APLICACIÓN DE LA METODOLOGÍA DE BIG DATA ANALYTICS A OPERACIONES DE PERFORACIÓN PARA ESTABLECER LOS PARÁMETROS ÓPTIMOS QUE REDUZCAN EL TIEMPO DE OPERACIÓN

PRESENTADO POR:

FREIDER NICOLAS TELLEZ DURAN

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERIAS FISICOQUIMICAS
ESCUELA DE INGENIERIA DE PETROLEOS
BUCARAMANGA
2018

APLICACIÓN DE LA METODOLOGÍA DE BIG DATA ANALYTICS A OPERACIONES DE PERFORACIÓN PARA ESTABLECER LOS PARÁMETROS ÓPTIMOS QUE REDUZCAN EL TIEMPO DE OPERACIÓN

Proyecto de grado presentado como requisito para optar al título de ingeniero de petróleos

PRESENTADO POR:
FREIDER NICOLAS TELLEZ DURAN

DIRECTOR
WILSON RAÚL CARREÑO VELASCO
Magíster en diseño, gestión y dirección

CO-DIRECTORES
HERNAN DARIO MANTILLA HERNANDEZ
Magíster en ingeniería de hidrocarburos

MAIKA KAREN GAMBUS ORDAZ
Doctora en ingeniería de petróleos

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERIAS FISICOQUIMICAS
ESCUELA DE INGENIERIA DE PETROLEOS
BUCARAMANGA
2018

DEDICATORIA

A Dios por todas las bendiciones que me brinda a diario, por darme la fortaleza y sabiduría para superar todos los obstáculos de la mejor manera y por brindarme una familia que, aunque pequeña en cantidad, inmensa en calidad

A mi madre querida, Xiomara Duran Salazar, porque por ella es que todo esto es posible, por todo su apoyo, su amor y dedicación y por siempre estar junto a mi impulsándome a superar mis metas y alcanzar todos los objetivos que me propongo
Te amo inmensamente

A mi tía Jeannette Duran Salazar, mi tío Hernán García Sarmiento, mi Prima, que es más como mi hermana, Karoll Marcela Álvarez Duran y mi nonita linda, Rosita Salazar de Duran, que nos cuida desde el cielo, por todas las palabras de ánimo y consejos y por ser mis pilares, incondicionales siempre

A mi padrino, Jesús Alfonso Ordoñez, por ser ese guía espiritual que me acompaña siempre y a José David Manrique Rueda, que ha sido un apoyo constante para mi mamá y para mi en las situaciones más críticas y especiales

Y a mis dos mejores amigos, a los que considero mis hermanos, Juan Sebastián Mayorga Quintero y Julián Ricardo Prada Gutiérrez, quienes me acompañan siempre en las buenas y en las malas sin excepciones y sin pedir nada a cambio
Gracias por acompañarme en todas mis locuras

AGRADECIMIENTOS

A mi hogar, la Universidad Industrial de Santander, el alma mater que me ha permitido vivir miles de experiencias increíbles y conocer gente fantástica.

A la escuela de ingeniería de petróleos y a todos los docentes que aportaron de una u otra manera en mi crecimiento tanto personal como académico y en quienes encontré grandes personas y mentores.

Al Grupo de Investigación en Estabilidad de Pozos, en cabeza de la Ph.D. Zuly Himelda Calderón Carrillo, por inculcar en mi la pasión por la investigación, por la guía y apoyo constante en la elaboración de este trabajo de grado y por darme la oportunidad de conocer compañeros con los que compartir gratos momentos.

A Ecopetrol S.A y al Instituto Colombiano del Petróleo por brindarme la información necesaria para poder llevar a cabo el desarrollo de este trabajo y, en especial a los ingenieros Reinel Corzo Rueda y Hernán Darío Mantilla por el tiempo, las sugerencias, el ánimo e interés dedicado durante la elaboración de este proyecto.

Al ingeniero Wilson Raúl Carreño Velasco, porque más que el director de este proyecto es un amigo y un excelente guía, quien con sus historias y consejos se convirtió en fuente de inspiración para mi crecimiento integral y para la mejor ejecución de este trabajo que me llena de orgullo.

A mis calificadores, el ingeniero Néstor Fernando Saavedra y el Ph.D. Emiliano Ariza, por aportar, con sus correcciones y sugerencias, a que este trabajo se realizará de la mejor manera posible.

A Carlos Andrés Berdugo y Felipe Romero Consuegra, buenos amigos, quienes, con sus ideas, brindaron un gran aporte para la feliz culminación de este arduo trabajo.

A esos amigos que siempre han estado para brindarme una mano, un consejo o simplemente una frase de ánimo: Andrés Rangel, Fabian, Tatiana, Camila Pachón, Nata, Jota, Danna, Mauro, Juanjo, Katty, Pipe Leal, Caro, Angelica Yunez, Soniale, Jhon Ramírez, Sebastián Moreno, Manuela Montt, Mafe Espejo, Sebastian Cornejo, Majo, Anggie, Erika, Camila Lizarazo, Jorge, Luisda, Centeno, Oscar, Verónica, Jose, Gabriel, Jaime, Mayra, Genghini, Sandra, Silvia, Luisca, Panda, Pachito, Angelica Gonzales, y finalmente, aunque no menos importante, a Yuri Stefany León, quien a pesar de llevar poco tiempo en mi vida, se ha convertido en una gran compañía y en fuente de ánimo e inspiración para enfrentar las más duras situaciones.

Y a todas aquellas personas que se me puedan escapar y que de una u otra manera han aportado en mi vida y en mi formación. Realmente muchas gracias a todos por hacer parte de este camino.

TABLA DE CONTENIDO

LISTADO DE FIGURAS

LISTADO DE TABLAS

RESUMEN

TITULO: APLICACIÓN DE LA METODOLOGÍA DE BIG DATA ANALYTICS A OPERACIONES DE PERFORACIÓN PARA ESTABLECER LOS PARÁMETROS ÓPTIMOS QUE REDUZCAN EL TIEMPO DE OPERACIÓN*.

AUTOR: FREIDER NICOLAS TELLEZ DURAN**

PALABRAS CLAVES: BIG DATA ANALYTICS, OPTIMIZACIÓN DE PARÁMETROS, MSE, PERFORACIÓN.

DESCRIPCIÓN:

El presente trabajo de grado muestra el desarrollo de un nuevo método de análisis Big Data Analytics, basado en la medición y optimización de la energía mecánica específica o MSE, la cual envuelve, implícitamente, diversos parámetros operacionales, inherentes a la perforación, por lo que sirve como parámetro guía para determinar aquellas variables operacionales, no tenidas en cuenta usualmente como un conjunto, sino evaluadas normalmente de manera individual, que podrían ser mejoradas para reducir los tiempos invisibles presentes durante una operación de perforación, logrando finalmente un ahorro de costos significativo. Para esto, son tratados tres partes fundamentales en el presente trabajo, la primera es el estudio del estado del arte del tema de Big Data Analytics en el mundo, lo que involucra los conceptos fundamentales, las principales compañías en manejo de datos a nivel mundial, los softwares, paquetes virtuales y metodologías existentes, además de casos de estudio, enfocados principalmente en el área de los hidrocarburos, que sirvan como guía para la elaboración del nuevo método a ser aplicado, seguido de la elaboración de este, mediante el uso de herramientas de programación como R Studio o Python, junto con el uso redes neuronales artificiales, las cuales resultan útiles para la generación de predicciones, y de su validación, al aplicarlo a diversos pozos de un campo colombiano.

---

*Trabajo de grado

**Facultad de ingenierías Físico-Químicas. Escuela de ingeniería de petróleos. Director: Wilson Raúl Carreño Velasco, magíster en diseño, gestión y dirección.

ABSTRACT

TITLE: APPLICATION OF BIG DATA ANALYTICS METHODOLOGY TO PERFORATION OPERATIONS TO ESTABLISH THE OPTIMAL PARAMETERS THAT REDUCE THE OPERATION TIME*.

AUTHOR: FREIDER NICOLAS TELLEZ DURAN**

KEYWORDS: BIG DATA ANALYTICS, PARAMETERS OPTIMIZATION, MSE, PERFORATION.

DESCRIPTION:

The present thesis shows the development of a new Big Data Analytics analysis method, based on the measurement and optimization of the specific mechanical energy or MSE, which implicitly involves various operational parameters, inherent in drilling, so it serves as a guiding variable to determine those operational variables, not taken into account usually as a set, but normally assessed individually, which could be improved to reduce the invisible times present during a drilling operation, eventually achieving significant cost savings. For this, they are treated three fundamental parts in the present work, the first is the study of the state of the art of the topic of Big Data Analytics in the world, which involves the fundamental concepts, the main companies in data management worldwide, the softwares, virtual packages and existing methodologies, in addition to case studies, mainly focused on the area of hydrocarbons, which serve as a guide for the elaboration of the new method to be applied, followed by the elaboration of this, through the use of programming tools such as R Studio or Python, along with the use of artificial neural networks, which are useful for the generation of predictions, and its validation, applying it to different wells of a Colombian field.

---

*Bachelor Thesis

** Facultad de ingenierías Físico-Químicas. Escuela de ingeniería de petróleos. Director: Wilson Raúl Carreño Velasco, magíster en diseño, gestión y dirección.

INTRODUCCIÓN

Big Data Analytics es un concepto moderno y una técnica poco explotada actualmente, la cual ha cobrado importancia al demostrar, mediante su aplicación en diversas industrias como bancos, centros comerciales o, incluso, grupos políticos, que es una técnica eficiente para la obtención de beneficios mediante la analítica extensa de los datos, ya que permite, no solo manejar los volúmenes masivos de datos actuales, sino que también, encontrar la información realmente relevante que se encuentra en ellas, como patrones o tendencias que no son evidentes y, por lo tanto, no son aprovechadas con el uso de técnicas de análisis convencionales.

Sin embargo, esta técnica no ha sido altamente explotada en la industria de los hidrocarburos, ya que se encuentra limitada a pocas aplicaciones realizadas en la producción de estos, el cuidado de los equipos en las instalaciones de campos petroleros y/o manejo de recursos y personal en las compañías, además esta problemática se extiende en mayor medida en Colombia, donde el empleo de técnicas avanzadas de análisis de datos se encuentra limitado o, incluso, atrasado respecto a otros países.

Es importante señalar que, a la hora de analizar volúmenes masivos de datos, que aumentan de forma exponencial con el paso del tiempo y el avance en la tecnología, un recurso muy valioso es el uso de redes neuronales artificiales, ya que estas permiten cálculos iterativos extensos, al tiempo que se adaptan a la tendencia o comportamiento de la información que se le suministra de mejor manera al aumentar el volumen de entrada con el que es alimentada.

Es por esto al ahondar en el estudio y aplicación de la analítica de Big Data, las redes neuronales se convierten en una herramienta inherentemente fundamental, debido a que al basarse en procesos de aprendizaje de máquinas y Deep learning, el programa “aprende” con cada nuevo paquete de datos que se le es ingresado, por lo que, con el aumento en la cantidad de la información

de entrada que se le suministra, se obtienen resultados más precisos y de forma más rápida que con la implementación de técnicas de análisis estadístico convencional, adaptándose de una mejor manera al comportamiento real de la operación que este siendo estudiada.

1. MARCO TEÓRICO SOBRE BIG DATA Y SU APLICACIÓN EN PERFORACIÓN

Para llegar a generar una nueva metodología de análisis Big Data que permita optimizar las operaciones de perforación es fundamental entender todos los conceptos básicos relacionados a la misma y estudiar el avance que en la actualidad se ha hecho en ella, de acuerdo con lo anterior es de gran importancia comenzar por entender el significado y potencial de esta herramienta.

1.1. Conceptos básicos

1.1.1. Big data

De acuerdo al instituto global McKinsey, Big Data se refiere a todo conjunto de datos cuyo tamaño sea mayor a la habilidad de captura de las herramientas software de bases de datos típicas para capturar, almacenar, manejar y analizar la información, la cual aumenta exponencialmente con el paso de los años[1], así, el análisis de Big Data se ha descrito como el proceso de examinar grandes cantidades de diversos tipos de datos en un esfuerzo por reconocer patrones, correlaciones y cualquier otra información importante, que lleven a procesos de cambio en la toma de decisiones, el ahorro de costos o la generación de mayores ingresos para las empresas[2]. La tecnología de *Big Data and analytics* describe una nueva generación de arquitecturas para la extracción de un valor económico de volúmenes muy grandes de una amplia variedad de datos, permitiendo su captura, descubrimiento y análisis a una gran velocidad[3].

La metodología de Big Data and Anlytics involucra principalmente cuatro factores que le infieren su importancia. Estos son:

- Amplia infraestructura
- Organización y manejo de datos

[1] SPATH, Jeff. Big Data!. Revista JPT. Enero de 2014.

[2] JONHNSTON, J; GUICHARD, A. New Findings in Drilling and Wells using Big Data Analytics. Offshore technology conference. Mayo de 2015.

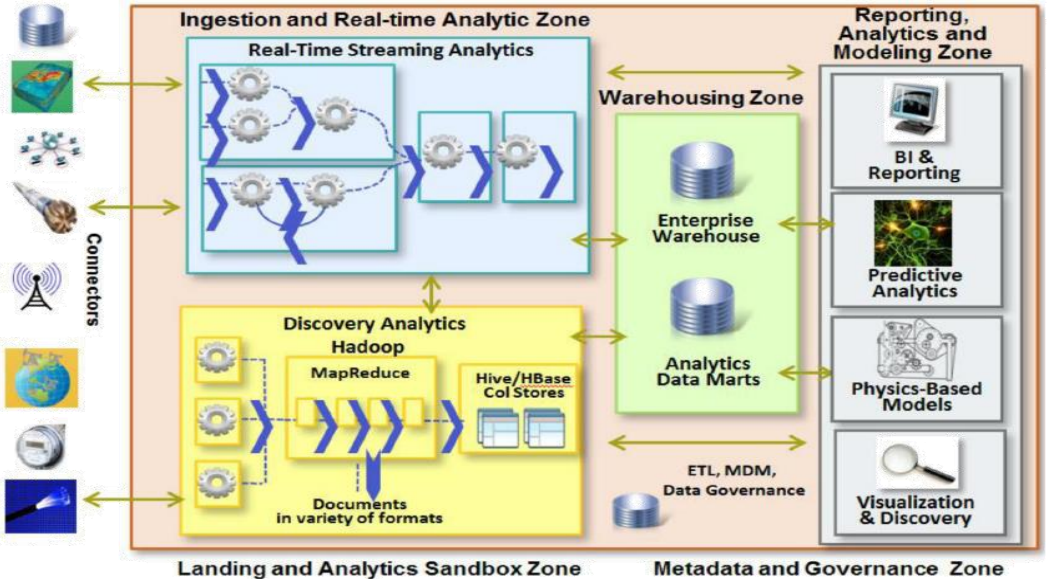
[3] FEBLOWITZ, Jill. Analytics in Oil and Gas: The Big Deal About Big Data. 2013 SPE digital energy conference. Marzo de 2013

- Análisis y descubrimiento de la información
- Soporte para la toma de decisiones

El proceso se basa en la toma, el filtrado y análisis de datos obtenidos directamente de diversas fuentes, los cuales posteriormente son almacenados y reportados (el esquema general se ilustra en la figura 1), también es importante tener en cuenta que, al estar hablando de manejo de datos, se deben aclarar dos conceptos básicos que afectan los mismos, estos son la ley de Moore y la ley de las 6 V's.

1.1.2. Ley de Moore: enuncia que el tamaño y la capacidad de los sistemas de almacenamiento, y por consiguiente de los datos generados, y su complejidad tiende a aumentar exponencialmente cada dos años, lo cual puede ser evidenciado actualmente en la industria, ya que se estima que en cada pozo se genera más de 1 TB(*) de información al día, lo que da cuenta de la importancia asociada al trato adecuado de la misma[4].

Figura 1. Esquema general del proceso de análisis Big Data



Tomado de: BRULÉ, Michael R. The Data Reservoir: How Big Data Technologies Advance Data Management and Analytics in E&P

[4] BECKWITH, Robin. Managing Big Data: Cloud Computing and Co-Location Centers. Revista JPT. Octubre de 2011.

1.1.3. Ley de las 6 V's: esta ley señala las variables de mayor relevancia que se deben tener en cuenta al hablar de manejo de datos[5]. Estas son:

- **Volumen:** se relaciona directamente con la ley de Moore, mencionada anteriormente, además, en la industria de los hidrocarburos los registros sísmicos son los que mayor flujo de datos generan, llegando fácilmente a producir cientos de Terabytes por registro, cantidad que puede aumentar al tratar con una perforación offshore.

- **Variabilidad:** se refiere al cambio constante que se presentan en los datos durante el proceso y el ciclo de vida del proyecto.

- **Veracidad:** se relaciona directamente con la calidad de los datos manejados, la cual, al aumentar, genera un análisis de mayor utilidad, lo que lleva a una mejor toma de decisiones.

- **Variedad:** se relaciona con la fuente de la información, la cual puede generar datos estructurados, semi-estructurados o sin estructurar.

