding of Soil Acidobacteria and Their Ecological Significance: A Critical Review*. *Frontiers in Microbiology*, 11. <https://doi.org/10.3389/fmicb.2020.580024>
- Ketmalee, T. 2019. *Prediction of Inflow Performance Relationship of a Gas Field Using Artificial Intelligence Technique*.
- Kissel, R. *Algorithmic trading methods applications usind advanced statistics, optimization, and machine learning techniques*. 2021. Second edition. Manhasser, NY, United States. Elsevier.
- Klamis, m. 2019. *Optimization of choke size for two-phase fow using artificial intelligence*.

- Klins, M.A; Majcher, M.W. 1992. *Inflow Performance Relationships for Damaged or Improved Wells Producing Under Solution-Gas Drive*, SPE 19852.
- Lee, K. et al. 2022. *Novel Application of AI: Harnessing Tacit Knowledge for LNG Plant Start-Up*.
- Liu, X. 2020. *Application Research on Intelligent Management Technology Based on Oil and Gas Pipeline*.
- Lopez, R. Fernández, J. 2008. *Las redes neuronales artificiales*. NETBIBLO, S. L. ISBN: 978-84-9745-246-5
- Mario A, P. et al. 2020. *Using Machine Learning for Building Multivariate IPR Models from High Frequency Streaming Data*.
- Martín del Brío, Bonifacio; Sanz Molina, Alfredo (2001). *Redes Neuronales y Sistemas Difusos*. Ed. Alfaomega Ra – Ma.
- Martinez, R. 2001. *Forecast Techniques for Lifting Cost in Gas and Oil Onshore Fields*. SPE. Buenos aire, Argentina. <https://doi.org/10.2118/69559-MS>
- Mejia, P. 2019. *Programación Genética y Evolución Gramatical: un enfoque combinado usando Straight Line Programs*. Universidad de Cantabria. [https://repositorio.unican.es/xmlui/bitstream/handle/10902/16926/TFM_PabloMejia.pdf?sequence=1#:~:text=La%20programaci%C3%B3n%20gen%C3%A9tica%20\(GP\)%20es,6%5D%2C%20%5B%5D](https://repositorio.unican.es/xmlui/bitstream/handle/10902/16926/TFM_PabloMejia.pdf?sequence=1#:~:text=La%20programaci%C3%B3n%20gen%C3%A9tica%20(GP)%20es,6%5D%2C%20%5B%5D)
- D
- Merzoug, A. et al. 2022. *A New Model for Optimized Gas Allocation in Gas Lift Operation Using Metaheuristic Algorithms*.
- Mezura-Montes, E. (2007). *A New Multimodal Multi-level Genetic Programming Model for Dynamic Optimization*. University of Birmingham.
- Mirjalili, S. Mohammad, S. Lewis, A., 2014. *Grey Wolf Optimizer, Advances in Engineering Software, Volume 69*. <https://doi.org/10.1016/j.advengsoft.2013.12.007>.

- Mohamed A. Abd El Moniem¹ · Ahmed H. El Banbi. 2017. *Development of an expert system for selection of multiphase flow correlations.*
- Mohammad, R. et al. 2019. *Machine Learning Application for Oil Rate Prediction in Artificial Gas Lift Wells.*
- Mohammad, R. et al. 2020. *Application of Artificial Intelligence to Estimate Oil Flow Rate in Gas-Lift Wells.*
- Mohammed, A. 2015. *Improving multiphase choke performance prediction and well production test validation using artificial intelligence: new milestone.*
- Montes, C. M. 2007. *A Genetic Programming infrastructure profiting from public computation resources.*
- Mueller, J. P., & Massaron, L. (2018). *Artificial Intelligence for dummies.* Wiley
- Müller, A. C., & Guido, S. (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists.* O'Reilly Media, Inc.
- Nacelle, A. 2009. *Redes neuronales artificiales.* Universidad de la República.
- Neuroth, M. et al. 2000. *Improved modelling and control of oil and gas transport facility operations using artificial intelligence.*
- Nurfarah Izwana, S. et al 2022. *The World's First Successful Digital Intelligent Artificial Lift (DIAL) for Dual String Application - A Paradigm Shift for Gas Lift System in Offshore Malaysia.*
- Ozdemir, S. (2016). *Principles of Data Science.* Packt Publishing Ltd.
- Parabavire. O. 2018. *¿Campos Digitales de Oil & Gas? Lo que debe saber.* <https://acortar.link/toPdTs>
- Parrott, A. y Warshaw, L. (2017). *Industria 4.0 y el gemelo digital: la fabricación encuentra su rival.*
- Pascual, D., Pla, F., & Sánchez, S. (2007). *Algoritmos de agrupamiento.* En R. Quirós, F. Pla, J. M. Badía, & M. Chover (Edits.), *Métodos informáticos avanzados (págs. 163-175).* Publicacions de la Universitat Jaume I.

- Peter J.F. Lucas & Linda C. van der Gaag. 1991. *Principles of Expert Systems*. Centre for Mathematics and Computer Science. <https://www.cs.ru.nl/~peterl/proe.pdf>
- Qiang Bai, Yong Bai. (2014). *Hydraulic and Thermal Analysis of Subsea Pipelines*. Subsea Pipeline Design, Analysis, and Installation. Pages 91-119. <https://doi.org/10.1016/B978-0-12-386888-6.00005-5>.
- R. Mosquera, O. Castrillon, L. Parra. 2018. *Support Vector Machines, Naïve Bayes Classifier and Genetic Algorithms for the Prediction of Psychosocial Risks in Teachers of Colombian Public Schools*.
- Rasmussen, C. Williams, C. 2006. *Gaussian Processes for Machine Learning*. <https://gaussianprocess.org/gpml/chapters/RW.pdf>
- Redutskiy, Y. 2017. *Conceptualization of smart solutions in oil and gas industry*. Elsevier. 10.1016/j.procs.2017.05.435
- Saeed Bahaloo, Masoud Mehrizadeh, Adel Najafi-Marghmaleki. 2022. *Review of application of artificial intelligence techniques in petroleum operations*. <https://doi.org/10.1016/j.ptlrs.2022.07.002>.
- Sakethraman, M. et al. 2020. *Using Digital Twin of Coriolis Meters for Multiphase Flow Measurement*.
- Salas, R. Redes neuronales artificiales. *Universidad de Valparaíso*.
- Salem, B. et al. 2018. *Using Artificial Intelligence to Predict IPR for Vertical Oil Well in Solution Gas Derive reservoirs: A New Approach*.
- Sarkar, D., Bali, R., & Sharma. (2018). *Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems*. Apress. <https://doi.org/10.1007/978-1-4842-3207-1>
- Scott, S. L., Liu, L. y Yi, J. 1999. *Modelling the Effects of a Deepwater Leak on Behaviour of a Multiphase Production Flow Line*. Ponencia presentada en la Conferencia Medio ambiental SPE/EPA E&P de 1999, Austin, Texas.