- **Velocidad:** en la actualidad la data puede ser obtenida en "tiempo real", lo que representa una transferencia de datos que puede alcanzar incluso los 2 MB/s(**).

- **Valor:** se vincula de cerca con la veracidad y se refiere a la utilidad que representará la información manejada para solucionar el problema estudiado o alcanzar la meta y objetivos planteados.

5 FEBLOWITZ, Jill. Analytics in Oil and Gas: The Big Deal About Big Data. 2013 SPE digital energy conference. Marzo de 2013

(*) Terabyte

(**) Megabites por segundo

1.1.4. Tipos de información: de acuerdo con el patrón de análisis que se quiera seguir y a los objetivos de la empresa, se pueden llegar a tener tres enfoques diferentes, en los que se maneja un tipo de información distinta, las cuales pueden ser:

1.1.4.1 Información en movimiento*:* este tipo de información se enfatiza en la velocidad y el volumen, para su manejo se recomienda el uso de Stream computing, metodología que se basa en el flujo constante de datos, orquestando un modelo y actualizándolo constantemente, este es el enfoque preferido para el análisis en tiempo real, para su manejo se recomienda el uso de almacenamiento Massively Parallel Processing (MPP), el cual utiliza múltiples procesadores para el manejo de la data[6].

El método de stream computing no provee los modelos, sino una nueva infraestructura que permite correr los modelos más complejos directamente en el taladro o en el campo, donde los datos se están generando, sin la preocupación que pueden generar la escala de los datos, la complejidad del modelo, el ancho de banda, la huella digital y otras barreras para la optimización y automatización de la perforación y la producción en tiempo real[7].

1.1.4.2 Información en pausa*:* en este caso se enfatiza en la variedad y el volumen, por lo cual se recomienda(*) el uso de *Hadoop/MaopReduce*, ya que provee una plataforma de operaciones integrada al obtener, limpiar y transformar la información sea estructurada o no, dentro de los estándares de la industria de los hidrocarburos[8].

(*) Recomendaciones brindadas por Michael Brulé, parte de la empresa IBM

[6] BRULÉ, Michael R. Big Data in E&P: Real-Time Adaptive Analytics and Data-flow Architecture. 2013 SPE digital energy conference. Marzo de 2013.

[7] BRULÉ, Michael R. The Data Reservoir: How Big Data Technologies Advance Data Management and Analytics in E&P. 2015 SPE digital energy conference. Marzo de 2015.

[8] BRULÉ, Michael R. Op. Cit. Marzo de 2013.

Hadoop es una red de software que permite el análisis de Big Data mediante el uso de un sistema de computadoras y modelos simples de programación. Cuenta con dos módulos base[9], que son:

- **Hadoop Distributed File System (HDFS):** en este módulo, la gran cantidad de data es organizada en bloques de 24 MB o más de ser necesario, que luego son organizados en varios nodos, los cuales se almacenan en memorias físicas de computadoras comunes, máquinas virtuales o servidores.

- **MapReduce parallel computing model: e**n este caso, se procesa una amplia cantidad de data en paralelo en varios clústeres de hardware comercial en una forma confiable y tolerante a fallas.

- **Información combinada:** este enfoque busca relacionar los dos tipos de información mencionados anteriormente para enfatizar tanto en la velocidad, como en el volumen y la variedad.

1.2 Problemas por solucionar: actualmente, la supervisión, análisis y optimización de la perforación o la producción incluye datos sónicos y vibracionales medidos a altas frecuencias mediante tecnologías de registros como DAS (Sensores sónicos distribuidos), DTS (Sensores de temperatura distribuidos) y registros de presión de fondo con medidores de fondo permanentes, que permiten realizar registros cada 1/10 ft, incluso en pozos horizontales. Esta amplia gama de fuentes de datos genera diversos problemas durante la operación[10], algunos de estos son:

- Pérdidas de producción.

[9] WU, Wenkuang; et al. Retrieving Information and Discovering Knowledge from Unstructured Data Using Big Data Mining Technique: Heavy Oils Fields Example. 2014

[10] BRULÉ, Michael R. The Data Reservoir: How Big Data Technologies Advance Data Management and Analytics in E&P. 2015 SPE digital energy conference. Marzo de 2015.

- Presencia de tiempos no productivos (NPT's).
- Alto OpEx.
- Los datos no pueden ser suministrados fácilmente a las aplicaciones.
- Muchos datos son generados a un gran precio, pero no son aprovechados.
- Necesidad de pre-procesamiento y manejo de grandes cantidades de datos en múltiples lugares remotos.
- Necesidad de mover datos desde un sitio remoto a una locación central, usualmente utilizando conexiones de comunicación pobres.
- En muchos casos, los operadores no están seguros de qué cambios realizar cuando las situaciones operacionales se vuelven complejas.
- Volúmenes muy grandes de data multidisciplinaria, data estructurada y semiestructurada de producción y data sin estructurar de geología deben ser analizadas y correlacionadas.

1.3 Aplicaciones*:* la importancia de un mejor análisis de datos radica en que nos permite obtener patrones más exactos y amplios que nos ayuden a tomar mejores decisiones durante las operaciones[11]. De esta manera las tres etapas de la metodología serán:

- Organizar datos que estén o no estructurados para proporcionar una visión estadística simple de estos.
- Crear una base unificada de datos que sirva como plataforma para el reconocimiento de patrones, que permitan tomar decisiones oportunas.
- Ofrecer señales en tiempo real, con límites estadísticos y científicos.

Así, las principales aplicaciones que tiene la metodología dentro de la industria de los hidrocarburos son[12]:

[11] ANAND, Pradeep. Big Data Is a Big Deal. Revista JPT. Abril de 2013.

[12] BRULÉ, Michael R. The Data Reservoir: How Big Data Technologies Advance Data Management and Analytics in E&P. 2015 SPE digital energy conference. Marzo de 2015.

- Uso más efectivo de la información de fibra óptica (DTS, DAS, DPS, DSS, etc.) para la optimización de la producción.
- Detectar fugas detrás de la tubería y gradientes composicionales con una velocidad y precisión mayor a la de los métodos convencionales.
- Monitorear los perfiles de fractura, inyección, frentes de flujo, etc, para determinar el estado de la producción del campo.
- Reducir los tiempos no productivos y los tiempos invisibles, permitiendo la optimización de la perforación al generar una supervisión de parámetros como la porosidad, permeabilidad, saturación, entre otros en tiempo real.

1.4 Principales compañías en el manejo de Big Data a nivel mundial*:*

actualmente la metodología de Big Data and Analytics se encuentra en un estado de desarrollo a nivel mundial, en Colombia la metodología es poco utilizada, por lo que se hace fundamental estudiarla para poder obtener las bases que permitan su correcta implementación en el país.

Así, se pueden tomar como base tres entidades que cuentan con una vasta experiencia en el manejo y desarrollo de la metodología, aplicada en diferentes disciplinas, especialmente en negocios y administración, las cuales son presentadas a continuación.

1.4.1. International Business Machine Corp. (IBM): empresa multinacional estadounidense de tecnología y consultoría fundada en 1911, la cual cuenta con el mayor número de patentes en EE.UU y es responsable de grandes inventos como el cajero automático, el disquete, el disco duro, entre otros, y cuyos trabajadores cuentan con premios nobel y Turing, entre otros.

En cuanto al tema de Big Data y Analytics, IBM ha llevado a cabo varios estudios, entre ellos podemos destacar el informe ejecutivo de “Analytics: el uso de big data en el mundo real Cómo las empresas más innovadoras extraen valor de

datos inciertos", realizado en conjunto con la escuela de negocios SaÏd de la universidad de Oxford.

El informe se centra en el estudio de cómo la gente en diversos sectores e industrias define qué es Big Data y Analytics y cómo es utilizada en las mismas, principalmente aquellas enfocadas a servicios al consumidor y bussines-to-bussines (B2B) [socios y proveedores], de esta manera cabe resaltar que "las soluciones de big data más eficaces identifican primero los requisitos de negocio y, a continuación, adaptan la infraestructura, las fuentes de datos y la analítica a fin de respaldar la oportunidad de negocio"[13].

A partir del estudio determinaron cinco recomendaciones clave para que las empresas puedan avanzar en sus iniciativas de Big Data y obtener el máximo valor de negocio, las cuales son:

- Dedicar los esfuerzos iniciales a obtener resultados centrados en el cliente.
- Desarrollar un plan de big data para toda la empresa.
- Comenzar con datos ya existentes para lograr resultados a corto plazo.
- Desarrollar capacidades analíticas sobre la base de prioridades de negocio.
- Crear un caso de negocio sobre la base de resultados cuantificables.

En el estudio, también se utilizó la expresión "adopción de big data" para referirse a una progresión natural de los datos, las fuentes, las tecnologías y las habilidades que resultan necesarias para crear una ventaja competitiva en un mercado integrado a escala global[14].

[13] INSTITUTE FOR BUSINESS VALUE. IBM. Analytics: el uso de big data en el mundo real Cómo las empresas más innovadoras extraen valor de datos inciertos. 2012.
(*) Se encuestaron a 1,141 profesionales y empresas a nivel mundial.
[14] Ibíd.

Para que la metodología sea eficiente es contar con un almacenamiento de datos escalable (Capacidad para manejar el crecimiento continuo de datos de manera fluida) de gran capacidad que permita integrar la información y sea segura.

El estudio reveló que la mayoría de las empresas encuestadas(*), alrededor de un 47%, se encuentran en una fase intermedia o de exploración, donde se están planeando actividades relacionadas con esta metodología, un 24% se encuentra en la fase de estudio de los conceptos (educar) y un 28% aplica actualmente la metodología (Interactuar- un caso de big data "22%", Ejecutar- dos o más caso de Big Data "6%")[15]. Además, se determinó que los objetivos de las empresas que aplican esta metodología son:

- Resultados centrados en el cliente (49%)
- Optimización operativa (18%)
- Gestión financiera/riesgos (15%)
- Nuevo modelo empresarial (14%)
- Colaboración entre empleados (4%)

Las fases de aplicación, así como los objetivos de las empresas son ilustrados en la figura 2 y 3, además, las figuras 4, 5 y 6 ilustran el respaldo de la empresa, la disponibilidad de datos y los principales obstáculos que se presentan en cada fase de la metodología[16].

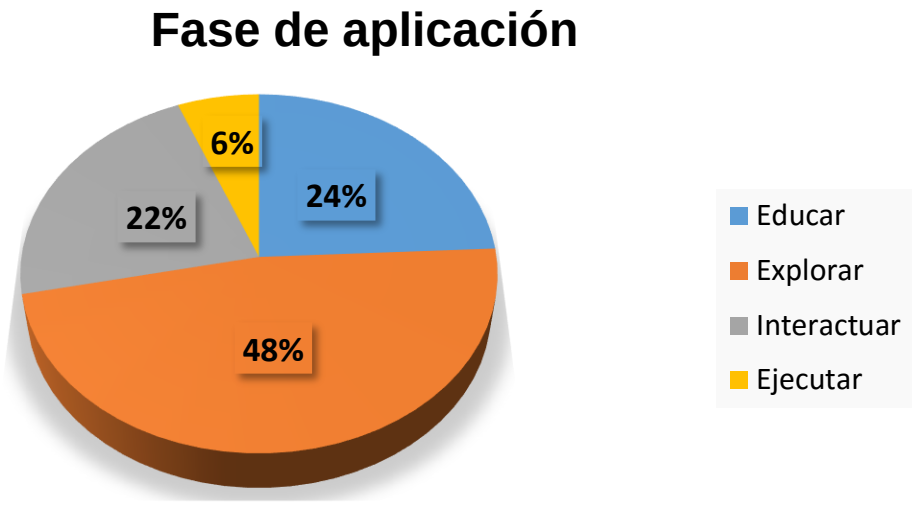
IBM desarrolló la herramienta *Máximo* como una plataforma especializada en la optimización de procesos en la industria de los hidrocarburos basado en el código Java, más específicamente Java 2 Enterprise Edition, esta herramienta se centra, principalmente, en el manejo de las facilidades o instalaciones en superficie, la reducción de riesgos debido a problemas que puedan ocurrir en las mismas, manejo de personal y materiales y transporte[17].

[15] Ibíd.

[16] Ibíd.

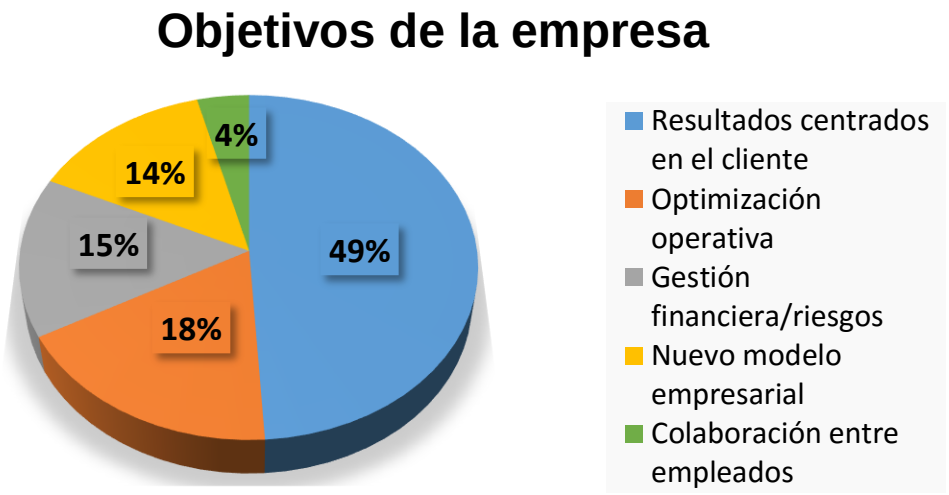
[17] INSTITUTE FOR BUSINESS VALUE. IBM. Maximo for oil and gas. 2016. [En línea]. Disponible en: <http://www-03.ibm.com/software/products/es/maximo-for-oil-and-gas>.

Figura 2. Fases de aplicación de la metodología de Big Data



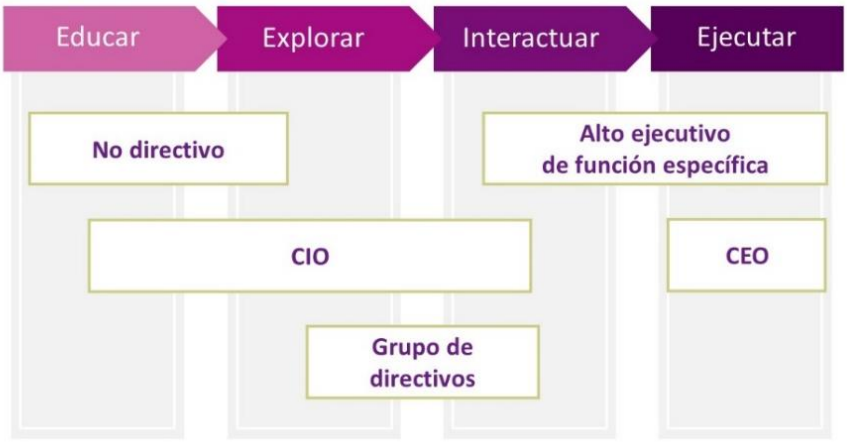
Adaptado de: http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf

Figura 3. Objetivos de las empresas que aplican la metodología



Adaptado de: http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf

Figura 4. Respaldo de la empresa requerido en cada fase

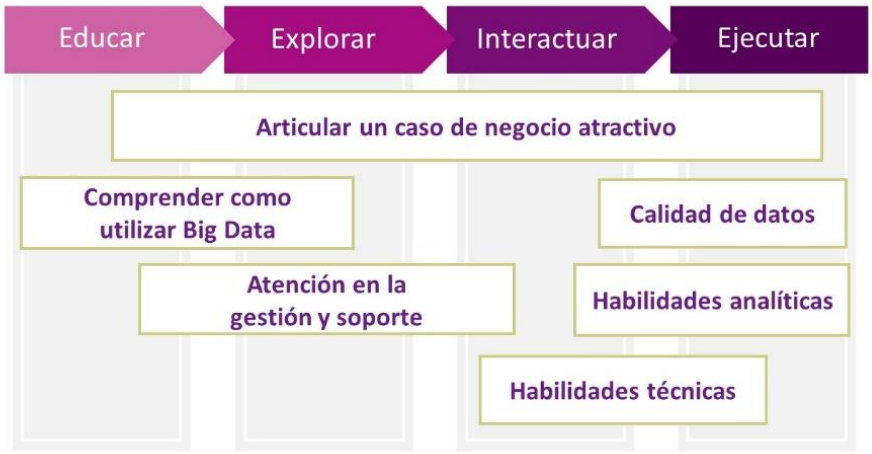


Adaptado de: http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf

Figura 5. Disponibilidad de datos requerido en cada fase

Adaptado de: http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf

Figura 6. Principales obstáculos presentados en cada fase



Adaptado de: http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf

1.4.2. SAS: empresa multinacional que opera actualmente en alrededor de 140 países y que fue fundada en 1976 en EE.UU., llegó a Colombia en 1998. Originalmente, SAS, que significa Statistical Analysis System, era simplemente un software desarrollado en la Universidad Estatal de Carolina del Norte como un proyecto para analizar la investigación agrícola[18].