- Shadravan, A., et al. 2015. *Intelligent Cement Design: Utilizing Machine Learning Algorithms to Assure Effective Long-term Well Integrity*.
- Shashipal, R. et al. 2019. *Decline Curve Analysis Using Artificial Intelligence*.
- Sheremetov, L., et al. 2005. *Intelligent computing in petroleum engineering*. Elsevier. doi10.1016/j.petrol.2005.01.001
- Spyropoulos, C. D. (2000). *AI planning and scheduling in the medical hospital environment*. *Artificial Intelligence in Medicine*, 101-111. [https://doi.org/10.1016/S0933-3657\(00\)00059-2](https://doi.org/10.1016/S0933-3657(00)00059-2)
- Standing, M.B. 1970. *Inflow Performance Relationships for Damaged Wells Producing by Solution – Gas Drive SPE 3237*.
- Sukarno, P.; Wisnogroho, A. 1995. *Generalized Two-Phase IPR Curve Equation under Influence of Non-linear Flow Efficiency*.
- Sun, T. 2023. *CO2 Pipeline Integrity Management: A Digital Twin Approach*.
- Tariq, Z., et al. (2021). *A systematic review of data science and machine learning applications to the oil and gas industry*. *Journal of Petroleum Exploration and Production Technology*. <https://doi.org/10.1007/s13202-021-01302-2>
- Thanawit, O. et al. 2019. *Artificial Lift Selection Using Machine Learning*.
- Timur, B. et al. 2019. *First Principles and Machine Learning Virtual Flow Metering: A Literature Review*.
- Vershinin, V. et al. 2020. *Neural Network as a Tool for Predicting and Controlling the Technological Regime of Production Wells*.
- Vogel, J.V. 1968. *Inflow Performance Relationships for Solution–Gas Drive Wells SPE 1476*.
- Wiggins, M. 1993. *Generalized Inflow Performance Relationships for Three-Phase Flow, SPE 25458, University of Oklahoma*.

Yan, yong. Et al. 2018. *Application of soft computing techniques to multiphase flow measurement: A review.*

YiFei Li, Han Cao. 2018. *Prediction for Tourism Flow based on LSTM Neural Network, Procedia Computer Science.* <https://doi.org/10.1016/j.procs.2018.03.076>.

Anexos

- **Ecuación. Ley de Darcy para flujo Pseudoestable.**

$$J = \frac{Q_o}{P_r - P_{wf}} = \frac{0.00708K_o h}{\mu_o \beta_o \ln(0.472 X + S)}$$

Fuente: Beggs, H.D. 2003. *Production Optimization Using NODAL Analysis, Tulsa, Oklahoma.* p. 5-

35

J: Índice de productividad, BOPD/psi

h: Espesor del yacimiento, ft.

Q_o : Caudal de petróleo, BOPD

μ_o : Viscosidad del petróleo, cP.

P_r : Presión del yacimiento, psi.

β_o : Factor Volumétrico del petróleo, BY/BN.

P_{wf} : Presión de fondo fluente, psi.

X: Factor de forma, adimensional.

K_o : Permeabilidad efectiva al petróleo, mD.

S: Factor de daño, adimensional

- **Ecuación. Correlación IPR de Vogel.**

$$\frac{Q_o}{Q_o \max} = 1 - 0.2 \left(\frac{P_{wf}}{P_r} \right) - 0.8 \left(\frac{P_{wf}}{P_r} \right)^2$$

Fuente: Vogel, J.V. 1968. *Inflow Performance Relationships for Solution-Gas Drive Wells SPE*

1476.

Q_o : Caudal de petróleo, BOPD.

P_{wf} : Presión de fondo fluente, psi.

$Q_o \max$: Caudal de petróleo máximo ($P_{wf}=0$), BOPD

P_r : Presión del yacimiento, psi

- **Ecuación. Correlación IPR de Standing para pozos con Factor de Daño.**

$$FE: \frac{\ln(0.472 X)}{\ln(0.472 X) + S}$$

$FE > 1$; Pozo Estimulado
 $FE < 1$; Pozo con Daño de formación

$$\frac{Q_o}{Q_o \max} = FE (1.8 - 1.8 A) - 0.8 FE^2 (1 - 2A + A^2)$$

$$A = \frac{P_{wf}}{P_r}$$

$$P_{wfi} = P_r - FE (P_r - P_{wf})$$

Fuente: Standing, M.B. 1970. *Inflow Performance Relationships for Damaged Wells Producing by Solution – Gas Drive SPE 3237.*

Q_o : Caudal de petróleo, BOPD.

X : Factor de forma, adimensional.

$Q_o \max$: Caudal de petróleo máximo ($P_{wf}=0$), BOPD

S : Factor de daño, adimensional.