SAS ha realizado avances en la aplicación que tiene la metodología Big data and Analytics en la industria de los hidrocarburos de manera profunda, esto se puede observar en el libro publicado por Keith R Holdaway, el experto en operaciones upstream para la unidad de negocios de petróleo y gas global de SAS, titulado "Harness Oil and Gas Big data with Analytics"[19].

[18] SAS. Company information. 2016. [En línea]. Disponible en: <http://www.sas.com/es_co/company-information.html#>

[19] Ibíd.

(*) Rate of penetration

(**) Revoluciones por minuto

La información más relevante que puede ser encontrada en este libro para el caso específico de la perforación es la búsqueda de la solución de problemas enfocados en la reducción de los NPT's y en la optimización de los parámetros de perforación como el ROP(*), RPM(**), torque y las condiciones de la broca[20].

1.4.3. Microsoft Azure: Azure es una plataforma de Microsoft que se basa en ofrecer servicios integrados en la nube, es decir que sirve como herramienta para la aplicación de la metodología Big Data. El software cuenta con dos versiones de compra, una estándar y una premium, la elección de cual usar depende del para que se quiera y su valor depende de la cantidad de clústeres o nodos que se requieran utilizar, la ventaja que presenta la versión Premium es la de ofrecer la herramienta predictiva, que es conocida como *R server*, para lo cual se deben utilizar los paquetes del A6 en adelante, estas herramientas son ilustradas en la figura 7[21].

Los paquetes se dividen en diferentes versiones y son compatibles con Linux y Windows. La primera es la versión A, en esta las clasificaciones de A1-A7 son las de uso general de *HDinsight*, mientras que las A10 y A11 son las de proceso intensivo las cuales cuentan con procesadores Intel Xeon E5 para clústeres de alto rendimiento, modelado y simulación, entre otros, aunque solo se encuentran disponibles en las regiones de EE.UU., Norte y Oeste de Europa y Japón Oriental. Las características de este paquete son ilustradas en la tabla 1[22].

Los otros paquetes son los de la serie D1V2, los cuales son nodos optimizados que se utilizan solo en computadoras de última generación ya que son un 35% más rápidos que los de la serie A y funcionan con los sistemas Intel Xeon E5 de 2.4 Ghz(*) e Intel turbo boost technology 2.0 de 3.2 Ghz, lo que no solo aumenta

[20] Ibíd.

[21] MICROSOFT AZURE. HDInsight. 2016 [En línea]. Disponible en: <https://azure.microsoft.com/es-es/services/hdinsight/>

[22] Ibíd.

(*) Gigahertz

su velocidad sino también su capacidad de almacenaje, el R server funciona con las versiones D4, D12, D13 Y D14, sus características se ilustran en la tabla 2[23].

Figura 7. Versiones estándar y premium de Microsoft Azure

Adaptado de: https://azure.microsoft.com/es-es/services/hdinsight/

Además, existe el servicio de máquinas virtuales, que hacen las veces de un emulador de computador, sin necesidad de hardware. En este caso se pueden encontrar en la serie A, del 0-4 el sistema básico, del 5-7 el estándar y del 8-11 el de alto procesamiento. El segundo sistema es el tipo DV1, en esta categoría podemos encontrar los estándares del 1-4 y los de alta capacidad de almacenamiento del 11-14. En el grupo DV2 los del 1-5 son de uso general mientras que los del 11-15 son de almacenamiento mejorado. Finalmente, los del grupo F son utilizados para procesos intensivos ya que emulan un sistema de 2GB de RAM y 16 GB de unidad de estado sólido (Sistema de almacenamiento no volátil) por nodo. Las características de los emuladores de cada grupo son mostradas en las tablas 3, 4, 5 y 6[24].

[23] Ibíd

[24] Ibíd.

Tabla 1. Paquetes de datos A de Microsoft Azure

Instancia	Núcleos	RAM (GB)	Tamaño de discos (GB)	Estándar (USD/m)	Premium (USD/m)
A1	1	1.75	70	59.52	74.40
A2	2	3.5	135	119.04	148.80
A3	4	7	285	238.08	297.60
A4	8	14	605	476.16	595.20
A5	2	14	135	260.40	290.16
A6	4	28	285	528.24	587.76
A7	8	56	605	1,049.04	1,168.08
A10	8	56	382	1,220.16	1,339.20
A11	16	112	382	2,261.76	2,499.84

Adaptado de: https://azure.microsoft.com/es-es/services/hdinsight/

Tabla 2. Paquetes de datos DV2 de Microsoft Azure

Instancia	Núcleos	RAM (GB)	Tamaño de discos (GB)	Estándar (USD/m)	Premium (USD/m)
D1V2	1	3.5	50	115.32	130.2
D2V2	2	7	100	231.38	261.14
D3V2	4	14	200	462.77	522.29
D4V2	8	28	400	924.79	1,043.83
D5V2	16	56	800	1,849.58	2,087.66
D11V2	2	14	100	282.72	312.48
D12V2	4	28	200	565.44	624.96
D13V2	8	56	400	1,017.79	1,136.83
D14V2	16	112	800	1,831.73	2,069.81

Adaptado de: https://azure.microsoft.com/es-es/services/hdinsight/

Tabla 3. Emulador tipo A de Microsoft Azure

Instancia	Núcleos	RAM (GB)	Tamaño de discos (GB)	Total (USD/m)
A0	1	0.75	20	74.40

A1	1	1.75	70	126.48
A2	2	3.50	135	252.96
A3	4	7	285	505.92
A4	8	14	605	1,011.84
A5	2	14	135	342.24
A6	4	28	285	684.48
A7	8	56	605	1,368.96
A8	8	56	382	1,480.56
A9	16	112	382	2,961.12
A10	8	56	382	1,279.68
A11	16	112	382	2,559.36

Adaptado de: https://azure.microsoft.com/es-es/services/hdinsight/

Tabla 4. Emulador tipo DV1 de Microsoft Azure

Instancia	**Núcleos**	**RAM (GB)**	**Tamaño de discos (GB)**	**Total(USD/m)**
D1V1	1	3.50	50	156.24
D2V1	2	7	100	312.48
D3V1	4	14	200	624.96
D4V1	8	28	400	1,249.92
D11V1	2	14	100	342.24
D12V1	4	28	200	684.48
D13V1	8	56	400	1,279.68
D14V1	16	112	800	2,398.66

Adaptado de: https://azure.microsoft.com/es-es/services/hdinsight/

Tabla 5. Emulador tipo DV2 de Microsoft Azure

Instancia	Núcleos	RAM (GB)	Tamaño de discos (GB)	Total(USD/m)
D1V2	1	3.50	50	156.24
D2V2	2	7	100	312.48
D3V2	4	14	200	624.96
D4V2	8	28	400	1,249.92
D5V2	16	56	800	2,499.84
D11V2	2	14	100	342.24
D12V2	4	28	200	684.48
D13V2	8	56	400	1,279.68
D14V2	16	112	800	2,398.66
D15V2	20	140	1,000	2,998.32

Adaptado de: https://azure.microsoft.com/es-es/services/hdinsight/

Tabla 6. Emulador tipo F de Microsoft Azure

Instancia	Núcleos	RAM (GB)	Tamaño de discos (GB)	Total(USD/m)
F1	1	2	16	136.15
F2	2	4	32	271.56
F4	4	8	64	543.86
F8	8	16	128	1,087.73
F16	16	32	256	2,176.20

Adaptado de: https://azure.microsoft.com/es-es/services/hdinsight/

1.5. Casos de aplicación actuales: para finalizar se presentará un caso en el que se aplicó la metodología para resolver un problema de la industria de los

hidrocarburos. En el 2013 Statoil implementó un proyecto basado en la metodología Big Data para el monitoreo del ambiente en el subsuelo 24/7, involucrando datos estructurados, semiestructurados y sin estructurar acerca del petróleo, la biología marina, regulaciones, etc.

El estudio fue realizado en conjunto por CGG y Teradata, el primero ofreció los datos de geociencia, pozo y perforación, mientras que Teradata ofreció un enfoque analítico iterativo que permitiera correr análisis complejos de información. Para este proyecto se siguió el esquema ilustrado en la figura 8[25].

La pregunta principal por responder era: ¿Por qué algunas perforaciones aumentan o disminuyen del tamaño nominal de la broca?
Para el estudio se utilizaron más de 300 pozos de la plataforma continental de Reino Unido y la información digital fue:

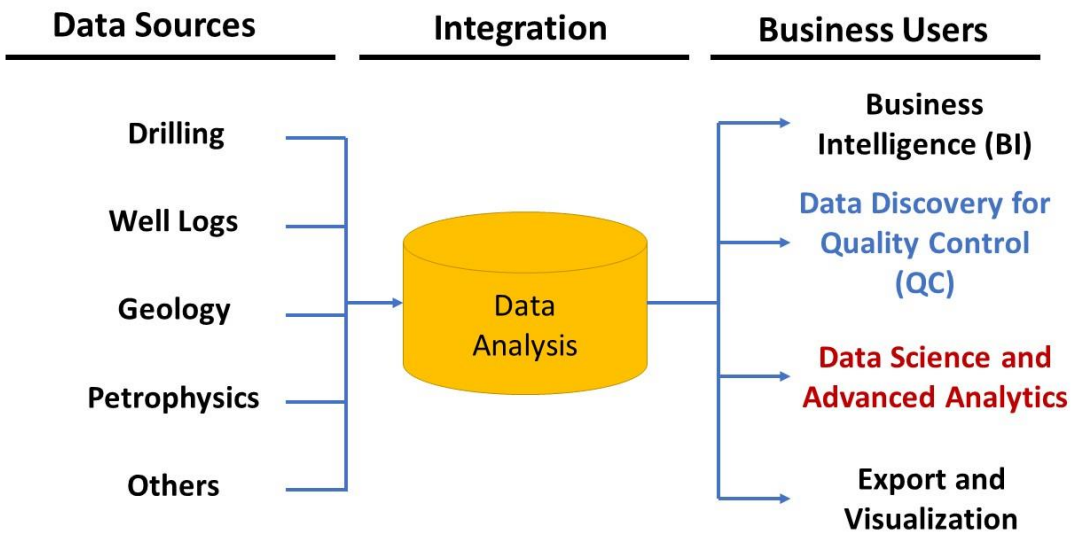
- Registros de pozo, incluido el caliper.
- Registros de perforación.
- WOB.
- ROP.
- Torque.
- Desviación.

Además de metadata(*) adicional como:

- Localización de los pozos.
- Formaciones litológicas.
- Tamaños de la broca.

[25] JONHNSTON, J; GUICHARD, A. New Findings in Drilling and Wells using Big Data Analytics. Offshore technology conference. Mayo de 2015.
(*) Datos que ofrecen información adicional relacionada a otros datos.

Figura 8. Esquema de trabajo aplicado por Statoil



Adaptado de: JONHNSTON, J; GUICHARD, A. New Findings in Drilling and Wells using Big Data Analytics

Los pozos se perforaron con tamaños de broca que disminuían al aumentar la profundidad junto a un conjunto de revestimientos en cada sección intermedia, los registros de pozo dan una medida del tamaño actual del pozo, el cual muchas veces es diferente al esperado, esto se puede deber a múltiples causas tanto geológicas como mecánicas[26]. Algunos ejemplos son:

- Formación compuesta por capas delgadas de arcillas y areniscas intercaladas, las cuales pueden colapsar, dejando sin soporte otras capas y generando washouts o ensanchamientos del hueco
- El uso de lodo base agua puede generar el ensanchamiento de arcillas si no se controla adecuadamente y generar una disminución en el diámetro del hueco

[26] Ibíd.

- Rugosidad de la pared, en donde el diámetro del hueco es similar al de la broca, pero en donde se presentan pequeñas variaciones periódicas que pueden hacer de algunas mediciones inservibles

En la figura 9 se presentan los resultados finales del estudio, en esta, la sección azul representa una calidad aceptable del hueco, mientras que la naranja una pobre calidad. El total de pozos fue de 347, de los cuales 85 en Chalk, 102 en Moray, 83 en Cromer y 77 en otras 6 formaciones.

Las herramientas utilizadas para este estudio permiten el análisis de los datos durante la marcha, sin necesidad de preconceptos o correlaciones precargadas, además de que el sistema trata todos los datos por igual, lo que permite que puedan ser comparados mediante una curva de desviación de entrada si/no. Algunos tipos de datos fueron: .TXT, .XLS, .PDF, .LAS[27].

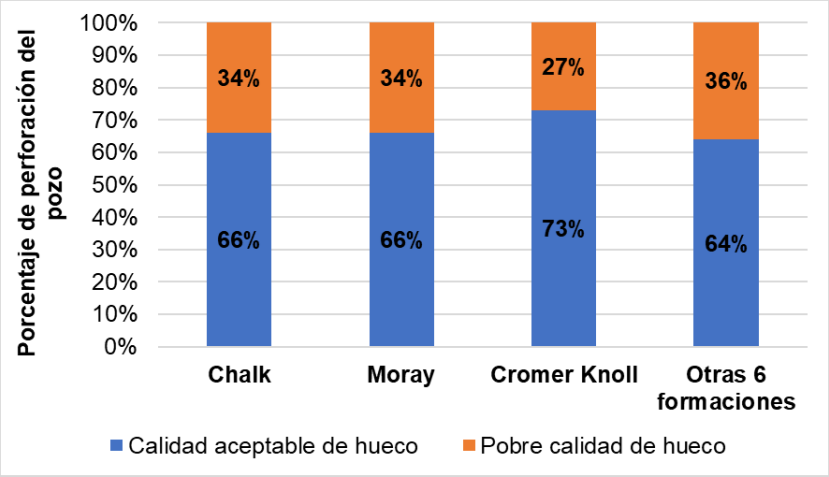
En el estudio se definió un Caliper diferencial (DCAL) que es igual a la diferencia entre el tamaño del hueco y el diámetro nominal de la broca. Cualitativamente se tomaron como pozos malos aquellos que tuvieran DCAL= Tamaño de la broca + 0.5” o tamaño de la broca – 0.2”. La parte más demorada del proceso fue la preparación de los datos, debido principalmente a la variedad de nombres que puede recibir cada registro, como, por ejemplo, el caliper que puede obtenerse como CAL, CALI, CAL1, CALR, CALS, C1 y muchos más, y que deben ser estandarizados antes de continuar el análisis. Para esto se utilizaron herramientas de inteligencia empresarial genéricas y una plataforma de Teradata Aster para la carga, limpieza y análisis de los datos[28].

En la figura 10 se puede observar un plot multivariable del WOB vs ROP para diversos pozos y formaciones, donde el tamaño de los puntos representa el diámetro del hueco:

[27] Ibíd.

[28] Ibíd.

Figura 9. Resultado final del estudio realizado por Statoil



Adaptado de: JONHNSTON, J; GUICHARD, A. New Findings in Drilling and Wells using Big Data Analytics

La lectura de la Figura 9 puede ser complicada para diversos pozos, por lo que se diseñó otro plot, ahora de WOB vs Torque, que puede ser observado en la figura 11, en este el tamaño de los puntos representa el DCAL y donde se representaron los pozos con buena calidad en color verde y los de mala calidad en naranja, esto aplicado a pozos tanto verticales como desviados, la prueba duró 6 semanas. Utilizando los resultados de la prueba se pueden generar un conjunto de parámetros de perforación optimizados para una formación específica en un área específica[29].

1.6. Big Data centrado a la industria de los hidrocarburos

Ahora, para continuar con el estudio de la aplicación de la metodología de Big Data Analytics a operaciones de perforación es necesario tener siempre presente los siguientes conceptos básicos explicados anteriormente:

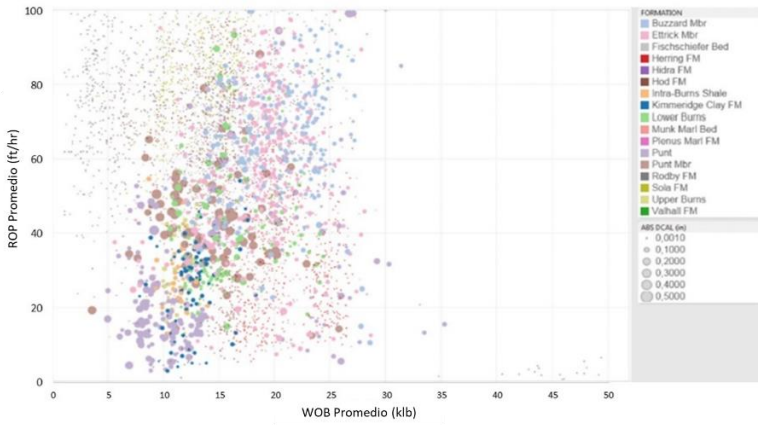
- **Big Data:** Todo conjunto de datos cuyo tamaño sea mayor a la habilidad de captura de las herramientas software típicas[30].

[29] Ibíd.

[30] [25] SPATH, Jeff. Big Data!. Revista JPT. Enero de 2014.

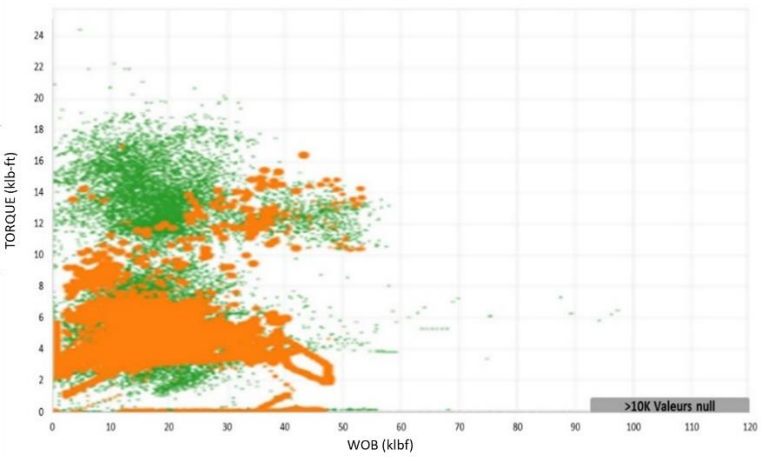
- **Metodología Big Data Analytics:** Describe a todo proceso en el que se examinan grandes volúmenes de datos de diversos tipos en un esfuerzo por reconocer patrones, correlaciones y otra información importante con el fin de extraer un valor económico[31].

Figura 10. WOB vs ROP



Adaptado de: JONHNSTON, J; GUICHARD, A. New Findings in Drilling and Wells using Big Data Analytics

Figura 11. WOB vs Torque



Tomado de: JONHNSTON, J; GUICHARD, A. New Findings in Drilling and Wells using Big Data Analytics

[31] FEBLOWITZ, Jill. Analytics in Oil and Gas: The Big Deal About Big Data. 2013 SPE digital energy conference. Marzo de 2013

Teniendo en cuenta lo anterior, se puede pasar a analizar el aporte realizado por el experto en perforación de SAS, el M.Sc. Keith Holdaway, autor del libro "Harness oil and gas Big Data with Analytics", donde explica ciertas generalidades de la aplicación de la metodología en la industria de los hidrocarburos, donde, para los fines de este trabajo, el capítulo de mayor interés es el quinto, en el cual habla de la parte de perforación y completamiento. Al estudiar esta sección lo primero que se debe tener en cuenta es que, para poder detectar ineficiencias en la perforación, y de esta manera poder optimizarla, es necesario analizar tendencias con el objetivo final de desarrollar las iniciativas estratégicas y soluciones operacionales que permitirán reducir los costos por sección, durante la construcción de pozos eficientes; es decir que la meta se resume en tres aspectos[32]:

- Reducir los NPTs.
- Optimizar los parámetros.
- Realizar interpretaciones en tiempo real.

Para alcanzar estos objetivos de la mejor manera, lo primero que se debe determinar son los calificadores de desempeño, que son aquellos parámetros que afectan de manera más directa a la eficiencia de la operación, a continuación, el efecto que tiene cada uno de ellos sobre los demás debe ser analizado para reconocer patrones ocultos y tendencias que ayuden en la toma de decisiones. Algunos de los parámetros más relevantes sugeridos por Holdaway son:

- Información Ofsset.
- Mecánica de las rocas.
- Propiedades del lodo.
- Diseño del BHA.
- Capacidad del taladro.

[32] HOLDAWAY, Keith. Harness oil and gas big data with analytics. 2014.

Estos son tomados como la información clave o mínima para realizar el análisis de cualquier problema de perforación, aun así, siempre resulta útil el uso de la mayor cantidad de información de la que se disponga, un ejemplo de ello puede ser el estudio de la reducción en la posibilidad de ocurrencia de una pega de tubería para la cual se pueden tener en cuenta datos como[33]:

- Información geomecánica.
- Información de los fluidos.
- Información litológica.
- Información de la dinámica del BHA.
- Información vibracional.
- MWD/LWD.
- Información de los equipos de superficie.

Finalmente se debe tener un mejoramiento continuo mediante la determinación de los rangos operacionales óptimos que permitan mejorar factores como:

- Calidad y estabilidad del pozo.
- Desempeño del taladro.
- Identificación de problemas en tiempo real.

Para el caso específico de la reducción de NPTs Holdaway recomienda seguir el siguiente procedimiento:

1. Integración de la información.
2. Interpretación.
3. Estimación de los NPTs.
4. Categorización del pozo.
5. Control y monitoreo de los NPTs.

En la parte de la categorización del pozo se deben tener en cuenta la frecuencia y probabilidad de ocurrencia del problema en ellos para determinar aquellos más críticos en donde la necesidad de remediación es mayor. Algunas compañías,

[33] Ibíd.

de acuerdo con sus objetivos, relacionan estos tiempos no productivos a determinadas empresas de servicios o cuadrillas para establecer la eficiencia de estas[34].

En cuanto a la optimización de parámetros se hace un énfasis en la determinación de la broca óptima a utilizar, para lo cual se deben tener en cuenta datos como:

- Registros.
- Topes de formación.
- Registros de lodo.
- Análisis de corazones.
- Mecánica de las rocas.
- Parámetros de perforación.
- Bit récords.

Además de un proceso evaluativo que cumpla pasos como:

- Evaluación de los tipos de formaciones esperados
- Acumulación de la información de pozos offset
- Determinación de esfuerzos compresivos no confinados en la roca, porosidad efectiva, características abrasivas y el potencial de impacto
- Identificación de los tipos de brocas potencialmente óptimos y varias características aplicables
- Predicción de costos por pies por cada broca potencial
- Recomendación de la broca optima

Posteriormente, se hace un análisis post-corrida que permita obtener una retroalimentación del desempeño de la broca seleccionada[35].

[34] Ibíd.

[35] Ibíd.

La realización de interpretaciones en tiempo real se refiere a comparar la información de perforación en tiempo real con las tendencias previas, esto permite:

- Evitar NPTs potenciales mediante la predicción de fallas, tales como una falla en el PDM (Motor de desplazamiento positivo)(*) debida a vibración excesiva
- Dirección geográfica: Realizar ajustes en tiempo real a la trayectoria del pozo
- Realizar cambios en tiempo real a los parámetros de perforación
- Prevenir reventones: Procesos iterativos multivariables para analizar presiones de formación, lodo y fluidos de perforación

Finalmente, lo que se busca es automatizar los procesos de[36]:

- Manejo de la información
- Calidad de la información
- Modelamiento predictivo y extracción de datos
- Reporte de resultados

A continuación, se presenta un caso de estudio que representa lo descrito anteriormente; en este se propone el uso de cadenas simbólicas de datos de perforación(**) para reconocer tendencias en los mismos y poder explotar realmente su valor, el estudio de estas cadenas simbólicas se basa en propiedades estadísticas como la media, varianza, asimetría, curtosis(***) y entropía, entre otras. Así el input(****) del problema son series de tiempo multivariable con nueve componentes {T1, T2…..T9}, donde Ti es una serie de

[36] Ibíd

(*) Motor de fondo que acciona diversas herramientas de fondo en perforaciones direccionales o de alto rendimiento

(**) Muy variados al involucrar múltiples procesos como el tiempo de conexiones, las rotaciones o el deslizamiento durante la perforación

(***) Concentración de variables alrededor de la media de distribución

(****) Información de entrada

números reales {X1, X2…. Xn} tomados secuencialmente en un periodo de tiempo específico[37].

En la tabla 7 se muestran algunas de los parámetros de entrada utilizados.

Tabla 7. Parámetros de entrada

PARÁMETRO	DESCRIPCIÓN
Flowinav	Velocidad de flujo promedio del lodo
Hkldav	Carga del gancho promedio
Mdbit	Profundidad medida de la broca
Mdhole	Profundidad medida del hueco
Prespumpav	Presión promedio de la bomba
Ropav	Velocidad de penetración promedio
Rpmav	Revoluciones por minuto promedio del taladro
Tqav	Torque promedio
Wobav	Peso sobre la broca promedio

Tomado de: HOLDAWAY, Keith. Harness oil and gas big data with analytics. 2014.

Estos datos de entrada requieren de un preprocesamiento y filtrado en donde se debe:

- Identificar y manejar los valores faltantes.
- Identificar y manejar los datos aislados.

En este caso específico la detección de los datos aislados se realizó mediante el método del rango inter-cuartil: IRQ= Q3-Q1, en donde un dato aislado será todo aquel que este 1.5 rangos intercuartiles por encimado o debajo del primer y tercer cuartil:

- X<Q1-1.5*IRQ

[37] Ibíd.

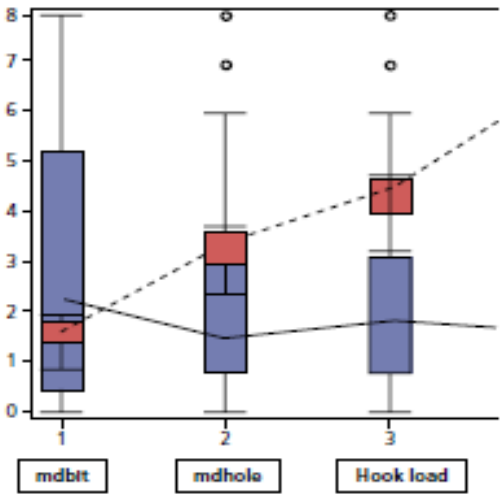
- X>Q3-1.5*IRQ

En la figura 12 se ilustra el procedimiento de filtrado del método de los rangos intercuartiles. Además de este filtrado se deben normalizar los datos para lo cual se dividió en la profundidad a aquellos que dependen de la misma. Teniendo en cuenta que i=1…10, los grupos principales de las medidas estadísticas calculadas para los Ti fueron[38]:

- Medidas de tendencia central: Media, mediana y moda.
- Medidas de variabilidad: Varianza, desviación estándar y rango.
- Medidas de forma: Asimetría, curtosis y segundo momento.
- Medidas de posición: Percentiles.
- Medidas de impureza: Entropía.

Un problema importante es que información con un alto nivel de dimensional puede generar una alta redundancia, lo que impactará de manera adversa el estudio, por ello se deben pasar las propiedades por un filtro de elección en donde se determinen las k propiedades más representativas[39].

Figura 12. Método de los rangos intercuartiles o diagrama de cajas y bigotes.



Tomado de: HOLDAWAY, Keith. Harness oil and gas big data with analytics. 2014.

[38] Ibíd

[39] Ibíd.

Para determinar cuántas propiedades utilizar para obtener la máxima precisión se deben correr varios modelos con diferentes propiedades y subsecuentemente aplicar un indicador de precisión en cada modelo, así que, finalmente se utilizan diferentes algoritmos de estudio como correlaciones o el método de chi cuadrado y, para este caso, cinco técnicas de clasificación, las cuales fueron:

- Red neuronal artificial (ANN).
- Inducción de reglas (RI).
- Árbol de decisión (DT).
- Bayes Naive (NB).
- Maquina apoyada en vectores (SVM).

Donde las técnicas más eficientes fueron la SVM y la RI[40].

Continuando con el estudio de Big Data Analytics resulta útil también tener en cuenta el aporte realizado por el Ph.D Erik Van Oort y RAPID, el grupo de investigación que dirige y que está centrado en este tema. RAPID es un grupo interdisciplinario de la universidad de Texas conformado por investigadores y estudiantes de ingeniería de petróleos, mecánica y aeroespacial (más de 30 integrantes) cuyas siglas significan Rig Automation and Performance Improvement in Drilling y cuyos objetivos y metas son[41]:

- Entregar soluciones de automatización para cualquiera y todos los aspectos de la construcción de un pozo.
- Reducir el tiempo de perforación y completamiento de pozos y su costo en más del 50%.
- Reducir el número de individuos en el taladro en más del 50%.

La educación dentro del grupo debe enfocarse en el manejo de múltiples datos, de esta forma, se centra en:

- Aprendizaje automático.

[40] Ibíd.

[41] VAN OORT, Eric. Drilling Optimization, Risk and Uncertainty Reduction, and Future Workforce Education Using Big Data Analysis. SPE. Febrero de 2016.

- Aprendizaje estadístico.
- Reconocimiento de patrones.
- Inteligencia artificial.
- Abastecimiento múltiple.
- Otros.

Además, se debe tener en cuenta que las principales fuentes de información son:

- Información en superficie.
- Sensores.
- Registros.
- Vibraciones.
- Datos en fondo.
- Entre otros.

Así, los principales obstáculos que se presentan y deben ser manejados son[42]:

- La información puede ser difícil y costosa de recolectar.
- Los estándares de los datos son usualmente deficientes.
- La recolección de la información puede ser una labor muy intensiva.
- Recursos dedicados de la compañía para el análisis pueden no estar disponibles.
- El tiempo de transformación puede ser muy largo para una optimización significativa.
- Debido a la antigüedad de los sensores, la información suele ser de baja calidad(*).
- La información recolectada por la industria del petróleo y gas es raramente utilizada debido a la ausencia de potencial humano capacitado para preparar y analizar los datos.
- Se almacenan inmensos volumenes de datos inservibles, que actúan como ruido, ocultando la información valiosa.

[42] Ibíd

(*) Dato indicado por un estudio de la universidad de Chesapeake

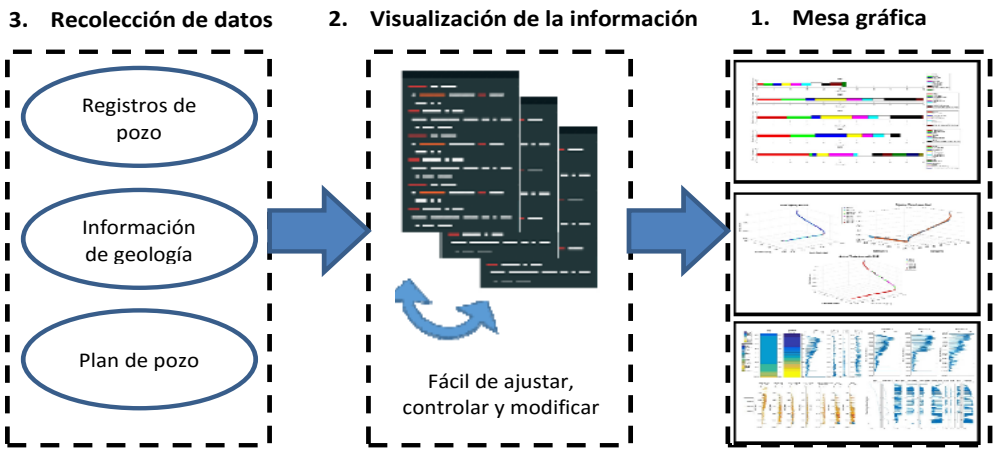
- Los algoritmos de análisis de datos y aprendizaje automáticos son inmaduros todavía en la industria del petróleo y gas.
- El principal problema relacionado a la información de perforación y completamiento es más el desorden presente en los datos que los grandes volumenes de información en sí.

El esquema de la metodología de trabajo que sigue el grupo se ilustra en la figura 13[43].

A continuación, se presentan un par de casos de estudio de análisis realizados a problemas de perforación:

A. Optimización de perforaciones en shale con datos de ingeniería comunes: El problema se centra en la variabilidad existente que se puede producir en la productividad de un pozo a otro, aun cuando compartan el mismo diseño en la misma formación; fenómeno conocido como "Naturaleza estadística del play(*)"[44].

Figura 13. Esquema de trabajo de RAPID



Adaptado de: VAN OORT, Eric. Drilling Optimization, Risk and Uncertainty Reduction, and Future Workforce Education Using Big Data Analysis. SPE. Febrero de 2016

[43] Ibíd

(*) Familia de yacimientos y/o prospectos los cuales tienen en común, la misma roca almacén, sello, y la misma historia de generación de hidrocarburos, migración y de carga

[44] LOGAN, W. D. Engineered Shale Completions Based on Common Drilling Data. SPE. 2015

Para abordar este problema se empieza por el cálculo de la Energía mecánica específica (MSE) y, a partir de esta, del UCS(*) que es igual a la MSE multiplicada por la eficiencia de la transmisión del poder de penetración del taladro a la roca (Deff). Esto con el fin de generar perforaciones en zonas con un UCS o MSE parecido[45]. Según Teale:

$$\text{MSE}= \left(\frac{480*T*RPM}{D^2*ROP}+\frac{4*WOB}{D^2*\Pi}\right) \quad \textbf{(Ecuación 1)}$$

Donde:

WOB: Peso sobre la broca (klb)

T: Torque (kft-lb)

RPM: Velocidad de rotación (rev/min)

D: Diámetro del hueco (in)

ROP: Velocidad de penetración (ft/hr)

Esta ecuación debe ser modificada en pozos donde se utilicen motores de fondo, así:

$$N' = RPM + (K_N * Q) \quad \textbf{(Ecuación 2)}$$

Donde:

K_N: Velocidad del motor de lodo a la rata de flujo (rev/gal)

Q: Flujo de lodo total (gal/min)

Y;

$$T' = \frac{T_{max}}{P_{max}} * \Delta P \quad \textbf{(Ecuación 3)}$$

Donde:

T_{max}: Torque máximo nominal del motor de lodo (ft-lb)

P_{max}:Delta de presión máxima nominal del motor de lodo (psi)

(*) Unconfined Compresive Strenght

[45] Ibíd.