P_{wf} : Presión de fondo fluyente, psi.

P_{wfi} : P_{wf} cuando $FE=1$, psi.

P_r : Presión del yacimiento, psi.

FE : Eficiencia de flujo, % o fracción

- **Ecuación. IPR de Wiggins para flujo de petróleo, gas y agua.**

$$\frac{Q_o}{Q_o \max} = 1 - 0.519167 \left(\frac{P_{wf}}{P_r} \right) - 0.481092 \left(\frac{P_{wf}}{P_r} \right)^2$$

$$\frac{Q_w}{Q_w \max} = 1 - 0.722235 \left(\frac{P_{wf}}{P_r} \right) - 0.284777 \left(\frac{P_{wf}}{P_r} \right)^2$$

Fuente: Wiggins, M. 1993. *Generalized Inflow Performance Relationships for Three-Phase Flow, SPE 25458, University of Oklahoma.*

Q_o : Caudal de petróleo, BOPD.

$Q_w \max$: Caudal de agua máximo BWPD.

$Q_o \max$: Caudal de petróleo máximo BOPD.

P_{wf} : Presión de fondo fluyente, psi.

Q_w : Caudal de agua, BWPD.

P_r : Presión del yacimiento, psi

- **Ecuación. Correlación IPR de Fetkovich.**

$$\frac{Q_o}{Q_o \max} = \left(1 - \frac{P_{wf}^2}{P_r^2}\right)^n$$

Fuente: Elias et al. 2009. *New Inflow Performance Relationship for Solution-Gas Drive Oil Reservoir*, SPE 124041, El Cairo University.

Q_o : Caudal de petróleo, BOPD.

P_{wf} : Presión de fondo fluyente, psi.

$Q_o \max$: Caudal de petróleo máximo BOPD.

P_r : Presión del yacimiento, psi.

N: pendiente de grafica Q_o vs $(Pr^2 - P_{wf}^2)$ en formato Log-Log

- **Ecuación. Correlación IPR de Klins y Majcher.**

$$\frac{Q_o}{Q_o \max} = 1 - 0.295 \left(\frac{P_{wf}}{P_r}\right) - 0.705 \left(\frac{P_{wf}}{P_r}\right)^{N1}$$

$$N1 = 0.28 + 0.72 \left(\frac{Pr}{Pb}\right) (1.235 + 0.001 Pb)$$

Fuente: Klins, M.A; Majcher, M.W. 1992. *Inflow Performance Relationships for Damaged or Improved Wells Producing Under Solution-Gas Drive*, SPE 19852.

Q_o : Caudal de petróleo, BOPD.

P_r : Presión del yacimiento, psi.

$Q_o \max$: Caudal de petróleo máximo BOPD.

N1: Parámetros del petróleo para la ecuación de Klins.

P_{wf} : Presión de fondo fluyente, psi.

• **Ecuación. Correlación IPR de Sukamo y Wisnagroho.**

$$\frac{Q_o}{Q_o \max_{(S=0)}} = FE \left(1 - 0.1489 \left(\frac{P_{wf}}{P_r} \right) - 0.4416 \left(\frac{P_{wf}}{P_r} \right)^2 - 0.4093 \left(\frac{P_{wf}}{P_r} \right)^3 \right)$$

$$FE = a_0 + a_1 \left(\frac{P_{wf}}{P_r} \right) + a_2 \left(\frac{P_{wf}}{P_r} \right)^2 + a_3 \left(\frac{P_{wf}}{P_r} \right)^3$$

$$a_i = b_{0i} + b_{1i}S + b_{2i}S^2 + b_{3i}S^3$$

Fuente: Sukarno, P.; Wisnagroho, A. 1995. *Generalized Two-Phase IPR Curve Equation under Influence of Non-linear Flow Efficiency.*

Q_o : Caudal de petróleo, BOPD.

P_r : Presión del yacimiento, psi.

$Q_o \max$: Caudal de petróleo máximo BOPD.

S: Factor de daño adimensional.

FE: Eficiencia de flujo, % o fracción.

$a_0 a_1 a_2 a_3$: Constantes adimensionales.

P_{wf} : Presión de fondo fluyente, psi.

$b_{0i} b_{1i} b_{2i} b_{3i}$: Constantes adimensionales.

• **Ecuación. Correlación IPR de Elias et al.**

$$\text{Si } P_r \leq 1600 \text{ psia}$$

$$\alpha = \frac{1}{-0.98 P_r - 152.585}$$

$$\text{Si } P_r > 1600 \text{ psia}$$

$$\alpha = -0.0043065 + 4.98E - 06P_r - 2.41E - 09P_r^2 + 5.69E - 13P_r^3 - 6.48E - 17P_r^4 + 2.85E - 21P_r^5$$

$$\frac{Q_o}{Q_o \max} = 1 - \frac{\ln(\alpha P_{wf} + 1)}{\ln(\alpha P_r + 1)}$$

Fuente: Elias et al. 2009. *New Inflow Performance Relationship for Solution-Gas Drive Oil Reservoir, SPE 124041, El Cairo University.*

Q_o : Caudal de petróleo, BOPD.

P_r : Presión del yacimiento, psi.

$Q_o \max$: Caudal de petróleo máximo BOPD.

α : Parámetro IPR del petróleo, adimensional.

P_{wf} : Presión de fondo fluyente, psi.

- **Ecuación. REGION I: BAJAS PRESIONES: Cuando el gas tiene presiones por debajo de 2000 psi**

$$Q_g: \frac{Kh(P_r^2 - P_{wf}^2)}{1422(\mu_g Z)_{avg} \left(\ln \left(\frac{r_e}{r_w} \right) \right) - \frac{3}{4} + S}$$

K: Permeabilidad, md.

h: Espesor de a arena, ft.

Q_g : Flujo de gas, Mscf/D

r_e : Radio de drenaje, ft.

r_w : Radio del pozo, ft.

P_r^2 : Presión promedio, psi.

P_{wf}^2 : Presión de fondo fluyente a nivel de las perforaciones, psi.

μ_g : Viscosidad de gas, cp.

Z: factor de compresibilidad del gas

S: Factor de daño.

- **Ecuación. REGION II: PRESIONES INTERMEDIAS: Cuando el gas tiene presiones entre 2000 y 3000 psi**

$$Q_g: \frac{Kh(\psi_r - \psi_{wf})}{1422 T \left(\ln \left(\frac{r_e}{r_w} \right) \right) - \frac{3}{4} + S}$$

K: Permeabilidad, md.

h: Espesor de a arena, ft.

Q_g : Flujo de gas, Mscf/D

r_e : Radio de drenaje, ft.

r_w : Radio del pozo, ft.

ψ_r : Pseudo-presión, psi.

ψ_{wf} : pseudo-presión de fondo fluyente a nivel de las perforaciones, psi.

S: Factor de daño.

- **Ecuación. REGION III: ALTAS PRESIONES: cuando el gas tiene presiones por encima de los 3000 psi**

$$Q_g: \frac{7.08 \times 10^{-6} K_g h (P_r - P_{wf})}{(\mu_g B_g)_{avg} \left(\ln \left(\frac{r_e}{r_w} \right) \right) - \frac{3}{4} + S}$$

K: Permeabilidad, md.

h: Espesor de a arena, ft.

Q_g : Flujo de gas, Mscf/D

r_e : Radio de drenaje, ft.

r_w : Radio del pozo, ft.

P_r^2 : Presión promedio, psi.

P_{wf}^2 : Presión de fondo fluyente a nivel de las perforaciones, psi.

μ_g : Viscosidad a la presión promedio, cp.

B_g : Factor volumétrico de la formación a la presión promedio, bbl/scf.

S: Factor de daño.