De esta manera la ecuación queda finalmente como:

$$MSE = \frac{480 * \frac{T_{max}}{P_{max}} * \Delta P * (RPM + (K_N * Q))}{D^2 * ROP * 1000} + \frac{4 * WOB}{D^2 * \Pi} \qquad \textbf{(Ecuación 4)}$$

También se deben tener en cuenta ciertas correcciones:

- Promedio: Se promedian las lecturas cada 3 ft (Se toman cada ft) para pozos horizontales (En verticales se pueden necesitar menos promedios realizados).
- Picos: Corregir los wirelogs/LWD si se puede, ya que son muy propensos a generar picos, sobre todo en zonas de gas.
- Determinar valores faltantes en las mediciones.
- Limitar las ROP de tal manera que no se acerque demasiado a cero ya que esto no permitiría el uso de la ecuación.
- Asegurarse de que las RPM sean cero cuando no se está rotando, si no, asumirlo.
- Ya que el delta de presión es fácilmente alterable y debe ser corregido por el perforador para asegurarse de que no den valores negativos cuando no se está rotando, en el estudio se prefirió tomar la presión de la tubería vertical (SPP o Standpipe pressure) y corregirlo por:
    - Fuerzas de fricción entre los fluidos y la tubería
    - Cambios en el caudal de lodo
    - Cambios en la densidad del fluido de perforación

En la tabla 8 se muestra una clasificación de la dificultad en la perforación de las diferentes litologías para este estudio y su MSE relacionado de acuerdo a una escala de colores[46].

[46] Ibíd.

Tabla 8. Escala de colores para el MSE

COLOR	DUREZA	MSE
	HD1	0 - 15K
	HD2	15K - 30K
	HD3	30K - 50K
	HD4	50K - 75K
	HD5	75K - 100K
	HD6	100K - 125K
	HD7	125K - 150K
	HD8	150K - 175K
	HD9	175K - 200K
	HD10	200K - 225K
	HD11	225K - 250K
	HD12	250K - 300K
	HD13	300K - 400K
	HD14	400K - 500K

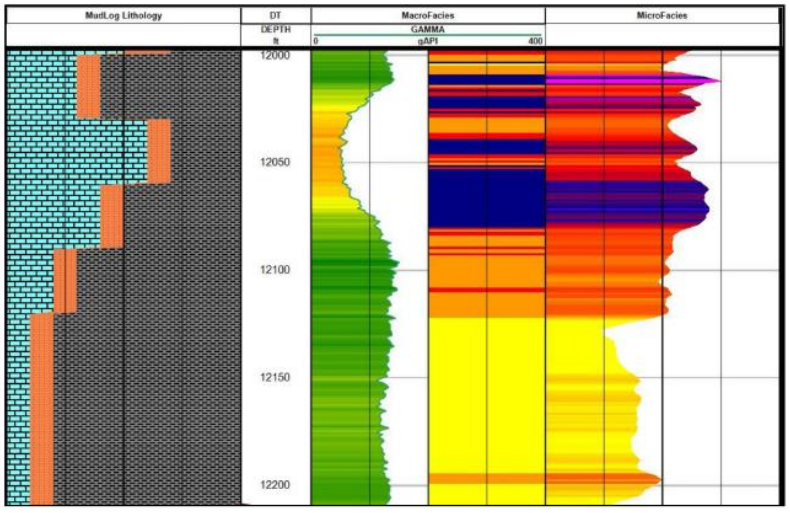
Aumento de la dureza

Adaptado de: LOGAN, W. D. Engineered Shale Completions Based on Common Drilling Data. SPE. 2015

Seguido a esto es necesario determinar el sistema de facies según la escala, tal y como se muestra en la figura 14. Posteriormente se definen para el completamiento:

- Longitud de cada etapa.
- Numero de perforaciones (Clusteres) por etapa.
- Distancia entre clústeres.

Figura 14. Sistema de facies relacionado

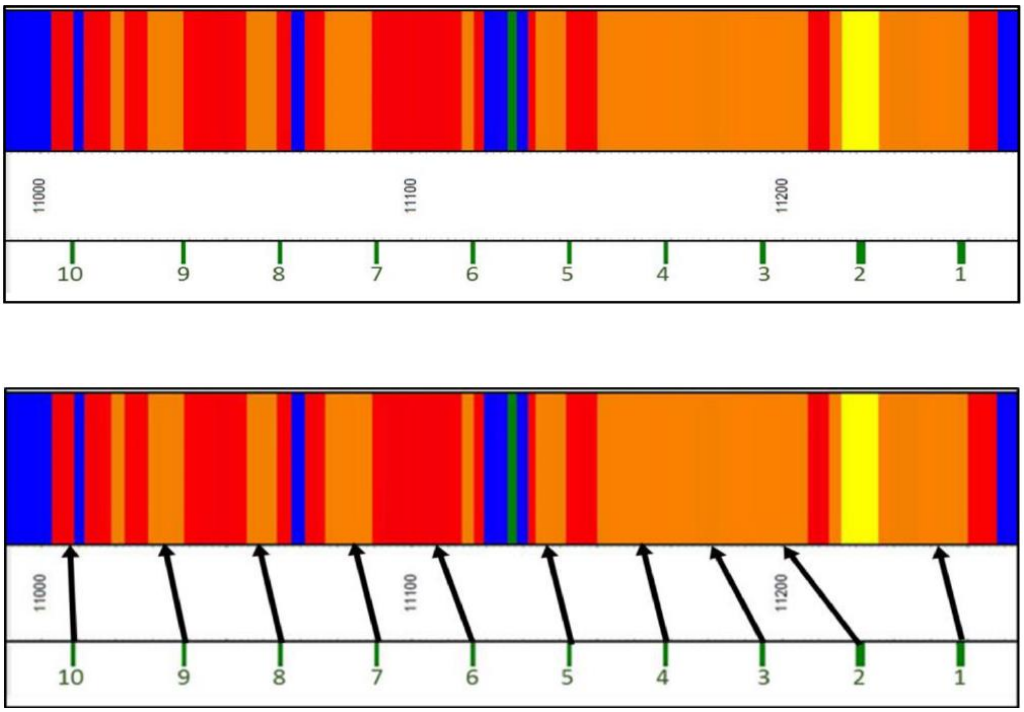


Adaptado de: LOGAN, W. D. Engineered Shale Completions Based on Common Drilling Data. SPE. 2015

Finalmente se sugiere el cambio en la ubicación de las perforaciones, etapa por etapa, enumeradas de la 1 a la 10, para que se ubiquen en zonas con MSE similares, de tal manera que la fractura no genere más esfuerzo en ciertas secciones que en otras(*) tal como se ilustra en la figura 15[47].

B. Big Data Anlytics para previsión pronóstica: En este estudio se señala la importancia que tiene la adaptación de las soluciones a las condiciones del campo como, por ejemplo, conexiones a internet lentas o inexistentes; y se centra en la evaluación para el mantenimiento preventivo de equipos, específicamente un compresor de gas[48].

Figura 15. Sugerencia final para las perforaciones



Tomado de: LOGAN, W. D. Engineered Shale Completions Based on Common Drilling Data. SPE. 2015

(*) Se debe tener especial cuidado con los datos de las zonas donde se deslizo la broca

[47] Ibíd.

[48] VON PLATE, Moritz. Big Data Analytics for Prognostic Foresight. SPE. 2016.

De esta manera, es importante tener en cuenta que actualmente la industria cuenta con varios mecanismos de predicción para determinar anomalías (Modelos basados en similitudes, reconocimiento de patrones e inferencia estadística), pero ellos no ofrecen previsiones objetivas específicas. Por lo tanto, a la hora de realizar pronósticos, la mayoría de los operadores se basan en el juicio subjetivo de los expertos[49].

Big Data Analytics ofrece soluciones de pronosticación y no de predicción, es decir que se basa en métodos estocásticos(*), evaluando diversos escenarios y dando un resultado cuantificable que permita determinar el óptimo. En la tabla 9 se pueden observar todo lo que involucra este proceso y en la tabla 10 se muestran algunos usos que se le puede dar al análisis de pronósticos[50].

Es importante aclarar que el principal aporte que ofrece este método es responder el "cuando", por ejemplo, cuando reemplazar un equipo, mientras que los métodos actuales responder el "que", "donde", "porque" y "como solucionarlo"[51].

El resultado final es el paso de una curva P-F(**), la cual explica el desarrollo en la falla de un equipo en retrospectiva; por una curva RUL(***), la cual señala las probabilidades de ocurrencia de una falla en un equipo, estas se ilustran en las figuras 16 y 17.

Es importante señalar que para la aplicación de la metodología de Big Data Analytics actual, se tomaran datos de un campo colombiano, se calculará el MSE en sus pozos para finalmente, mediante un análisis multivariable, determinar cuáles son los parámetros operacionales que deben ser modificados en los mismos para optimizar el tiempo de perforación en el campo.

[49] Ibíd

(*) Tienen en cuenta la probabilidad y se evalúan mediante métodos estadísticos

[50] Ibid.

[51] Ibíd

(**) Potencial de falla(P)-Falla(F)

(***) Remaining useful time o tiempo de vida remanente

Finalmente, se toma la decisión de seguir un procedimiento de filtrado estadístico inicial de los datos mediante el método de rangos intercuartiles sugerido por Holdaway, para posteriormente con estos datos realizar la medición de la Energía Mecánica Especifica (MSE) como parámetro base en los diferentes pozos de un campo, como es realizado en uno de los casos de estudio presentados anteriormente, agrupándolos de acuerdo con su parecido en cuanto a dirección, profundidad y formaciones atravesadas.

Tabla 9. Fases del proceso de pronósticos

	Monitoreo	Diagnóstico	Pronóstico
Fuente	• Sensores y equipos de recolección de datos, mediciones laser, lecturas de temperatura y vibración • Software para ilustración de data dura y mapeado	•Software avanzado basado en modelos estadísticos genéricos o específicos del equipo •Consultores especializados, expertos en los equipos y técnicos de campo	•Aplicaciones de procesos estocásticos transferidos de otros sectores como finanzas y cuidado de la salud aplicado al manejo de activos industriales •Inteligencia en comportamiento de activos provista por expertos
Funciones	• Operadores de alerta ante cualquier cambio en las condiciones que pueda requerir atención	•Procesamiento de la data dura para explicar el estado y comportamiento actual del activo	•Análisis continuo, condición histórica y procesamiento de la data para pronóstico disponible en el tiempo

	• Sirve como input para diagnóstico y pronóstico	•Muestra tendencias de tiempo •Genera advertencias poco antes de un malfuncionamiento o defecto	•Computa perfiles de riesgo de malfuncionamiento futuro •Dispone diferentes escenarios para la optimización de operaciones y cronogramas de mantenimiento
Horizonte	Ninguno	Horas a días	Semanas a meses
Beneficio	Alerta y alarma para respuesta inmediata	Visión para dar órdenes de trabajo	Previsión para planeación a largo tiempo

Adaptado de: VON PLATE, Moritz. Big Data Analytics for Prognostic Foresight. SPE. 2016

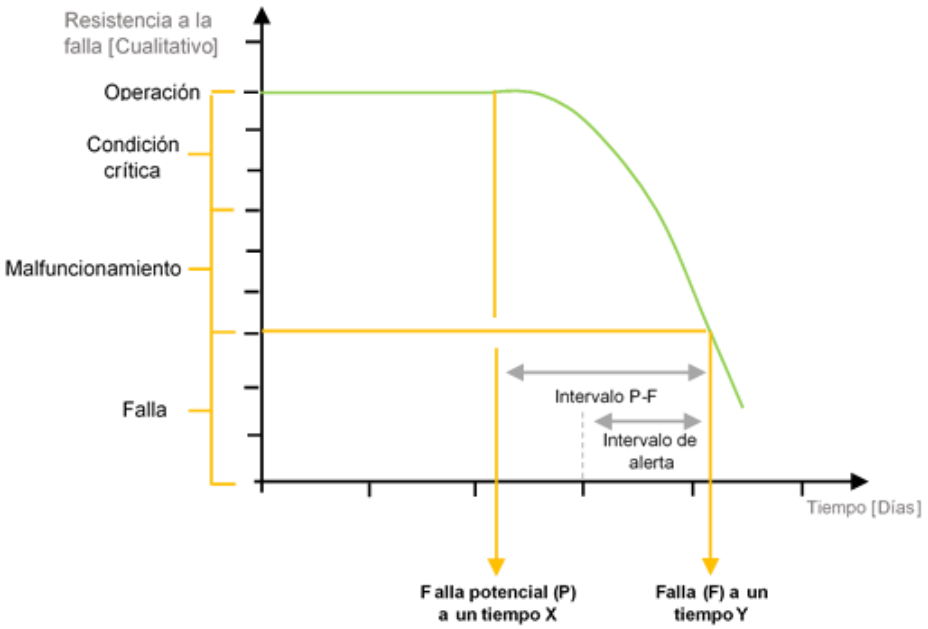
Tabla 10. Usos de los pronósticos

Categoría	Casos de estudio
Mantenimiento y reparación	• Cronograma de mantenimiento a largo plazo •Preparación de mantenimiento reactivo a corto plazo •Planeación y localización del personal de mantenimiento
Operaciones	•Planeación de la producción de acuerdo al perfil futuro disponible •Aumento en la disponibilidad de activos y minimización del tiempo de inactividad en actividades upstream

Finanza	• Oportunidades de ahorro de beneficio por la reducción de la inactividad • Costos de mantenimiento anual disminuido • Aumento en los beneficios totales esperados por los activos petroleros de las operadoras • Política y costos de aseguramiento optimizados
Manejo del ciclo de vida	• Planeación de remplazo y readaptación • Explotación RUL optima

Adaptado de: VON PLATE, Moritz. Big Data Analytics for Prognostic Foresight. SPE. 2016

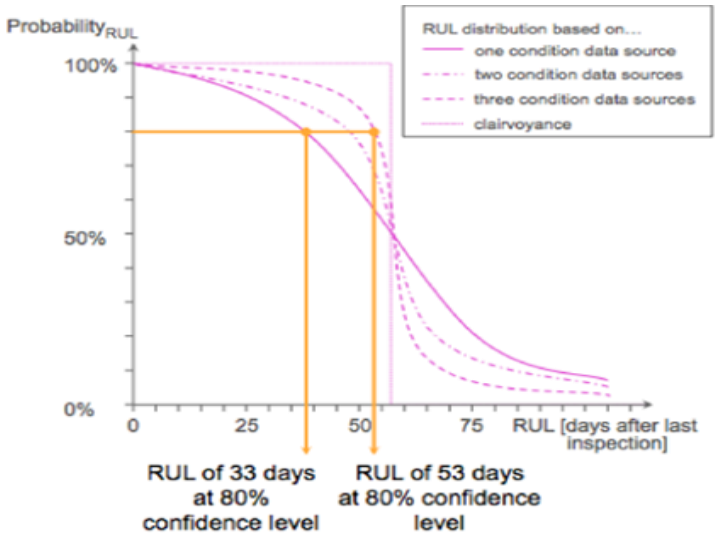
Figura 16. Curva P-F



Adaptado de: VON PLATE, Moritz. Big Data Analytics for Prognostic Foresight. SPE. 2016

El siguiente paso consiste en seleccionar la MSE óptima, es decir, la menor de cada grupo, en cada sección de los pozos, mediante un análisis de sensibilidad, para con base en esta poder seleccionar parámetros a ser mejorados en los diferentes proyectos, que permitan un mayor acercamiento al valor de MSE seleccionado, lo que se traduciría en una mayor eficiencia y efectividad en las perforaciones, y con ello en menores costos para la empresa.

Figura 17. Curva RUL



Tomado de: VON PLATE, Moritz. Big Data Analytics for Prognostic Foresight. SPE. 2016

Cabe aclarar que los parámetros a ser mejorados serán seleccionados más adelante, basados en la información disponible y en una preselección de aquellos que resulten más relevantes para las operaciones.

1.7. Redes neuronales

Dos opciones evaluadas para la programación de la metodología a ser propuesta fueron las del uso de redes neuronales artificiales o de la programación genética,

las cuales tienen una estrecha relación entre sí ya que tienen como fin desarrollar en la maquina un proceso de aprendizaje análogo al del cerebro humano mediante el uso de algoritmos genéticos, que no son más que algoritmos más complejos y estructurados, en estos las variables iniciales son relacionadas entre sí mediante el uso de varias capas, conformadas a su vez por diferentes "neuronas", donde se integran operaciones aritméticas, funciones matemáticas, funciones lógicas y/o funciones propias del dominio, de tal forma que, dependiendo del problema particular, se pueden tener diferentes tipos de redes, de acuerdo al tipo de dato resultante al final de la red, como son[52]:

- Valor booleano
- Valor integrado
- Valor real
- Valor complejo
- Valor tipo vector
- Valor simbólico
- Valor múltiple

Para entender de forma más clara el concepto base de una red neuronal artificial es importante relacionar su estructura con un órgano humano muy familiar, nuestro cerebro. El cerebro humano está compuesto por millones de neuronas interconectadas entre sí, las cuales se comunican a través de impulsos eléctricos. De acuerdo con el esquema morfológico del cerebro humano, las neuronas biológicas poseen diferentes partes, cada una de ellas con un propósito específico, en la figura 18 se presenta un esquema general de las partes de dicha célula[53].

De esta forma es posible definir una Red Neuronal Artificial, o por sus siglas en ingles Artificial Neural Network (ANN) como un modelo computacional no linear

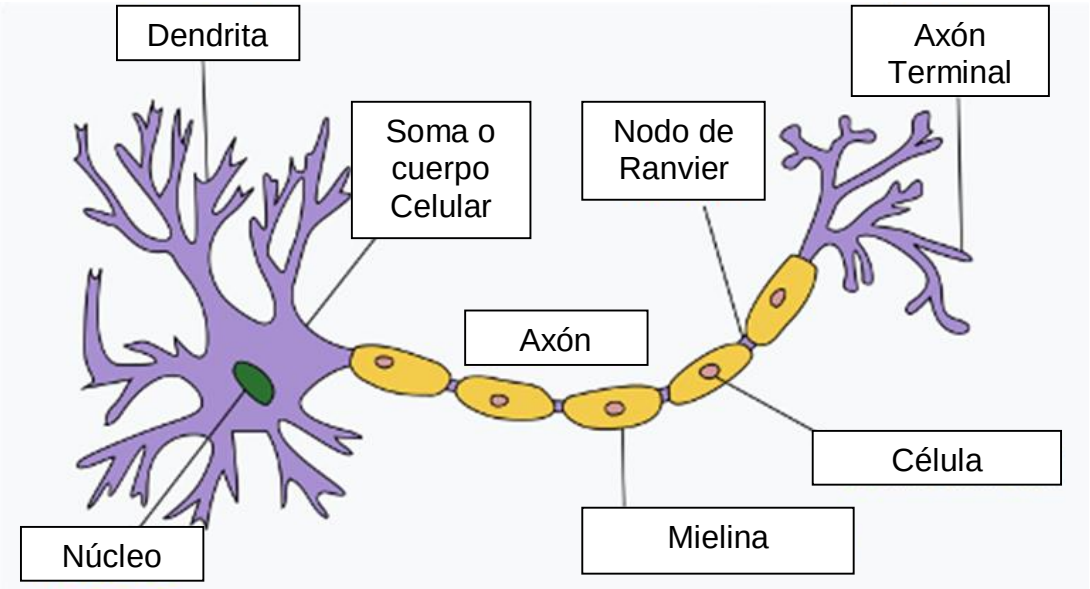
[52] KOZA, Jhon R. Genetic programming: on the programming of computers by means of natural selection. Enero 1992.

[53] DAVYDOVA, Olga. 7 types of Artificial Neural Networks for Natural Language Processing. Septiembre de 2017. [En línea]. Disponible en: < https://medium.com/@datamonsters/artificial-neural-networks-for-natural-language-processing-part-1-64ca9ebfa3b2>

basado en el funcionamiento de la Red Neuronal biológica en el cerebro humano, la cual es capaz de aprender a resolver tareas como la clasificación, predicción, toma de decisiones, visualización y otras tareas con el simple hecho de considerar ejemplos previos[54].

En la figura 19 se muestra el esquema general de una neurona computacional cuyas partes se relacionan con la célula humana. Respecto a esta última, como se observa en la figura 18, las dendritas se ramifican desde el soma en forma de árbol y se vuelven más delgadas con cada rama, recibiendo señales o impulsos de otras neuronas en la sinapsis. El axón también abandona el soma y generalmente tiende a extenderse por distancias más largas que las dendritas, y tiene como función enviar las señales de una neurona a otra[55].

Figura 18. Esquema de una célula neuronal humana.



Modificado de: Neuronal Networks. Python course. 2018

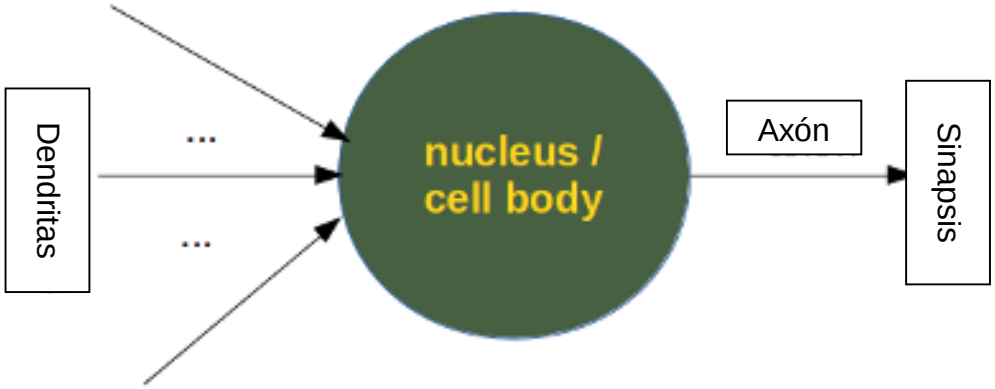
De igual manera, en la neurona computacional, la información o datos input llegan a través de una entrada, que se comportaría de manera análoga a las dendritas de las neuronas humanas, donde cada variable input posee un valor

[54] Ibíd.

[55] Ibíd

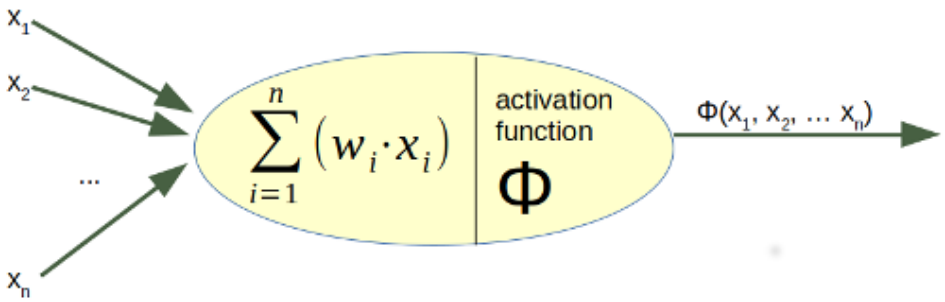
de peso, específico para el proceso. Si por un mismo canal de entrada están ingresando diferentes variables, se conoce como vector el cual tendrá su respectivo valor peso en el proceso, en relación con las variables que por él estén entrando al núcleo de la neurona, los cuales son modificados por la neurona a lo largo del proceso de aprendizaje. Después de esto, las señales de entrada modificadas se suman para, finalmente, determinar la salida real. Para este fin se aplica una función φ de activación o función de trasferencia, la cual asigna un valor de peso específico a la suma ponderada de los valores de entrada y entrega a la siguiente neurona la data Output. La figura 20 presenta un esquema grafico del proceso previamente mencionado[56].

Figura 19. Esquema de una neurona computacional análoga con partes de una neurona humana.



Modificado de: Neuronal Networks. Python course. 2018

Figura 20. Neurona computacional: Input de data, procesamiento en el núcleo y Output de data.



[56] Ibíd

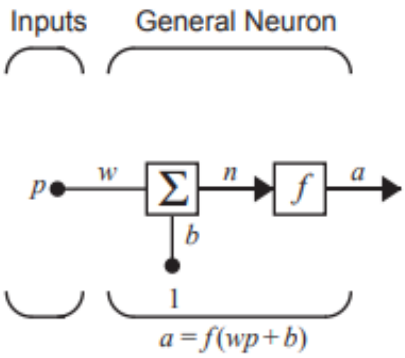
Modificado de: Neuronal Networks. Python course. 2018

Sin embargo, independientemente del tipo de red a utilizar, en general, la programación de toda red neuronal artificial se basa en dos pasos fundamentales que son la codificación inicial, donde se asignan pesos a las diferentes neuronas, y el entrenamiento, siendo este el factor diferenciador de esta técnica. En la fase de entrenamiento se busca variar los pesos de las diferentes neuronas de manera dinámica de tal manera que la maquina sea capaz de alcanzar una solución que presente el mayor ajuste posible a la realidad, al utilizar tanto los valores de validación suministrado como valores nuevos introducidos[57].

1.7.1. Tipos principales de redes neuronales artificiales

1.7.1.1. Neurona con una sola entrada: En la figura 21 se puede apreciar el esquema de una neurona computacional con una sola entrada de datos, o single-input. El valor escalar p es multiplicado por el valor de peso w, transformándose en un escalar diferente wp el cual es enviado al cuerpo de la neurona. La otra entrada que se ve en la imagen, cuyo valor es uno, corresponde a un factor de ajuste llamado bias. Este factor es usado en las redes neuronales para ajustar el proceso de aprendizaje y lograr resultados específicos que se asemejen a los valores esperados en el output de la red.

Figura 21. Neurona single-input.

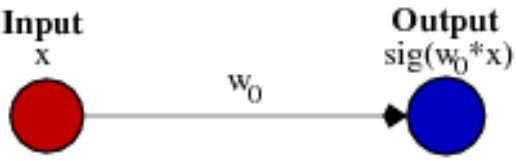


Tomado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

[57] GALVIS C, Laura V; et al. Estimación de propiedades mecánicas de roca utilizando inteligencia artificial. Diciembre 2011.

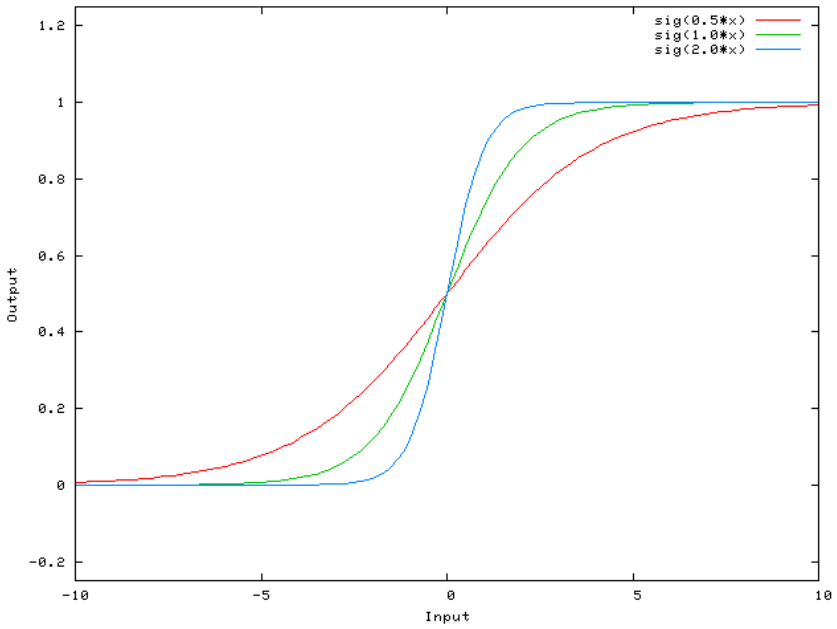
En las figuras 22, 23, 24 y 25 se presentan ejemplos sobre el funcionamiento del bias en la red neuronal. Para entender mejor este concepto, en la figura 22, se recurre al uso de una red neuronal muy simple, conformada por una única neurona a la cual llega solo un valor input llamado “x”. El dato de salida output es computado mediante la multiplicación por un valor de peso Wo a través de la función de transformación usada, sig. La grafica de la figura 23 presenta las diferentes curvas para 3 valores distintos de Wo[58].

Figura 22. Neurona con un solo Input sin el uso del Bias.



Tomado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

Figura 23. Curvas de resultados para 3 valores de peso Wo sin el uso del Bias.



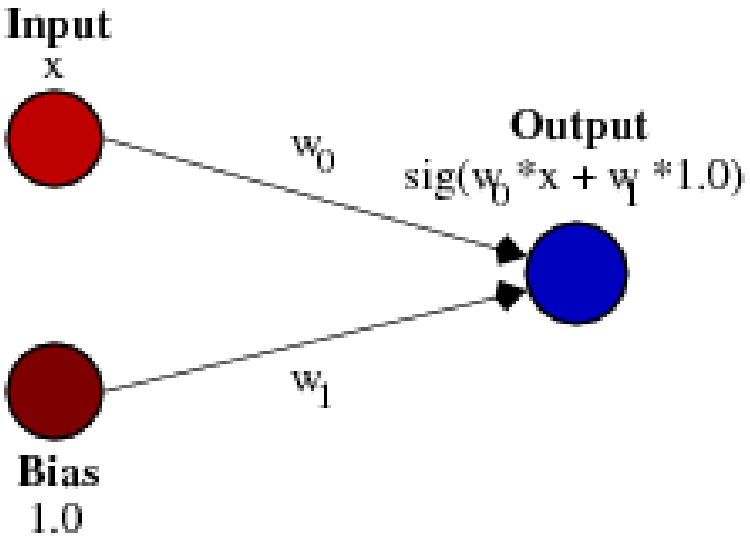
Tomado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

[58] Ibíd

El cambio en la ponderación del valor Wo le confiere cierta flexibilidad al proceso de aprendizaje de la red, sin embargo, si se requiere alcanzar un output específico, el cambio únicamente en la ponderación de este valor se hace ineficiente, por lo que es necesario introducir el bias, tal como se muestra en la figura 24[59].

Este nuevo termino agregado a la neurona altera la ecuación de transformación permitiendo a la red, para el caso citado, trasladar las curvas resultado hasta que el mismo proceso garantice, para un valor de entrada "x" especifico, obtener el valor de output deseado. Esto se puede apreciar en la gráfica de la figura 25 en donde para un mismo valor de peso Wo la red neuronal modifica el factor Bias hasta obtener el valor output deseado con un mismo valor "x" de entrada[60].

Figura 24. Neurona con un solo Input sin el uso del Bias.



Tomado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

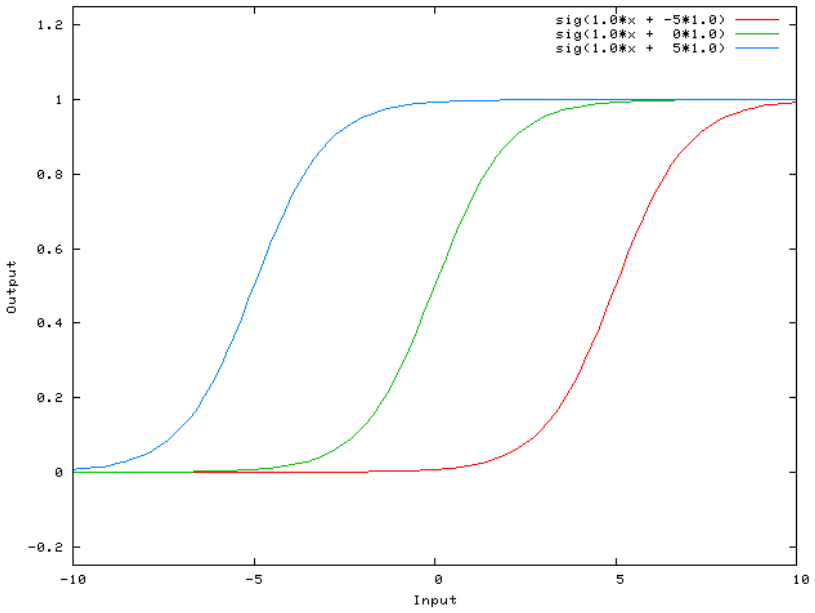
[59] Ibíd.

[60] HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

También es importante tener en cuenta que[61]:

- El factor Bias se agrega para aumentar la flexibilidad del modelo a la hora de ajustarse a los datos. Específicamente, permite que la red se ajuste a los datos cuando todas las características de entrada son iguales a 0, lo que, muy probablemente, disminuya el sesgo de los valores ajustados en otra parte del espacio de datos en la red neuronal.
- Normalmente, se agrega un solo bias para la capa de entrada y cada capa oculta en una red de tipo “feedforward”. Nunca se debe agregar dos o más bias a una capa determinada, pero una capa podría no contar con el factor bias si se decide eliminar de la red. El número total está, por lo tanto, determinado en gran medida por la estructura de su red.

Figura 25. Curvas de resultados con el valor de peso Wo constante y distintos valores del factor Bias.



Tomado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

[61] Ibíd.

Teniendo en cuenta todo lo anterior, es posible resumir que la neurona computacional transforma el dato input p junto con valor de peso w y el factor bias en un resultado bruto denominado n. Este valor output n pasa a través de una función de transformación la cual se encarga de convertir el resultado n en el valor de salida a. Este resultado final es calculado mediante la siguiente función[62]:

$$a=f(wp+b) \qquad \textbf{(Ecuación 5)}$$

Donde ambos escalares, w y b, son parámetros modificables de la neurona. Normalmente la función de transferencia para la totalidad de la red neuronal debe ser escogida por el programador y luego los parámetros w y b serán ajustados en la fase de entrenamiento de la red enfocados a un resultado específico[63].

Es importante hablar también, en este sentido, de las funciones de transferencia, las cuales tienen como objetivo transformar un resultado simple “n”, en un comportamiento o tendencia “a”[64].