- **Figuras. Paso a paso de implementación técnica ciencia de datos.**

FORECASTING


1. **PRIMER PASO**
Recolección de datos: Recopilar los datos históricos pertinentes para la serie temporal que deseas prever.
2. **SEGUNDO PASO**
Indagación de datos: realizar un análisis preliminar de los datos para comprender su estructura y cualquier patrón o tendencia evidente
3. **TERCER PASO**
Elección del enfoque: se selecciona un modelo de previsión apropiado para tu serie temporal.
4. **CUARTO PASO**
Partición de datos: Se dividen los datos en entrenamiento y en prueba
5. **QUINTO PASO**
Entrenamiento del modelo: se usan los datos de entrenamientos para ajustar el modelo seleccionado
6. **SEXTO PASO**
Verificación del modelo: Se ensaya el modelo utilizando los datos de prueba.
7. **SÉPTIMO PASO**
Ajuste del modelo: si es necesario se realiza un ajuste de los hiperparámetros del modelo.
8. **OCTAVO PASO**
Producción de predicciones: Aplica el modelo capacitado para generar estimaciones para el lapso temporal especificado.
9. **NOVENO PASO**
Evaluación y mejora: Se compara los pronósticos generados con valores reales observados y se ajusta el modelo según sea necesario.

SISTEMA LÓGICA DIFUSA (ANFIS)

ANFIS se basa en dos conjuntos de parámetros: premisa y consecuencia. Su objetivo es encontrar de forma óptima estos parámetros para establecer una correlación efectiva entre los datos de entrada y generar resultados precisos. ANFIS ajusta estos parámetros de manera adaptable para reflejar el comportamiento de los datos, permitiendo una modelización robusta y capacidad de predicción adecuada.


1. Recopilación y preprocesamiento de datos:
 - Reúne un conjunto de datos que represente el problema que estás tratando de resolver.
 - Re procesa los datos según sea necesario, lo que puede incluir normalización, eliminación de valores atípicos, etc.
2. División de datos:
 - Divide tus datos en conjuntos de entrenamiento, validación y prueba. El conjunto de entrenamiento se utiliza para ajustar los parámetros del modelo, el conjunto de validación se utiliza para ajustar los hiperparámetros del modelo y el conjunto de prueba se utiliza para evaluar el rendimiento final del modelo.
3. Diseño del modelo ANFIS:
 - Un modelo ANFIS típico tiene varias capas que representan diferentes componentes del sistema:
 - Capa de nodos de entrada: cada nodo representa una variable de entrada.
 - Capas de nodos de pertenencia difusa: cada nodo calcula la membresía de la variable de entrada en un conjunto difuso.
 - Capa de reglas: combina las salidas de los nodos de pertenencia para formar reglas difusas.
 - Capa de nodo de salida: calcula la salida del modelo.
4. Inicialización de parámetros:
 - Inicializa los parámetros del modelo, como los centros y las anchuras de las funciones de pertenencia difusa, de manera aleatoria o utilizando algún método heurístico.
5. Entrenamiento del modelo:
 - Utiliza el conjunto de entrenamiento para ajustar los parámetros del modelo.
 - Puedes utilizar algoritmos de optimización como el descenso de gradiente para minimizar una función de pérdida que mida la discrepancia entre las salidas del modelo y los valores reales.
6. Validación del modelo:
 - Utiliza el conjunto de validación para ajustar los hiperparámetros del modelo, como el número de nodos en las capas ocultas, la tasa de aprendizaje, etc.
 - Realiza ajustes en el diseño del modelo según sea necesario para mejorar el rendimiento en el conjunto de validación.
7. Prueba del modelo:
 - Evalúa el rendimiento final del modelo utilizando el conjunto de prueba.
 - Calcula métricas de evaluación como el error cuadrático medio, el coeficiente de determinación, etc., para evaluar la precisión del modelo.
8. Ajuste fino y optimización:
 - Realiza ajustes adicionales en el modelo según sea necesario para mejorar su rendimiento, como la selección de diferentes funciones de pertenencia, la optimización de los parámetros del modelo, etc.