La función de activación o función de transferencia, f, en la figura 20, puede ser del tipo linear o no linear para el dato de salida n. Esta función de transformación es escogida con el propósito de satisfacer alguna especificación requerida del problema que intentara resolver la red neuronal artificial. Existen diferentes tipos de funciones de transferencia de acuerdo con la finalidad de la red artificial, pero tres de estas son las más comúnmente usadas[65]:

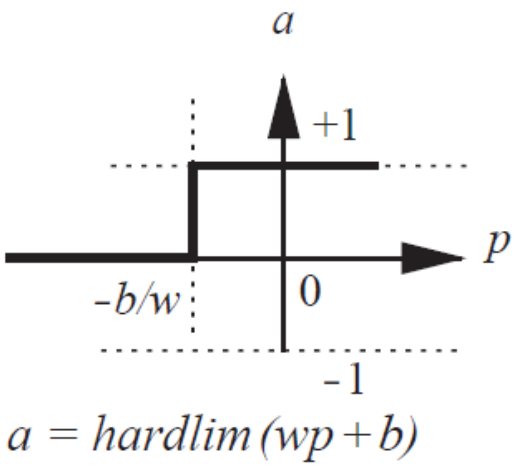
- La Función de transferencia de Limite Rígido (Hard Limit) permite transformar el dato de salida de la red en un valor de cero (0) si el argumento o variable independiente de la función es menor a dicho valor, o en un valor uno (1) si el argumento es igual o mayor a cero.

[62] Ibíd.
[63] Ibíd.
[64] Ibíd.
[65] Ibíd.

De acuerdo con la figura 20, nuestro argumento será n = wp+b. La figura 26 muestra en que valor nuestro argumento comenzara a tomar valores iguales o mayores a cero para n = -b/w.

Las funciones de transferencia del tipo Hard Limit son usadas cuando se quiere crear redes neuronales que clasifiquen datos de en dos únicas categorías diferentes ya que se basan en el uso del código binario como base de su funcionamiento.

Figura 26. Neurona con una sola entrada usando la función Hard Limit



Tomado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

- La función de transferencia linear arroja un output igual al dato de salida de la neurona sin transformar, es decir, a=n. Para explicar mejor este tipo de función se tiene ayuda de la figura 27 la cual muestra una neurona con dos entradas, p1 y p2, donde el dato de salida estaría condicionado como una función linear de los datos de entrada de la siguiente manera:

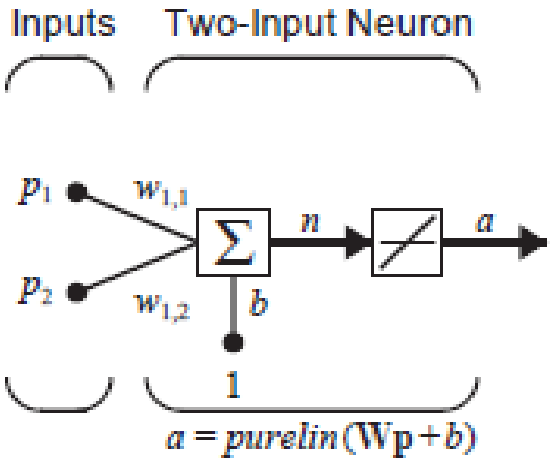
$$a = f.Linera(n) = f.Linear\left(\left[\sum w_i p_i\right] + b\right) = \left[\sum w_i p_i\right] + b = w_{1.1}p_1 + w_{1.2}p_2 + b \quad \textbf{(Ecuación 6)}$$

- La función de transferencia *Log-Sigmoid* permite convertir cualquier dato Input, entre – infinito e infinito, en un dato Output dentro del rango de 0 a 1 de acuerdo con la expresión mostrada a continuación:

$$a = \frac{1}{1 + e^{-c*n}} \qquad \textbf{(Ecuación 7)}$$

Este tipo de función es usada comúnmente en redes neuronales más complejas que usan el algoritmo *Backpropagation* como método de entrenamiento. La tabla 11 presenta las funciones previamente mencionadas junto con otros tipos comúnmente usadas.

Figura 27. Neurona con dos entradas usando la función de transferencia linear.



Tomado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

1.7.1.2. Neurona con múltiples entradas: Normalmente en las redes artificiales, las neuronas tienen más de una entrada. Cada una de estas entradas está ajustada a un peso específico. Si bien se pueden tener varios datos de entrada, cada neurona solo posee un factor *Bias*. El dato de salida sin ajustar por la función de transformación correspondería a una suma del Bias con los diferentes datos de entrada con sus respectivos pesos. Debido a que una

neurona puede tener muchas entradas, dichos Inputs son agrupados en una matriz llamada Wp como se presenta a continuación[66]:

Tabla 11. Funciones de transferencia más comunes.

NOMBRE	RELACION INPUT/OUTPUT	ICONO
Limite Rígido (Hard Limit)	a = 0 n < 0 a = 1 n ≥ 0	
Limite Rígido Simétrico	a = -1 n < 0 a = +1 n ≥ 0	
Lineal	a = n	
Lineal Saturado	a = 0 n < 0 a = n 0 ≤ n ≤ 1 a = 1 n > 1	
Lineal Saturado Simétrico	a = -1 n < -1 a = n -1 ≤ n ≤ 1 a = +1 n > 1	
Log-Sigmoid	$a = \frac{1}{a + e^{-n}}$	
Tangente Hiperbolica-Sigmoid	$a = \frac{e^{n} - e^{-n}}{e^{n} + e^{-n}}$	
Lineal Positivo	a = 0 n < 0 a = n 0 ≤ n	
Competitivo	a = 1 neurona con max n a = 0 Todas las otras neuronas	

Adaptado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

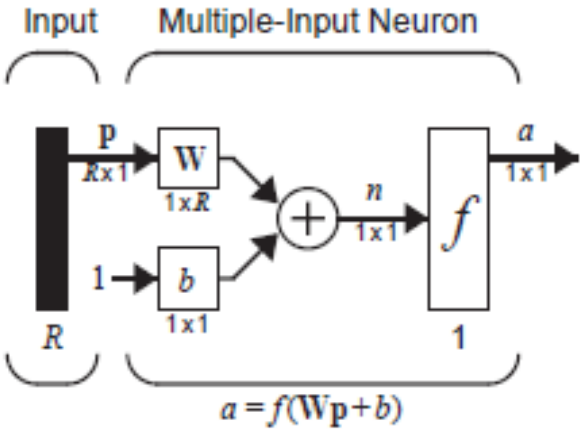
[66] Ibíd.

$$n = w_{1.1}p_1 + w_{1.2}p_2 + \cdots + w_{1.R}p_R + b$$
$$n = Wp + b$$
$$a = f(Wp + b)$$

(Ecuación 8)

La figura 28 presenta el esquema de una neurona con un vector de entrada R x 1 de datos Input llamado P. Este vector de datos de entrada atraviesa la matriz de pesos específicos W la cual tiene R columnas, pero solo una fila, posteriormente se multiplicará escalarmente el factor *Bias,* el primer resultado que se obtiene representado con la letra *n* es la suma del bias por la multiplicación Wp. Por último, el resultado *n* es transformado a un valor escalar *a* gracias a la función de activación[67].

Figura 28. Neurona con múltiples entradas.



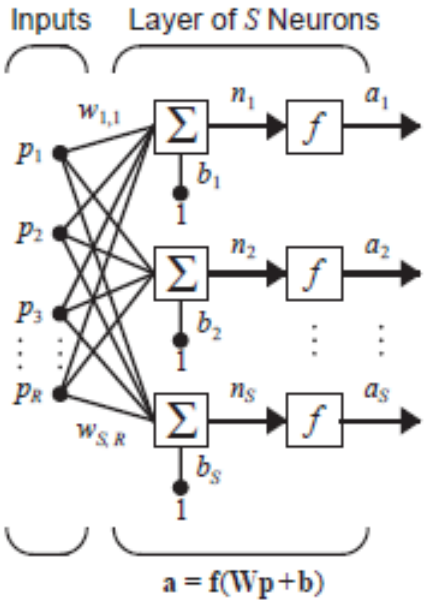
Tomado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

1.7.1.3. Capa de neuronas: Comúnmente una sola neurona, aunque posea múltiples entradas el manejo de esta no sería suficiente para obtener un resultado específico y probablemente se necesiten cuatro o cinco neuronas más que se encuentren conectadas entre sí en paralelo. A esto se le llama Capa. La figura 29 muestra un esquema general de una capa de neuronas. Como se ve en la imagen todos los datos de entrada Input están conectados con cada una

[67] Ibíd.

de las neuronas y debido a eso la matriz de pesos específicos, W, tiene ahora múltiples filas. Las coordenadas en cada uno de los pesos específicos representan la variable Input de procedencia y la neurona a la que se dirige de esta forma: El dato w3,2 indica el peso específico conectado del segundo dato de entrada hacia la tercera neurona[68].

Figura 29. Capa de neuronas.



Tomado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

Como se puede apreciar en la imagen, cada una de las neuronas posee su respectivo valor Bias, su valor Output a y su función de transferencia, la cual puede ser diferente para cada neurona.

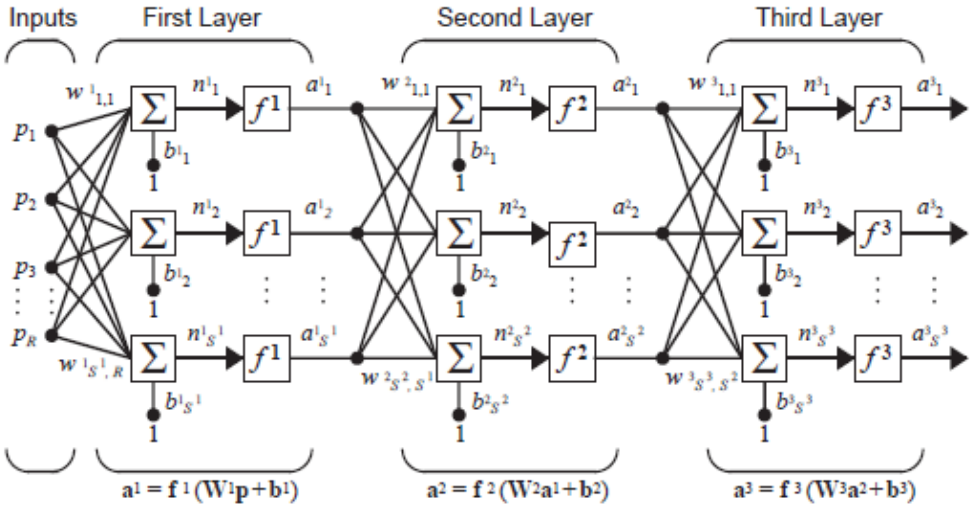
1.7.1.4. Múltiples capas de neuronas: En algunos casos las redes neuronales se construyen de formas más complejas debido a la cantidad y calidad de los datos que deben manejar, de hecho, en la realidad la mayoría de las redes neuronales artificiales están compuestas por múltiples capas de neuronas. Como se puede observar en la figura 30, cada capa de neuronas

[68] Ibíd.

posee su propia matriz de pesos específicos y su propio vector Bias, además, los datos de salida finales de la primera capa serán los datos de entrada de la segunda y así sucesivamente hasta la última capa. En el diagrama esquema de la red neuronal se usan superíndices para diferenciar una capa de otra[69].

A parte de la primera y última capa, las demás que se encuentran en medio se conocen como capas ocultas. Si bien el uso de estas capas ocultas es importante para mejorar la representatividad de los resultados, no se tiene un numero especifico u óptimo de capas ocultas o de neuronas dentro de estas, aunque la literatura recomienda mínimo usar dos o tres capas. A diferencia de esto, la última capa de neuronas está condicionada al número de resultados que se necesiten[70].

Figura 30. Red Neuronal con múltiples capas de neuronas.



Tomado de: HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

De acuerdo con su estructura y funciones de transferencia las redes neuronales de múltiples capas se dividen en varios tipos, los más representativos se nombran a continuación[71]:

[69] Ibíd.

[70] Ibíd.

[71] Ibíd.

- **Perceptrón:** El perceptrón es la red neuronal más básica, la cual consta de una sola neurona con una matriz Input, una de peso específico y un vector Bias.

- **Feed-Forward Network:** Este tipo de red es la más antigua dentro de los tipos de redes más complejas. Entre sus características principales está que todas las neuronas se encuentran conectadas entre sí, solo poseen una capa oculta de neuronas y el método de entrenamiento de este tipo de red artificial suele ser el algoritmo de Backpropagation. Su funcionamiento es el de clasificar en “Si o No” una serie de datos mediante el uso del código binario.

- **Red Neuronal recurrente:** A diferencia de la feed-forward, en donde los cálculos realizados eran de tipo linear, en este tipo de red las conexiones entre neuronas manejan una trayectoria circular lo cual genera una dependencia del resultado final, no solo de los valores Input iniciales, sino que además de los resultados internos de las capas ocultas. Este tipo de arreglos informáticos son utilizados en programas de reconocimiento de escritura a mano o de voz.

- **Red neuronal tipo Hamming:** Esta red está diseñada para resolver problemas de selección basados en el código binario con el detalle de que su estructura se encuentra conformada por una mitad del tipo Feed-Forward y la otra mitad del tipo Recurrent Network.

  En la capa Feed-Forward las neuronas trabajan bajo la función de transferencia linear generando la multiplicación de los vectores Input con la matriz de pesos específicos. Dependiendo de la cantidad de datos en los diferentes vectores de entrada, cada neurona de esta capa arrojará como resultado un vector Output interno, la red está programada para seleccionar el mayor de estos vectores Output

internos y convertirlo en el vector prototipo que alimentará la capa Recurrent Network.

Posteriormente, la capa Recurrent Nerwork, que está programada con la función de transferencia Competitive, las cuales compiten entre sí mismas hasta que en una de ellas la función de transformación de un resultado diferente a cero.

- **Red neuronal tipo Hopefield:** Este tipo de red funciona de manera similar a la capa Recurrent Network de las del tipo Hamming, con la diferencia de que su única capa de neuronas, por la manera en que están relacionadas entre sí, cumplen el trabajo de las dos capas usadas en esta. Estas neuronas comienzan con los datos de entrada del vector Input, para posteriormente, usando las funciones de transferencia de tipo lineal saturado simétrico, realizar iteraciones hasta que el resultado corresponda exactamente con el vector prototipo, es decir, que la principal diferencia entre este tipo de redes y las de tipo Hamming radica en la forma como es construido el vector prototipo.

1.7.2. Principales métodos de entrenamiento

1.7.2.1. Métodos de Kernel: Los métodos de entrenamiento de Kernel presentan buen rendimiento general, permitiendo alcanzar resultados bastante certeros, ya que pueden ser fácilmente ajustados mediante el uso de la teoría de aprendizaje estadístico o el uso de argumentos Bayesianos. Estos pueden ser utilizados para la automatización de regresiones, clasificaciones, aprendizaje no supervisado o detección de novedades, entre otros. Debido a que el objetivo del trabajo es optimizar variables, se enfocará principalmente en los métodos de regresión, entre los que encontramos[72]:

[72] HOWLETT, Robert; JAIN, Lakhmi. Radial basis function networks 1. 2001

- **Regresión tipo Ridge:** Se basa en el método de mínimos cuadrados, donde el objetivo es minimizar el vector w presentado en la siguiente ecuación (KALNISHKAN, 2009):

$$\boldsymbol{L_{RR}(W)} = \mathrm{A} * \left|\left|w\right|\right|^2 + \sum_{i=1}^{T} (w' * x_i - y_i)^2 \qquad \textbf{(Ecuación 9)}$$

  Según autores como Campbell, este tipo de regresión es muy útil al trabajar con funciones de base radial, y con máquinas soportadas por vectores o SVMs, por sus siglas en inglés, ya que permite generar proyecciones no lineales de una gran cantidad de datos, para clasificarlas de manera binaria, logrando reconocer, de forma más sencilla, los datos útiles.

- **Regresión no lineal**

1.7.2.2. Método de retropropagación o Backpropagation: Este es un método numérico que se centra en la búsqueda del mínimo en la función error en el espacio de pesos w, mediante el uso de gradientes descendentes, de tal manera que la combinación de pesos que minimice la función error es considerada una solución para el problema de aprendizaje. Ya que debe tener la capacidad de iterar el gradiente de la función error en cada paso, es necesario asegurar que la función base sea continua y diferenciable de la función error, para ello es posible utilizar cualquier función de transferencia que cumpla ambas condiciones, aunque, según diversos autores, la más recomendable es la función sigmoid, en la cual la constante c puede ser seleccionada arbitrariamente y cuyo reciproco 1/c es conocido como parámetro de temperatura para redes neuronales estocásticas[73].

También es posible el uso de redes neuronales de rectificador profundo mediante este método, de tal manera que, en lugar de utilizar una función sigmoid, se

[73] ROJAS, Raúl. Neuronal networks: A systematic introduction. Marzo de 1996.

aplica una función de rectificación para todas las neuronas ocultas, cuyo rango esta entre 0 y x, max(o,x), y no entre 0 y 1, lo que evita que la red se sature, por lo que es recomendable al aplicar redes muy profundas[74].

Sin embargo, el principal problema que presenta este método de entrenamiento es que, bajo ciertas condiciones, se pueden llegar a presentar mínimos locales en la función error, que impidan encontrar el resultado real, razón por la cual se sugiere el uso de modificaciones en el método como puede ser la implementación de momentos, que modifican ligeramente la ecuación y disminuye la posibilidad de fallo[75].