BIG DATA

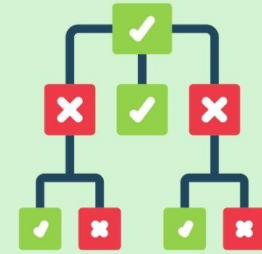


1. Definir el problema y los objetivos: Identifica claramente el problema que estás tratando de resolver y establece los objetivos que deseas lograr con el análisis de Big Data. Esto ayudará a enfocar tus esfuerzos y a determinar qué datos son relevantes.
2. Recolección de datos: Reúne todos los datos relevantes para tu problema. Esto puede incluir datos estructurados (por ejemplo, de bases de datos relacionales), datos semi-estructurados (por ejemplo, archivos CSV) y datos no estructurados (por ejemplo, texto, imágenes, videos).
3. Almacenar los datos: Decide cómo almacenarás los datos. Esto puede ser en un sistema de almacenamiento distribuido o en una nube.
4. Preprocesamiento de datos: Limpia y prepara tus datos para el análisis. Esto puede implicar eliminar valores faltantes, corregir errores, normalizar datos y transformar datos en un formato adecuado para el análisis.
5. Procesamiento y análisis de datos: Utiliza herramientas de procesamiento de Big Data. Esto puede incluir la realización de consultas complejas, el procesamiento de datos en tiempo real, el análisis de grafos, etc.
6. Visualización de datos: Visualiza los resultados de tu análisis porque ayudan a comprender los patrones en los datos y a comunicar los resultados de manera efectiva.
7. Modelado y predicción: Si es necesario, utiliza técnicas de modelado predictivo o de aprendizaje automático para construir modelos que puedan predecir resultados futuros o identificar patrones en los datos. Esto puede incluir el uso de algoritmos como regresión, clasificación, agrupación, etc.
8. Optimización y ajuste fino: Ajusta tus modelos y técnicas de análisis según sea necesario para mejorar la precisión y el rendimiento.
9. Implementación y despliegue: Implementa tus soluciones en producción y desplégalas para su uso en entornos reales. Asegúrate de que tus soluciones sean escalables, seguras y eficientes en términos de recursos.
10. Monitoreo y mantenimiento: Monitorea el rendimiento de tus soluciones en producción y realiza mantenimiento regular para garantizar que sigan siendo efectivas con el tiempo. Esto puede implicar la monitorización de métricas clave, la detección y corrección de problemas, y la actualización de modelos y análisis según sea necesario.

ARBOL DE DECISION



1. **IMPLEMENTACIÓN DE DATOS**
Es importante comenzar con un conjunto de datos limpio y bien organizado. Estos conjuntos de datos consisten en características o atributos que caracterizan un objeto específico. Pueden incluir datos numéricos como precios, dimensiones, peso y tiempo, entre otros, asociados con un objeto o clase particular.
2. **SELECCIÓN DE VARIABLES**
Al crear un árbol, existen diferentes métodos para determinar la mejor manera de dividir cada nodo. Esto puede incluir evaluar el Error de Clasificación, el índice Gini o la Entropía. El valor de estas métricas en cada nodo guiará la forma en que se dividen los caminos del árbol.
3. **ELEGIR EL NODO INICIAL DEL ÁRBOL**
Esto implica utilizar la variable de clasificación en todo el conjunto inicial para realizar la primera división.
4. **ESTABLECER NODOS FINALES**
Es crucial determinar la profundidad adecuada del árbol, especialmente en conjuntos de datos grandes, para evitar que se convierta en una estructura excesivamente compleja. Esto se logra utilizando criterios como la profundidad máxima del árbol o el número total de nodos para detener su crecimiento. Cada vez que se evalúa dividir un subconjunto, se considera este criterio para decidir si continuar o detenerse en un nodo terminal. Una vez que el árbol no puede dividirse más, se elige un criterio para seleccionar la predicción final, a menudo basada en el dato más común entre las opciones posibles.
5. **ELABORAR EL ÁRBOL DE DECISION**
Se puede lograr utilizando un algoritmo recursivo que, dado un conjunto de registros de entrenamiento en un nodo, determina si pertenecen a una sola clase (en cuyo caso se considera un nodo terminal) o a múltiples clases. En el segundo caso, se dividen los datos en subconjuntos más pequeños según la variable de clasificación elegida. Este proceso se repite hasta que se alcancen todos los nodos terminales necesarios.
6. **HACER LAS PREDICCIONES**
Se pueden llevar a cabo mediante la aplicación de un algoritmo recursivo que analiza cada nodo del árbol según los datos de entrada, seleccionando el camino correcto en cada nodo hasta llegar a un nodo terminal. En este nodo terminal se obtiene la predicción final.

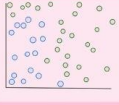


SVM MÁQUINAS DE VECTORES SOPORTE

Support vector machine

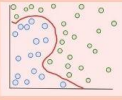
1. CORRELACIÓN DE DATOS

Colocación de datos a un espacio de características de grandes dimensiones. En la figura se observa un ejemplo (los puntos de datos corresponden a dos categorías)



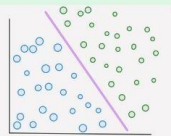
2. AJUSTE

Se ajusta a un hiperplano de separación entre las dos clases. En la figura se observa los datos con un separador añadido



3. SELECCIÓN DEL MODELO

SVM, Se elige este algoritmo de aprendizaje supervisado que se utiliza en muchos problemas de clasificación y regresión



Tras la transformación, el límite entre las dos categorías se define por un hiperplano, como se muestra en la figura

4. FASE DE ENTRENAMIENTO

Se selecciona un hiperplano óptimo que busca maximizar la distancia euclidiana a los puntos de datos más cercanos en cada lado del plano. Estos puntos de datos más cercanos a cada lado del hiperplano son referidos como **vectores de soporte**.

5. VALIDACIÓN DEL MODELO

Viene dada por la función matemática de Kernel, Según M.R. Alam:

$$S(x) = \sum_{i=1}^L \alpha_i c_i K(x, x_i) + d$$

GEMELOS DIGITALES

1.

Definición del Gemelo Digital: Comienza por definir claramente qué objeto físico o sistema desea replicar digitalmente.

2

Recolección de Datos: Recolecta datos del objeto físico utilizando diversas fuentes.

3

Digitalización: se usan técnicas como escaneo 3D, fotogrametría, LIDAR (Detección y Teledetección por Luz y Alcance), etc.

4

Modelado 3D: Se utiliza un software de modelado 3D para crear un modelo digital preciso basado en los datos recopilados.

5

Integración de Datos: Integra los datos recopilados, como información de sensores, registros de mantenimiento y cualquier otra información relevante, en el modelo digital

6

Implementación de Algoritmos y Simulaciones: Implementar algoritmos y simulaciones que permitan al gemelo digital emular el comportamiento del objeto físico en diferentes condiciones y escenarios.

7

Conectividad y Actualización en Tiempo Real: Establece una conexión entre el gemelo digital y el objeto físico para permitir la recopilación continua de datos en tiempo real y la actualización.

8

Validación y Verificación: Realiza pruebas exhaustivas para validar y verificar la precisión y la eficacia del gemelo digital en la emulación del objeto físico en diferentes condiciones y escenarios.

9

Aplicación y Uso: Se aplica el gemelo digital en entornos prácticos donde pueda utilizarse. Ej: optimización de procesos, análisis de rendimiento, mantenimiento predictivo, formación y simulación, Etc.

10

Revisión y Mejora Continua: Se revisa el rendimiento del gemelo digital en uso práctico y realiza mejoras continuas según sea necesario para garantizar que siga siendo preciso y efectivo en su función.