1.7.2.3. Métodos pre-training: Este tipo de método se subdivide en dos clases principales, a saber[76]:

- **Deep Belief Network o DNB:** Es un tipo de algoritmo no supervisado, muy eficiente para determinar los pesos en una maquina Boltzmann restringida multicapa, RBMs por sus siglas en inglés, en las cuales, las neuronas deben formar gráficas bipartidas. En este tipo de redes, las conexiones deben realizarse entre una capa oculta y una capa visible y su entrenamiento puede ser complementado con el uso de un algoritmo de divergencia contrastiva de un solo paso.

- **Discriminativo o DPT:** Este es un método alternativo a la DNB, donde primero se entrena una capa oculta hasta su convergencia total mediante el apoyo del uso de backpropagation, esta capa pasara a conocerse como capa softmax y será reemplazada por otra capa oculta a entrenar, repitiendo el proceso sucesivamente hasta alcanzar el número de capas ocultas deseado, esta alteración permite reducir

[74] TÓTH, Lázló; GRÓSZ, Tamáz. A Comparison of Deep Neural Network Training Methods for Large Vocabulary Speech Recognition. 2013.
[75] ROJAS. Op. Cit.
[76] TÓTH. Op Cit.

la cantidad de iteraciones de una DBN, lo que representa mayor rapidez y con ella mayor efectividad de la red.

2. PLANTEAMIENTO DE LA METODOLOGÍA BIG DATA ANLYTICS PARA LA OPTIMIZACIÓN DE PARÁMETROS DE PERFORACIÓN

Teniendo presentes los conceptos, softwares y metodologías estudiadas anteriormente, se procede a proponer una metodología nueva que permita aplicar el concepto de Big Data Analytics a la optimización de los parámetros de perforación más representativos, como el peso sobre la broca o las revoluciones por minuto de esta, entre otras, de acuerdo con la operación, tal como se menciona al final del capítulo 1.6. Para esto lo primero consiste en filtrar los datos y normalizarlos mediante métodos estadísticos, con el fin de eliminar datos vacíos o aislados, los cuales pueden generar un gran error al calcular la Energía Mecánica Especifica o MSE por sus siglas en inglés, la cual será utilizada como parámetro base en la optimización, para lo que se aplica la ecuación generalizada de la Energía Mecánica Especifica, mostrada a continuación:

$$MSE = \frac{WOB}{A} + \frac{120 * T * N * \pi}{A * ROP}$$ [77] **(Ecuación 10)**

Donde:

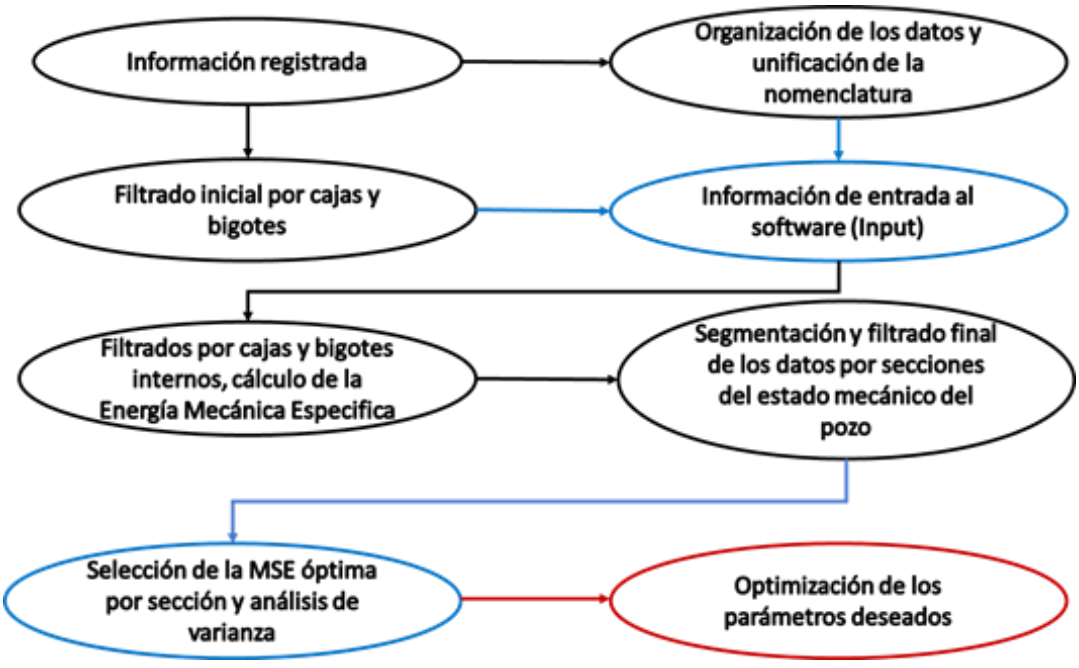
- WOB: Peso sobre la broca [lb]
- A: Área del hueco [in^2]
- T: Torque [ft-lb]
- N: Revoluciones por minuto de la broca [rpm]
- ROP: Velocidad de penetración [ft/hr]

[77] LOGAN, W. D. Engineered Shale Completions Based on Common Drilling Data. SPE. 2015

El paso siguiente es realizar un segundo filtrado de datos sobre la MSE calculada en cada sección del estado mecánico, mediante el método de diagramas de cajas y bigotes, para posteriormente elegir el valor óptimo, no siendo siempre el menor, sino aquel que tenga la mayor frecuencia o moda dentro del intervalo tomado, debido a que alcanzar el valor menor de manera inmediata en toda la zona puede ser demasiado ambicioso o complicado de obtener en la práctica.

Posteriormente se pasará a tomar las variables disponibles en la base de datos de la operación, tanto operacionales como geológicas, reológicas o administrativas, estas últimas relacionadas con las compañías de servicios involucradas en el desarrollo de cada pozo. Una vez tomadas todas las variables disponibles se procederá a determinar su relación o impacto sobre la variable base mediante diferentes métodos como la aplicación de correlaciones, redes neuronales artificiales o cross-plots, para poder establecer, así, las variables que más la afectan, las cuales serán tomadas como variables críticas a ser optimizadas. el diagrama de flujo de la metodología general se presenta en la figura 31.

Figura 31. Diagrama de flujo de la metodología



Finalmente se realizarán los diferentes escenarios o posibilidades de optimización del conjunto de variables críticas obtenido, de tal manera que se obtenga el valor de MSE deseado en todo el intervalo, lo que asegurará una reducción en el tiempo total de perforación y con este una disminución en los costos de la operación.

Para poder comprender de una mejor manera la metodología propuesta, en el capítulo 3, presentado a continuación, se explicará de manera más detallada el paso a paso presentado en el diagrama general, a la vez que es programada y validada mediante el ingreso de datos reales de diferentes pozos de un campo colombiano de crudo pesado ubicado en los llanos orientales, el cual no puede ser especificado por motivos de confidencialidad.

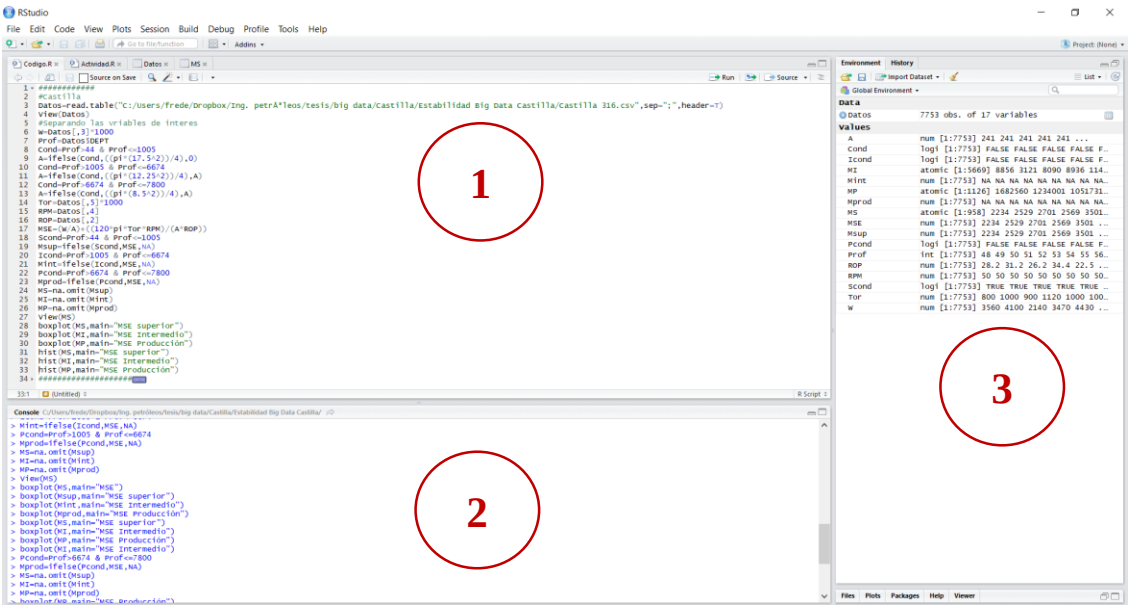
3. PROGRAMACIÓN DE LA METODOLOGÍA

3.1. R Studio

Debido al tratamiento estadístico que debe realizarse a lo largo del proceso, la primera opción para realizar una programación de la metodología es R Studio, ya que, al ser de código abierto, cuenta con una amplia gama de librerías, lo que permite realizar una programación muy versátil, además de contar con opciones de visualización y gráficos personalizables, lo que facilita la interpretación de los datos.

De esta manera se procede a realizar la programación de un pozo, llegando hasta la etapa de filtrado y selección de la MSE óptima, para validar el código con datos reales fundamentados previamente y para verificar la eficiencia de visualización de los resultados obtenidos con el programa. La mesa de trabajo del programa es ilustrada en la figura 32.

Figura 32. Mesa de trabajo de R Studio



En la figura 37 se puede observar una comparación de los diagramas de cajas y bigotes para un intervalo obtenidos con R Studio y con el análisis estadístico realizado y validado previamente con la ayuda de Excel.

Como se puede evidenciar en la figura 32, la mesa de trabajo del programa está conformada por 3 zonas principales, estas son:

1. **El editor de código:** Es la ventana principal de programación, donde se escribe y guarda el código, además de permitir la entrada de comentarios para facilitar la organización de las funciones realizadas en cada parte de la codificación, un ejemplo de la forma en que se presenta el código en esta ventana se presenta en la figura 33.

2. **La consola:** Esta es la ventana donde se ejecutan los comandos, mostrando los resultados del código, tal como se muestra en la figura 34.

3. **Multipropósito:** En esta ventana se pueden visualizar las variables, constantes, matrices, data frames, vectores definidos a lo largo del código,

de esta manera se puede saber el tipo de variable y las características que están siendo manejadas a lo largo del código, como se puede observar en la figura 35.

4. **Multifunción:** Ventana complementaria que puede ser desplegada debajo del multipropósito y en la cual se puede encontrar el menú de ayuda del programa.

Figura 33. Editor de código de R Studio

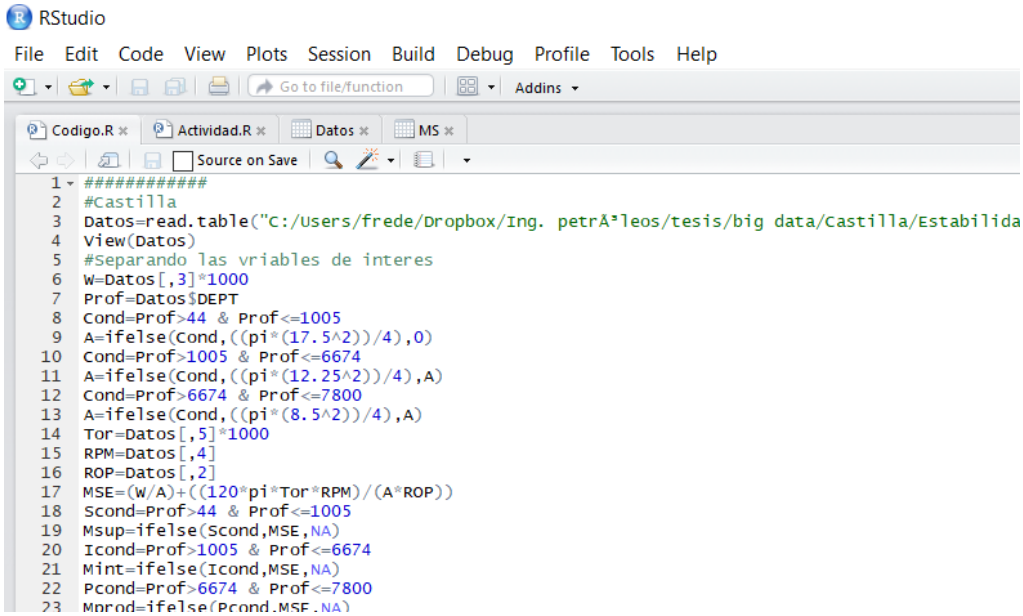


Figura 34. Consola de R Studio

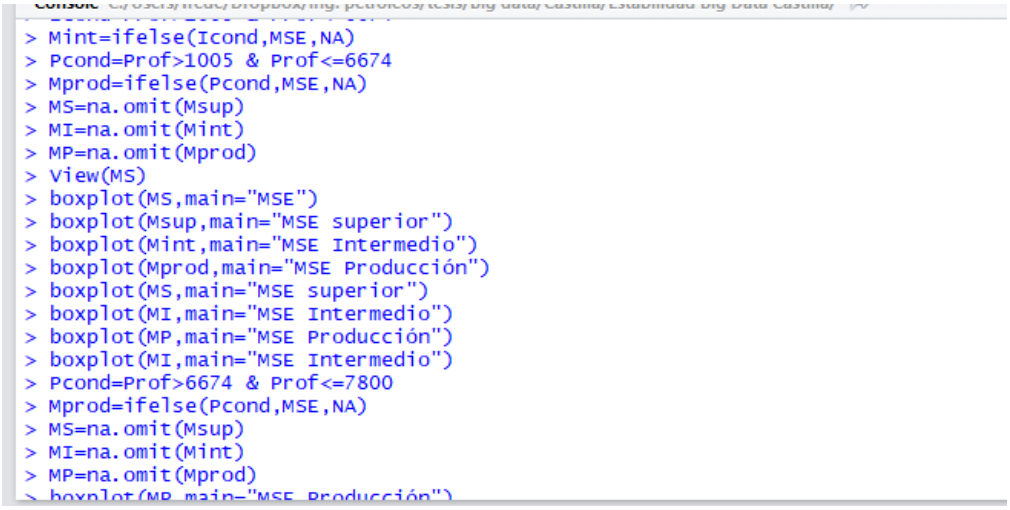
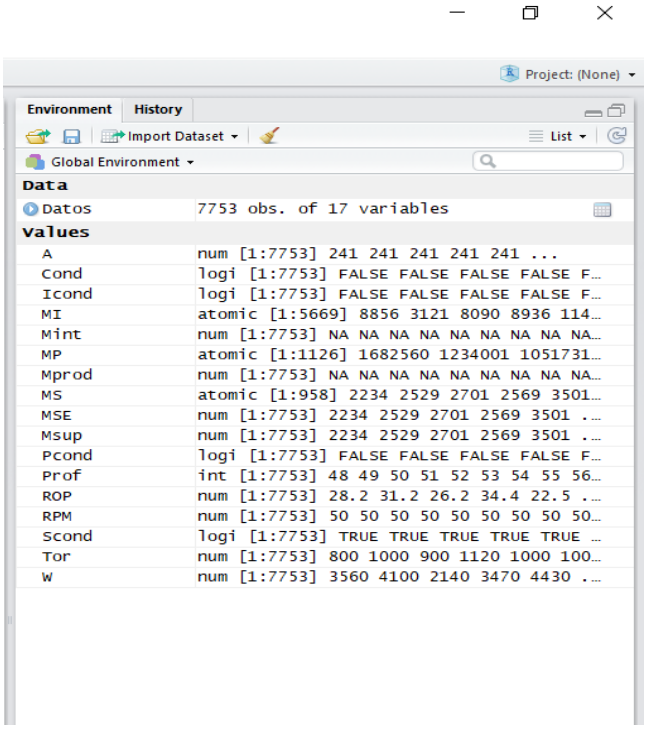


Figura 35. Multipropósito de R Studio



Conociendo el entorno o mesa de trabajo del programa podemos pasar a codificar el algoritmo o flujo de trabajo planteado inicialmente, el cual se divide en las siguientes partes:

1. **Definición de la base de datos:** Lo primero es llamar los datos de entrada con los que vamos a trabajar, es importante recordar, que estos datos deben ser previamente filtrados para evitar errores debido a datos aislados o faltantes, para nuestro caso particular los datos serán guardados y tratados como data frames por el programa, teniendo presente que un data frame es una matriz de doble variable, y funcionan al igual que una hoja de calcula de Excel, como se puede evidenciar en la figura 36.

2. **Definición de las variables, constantes y ecuaciones a utilizar:** Es fundamental tener claridad de las diferentes variables que serán manejadas a lo largo del código, otorgándoles nombres representativos para evitar confusiones y errores durante la codificación.

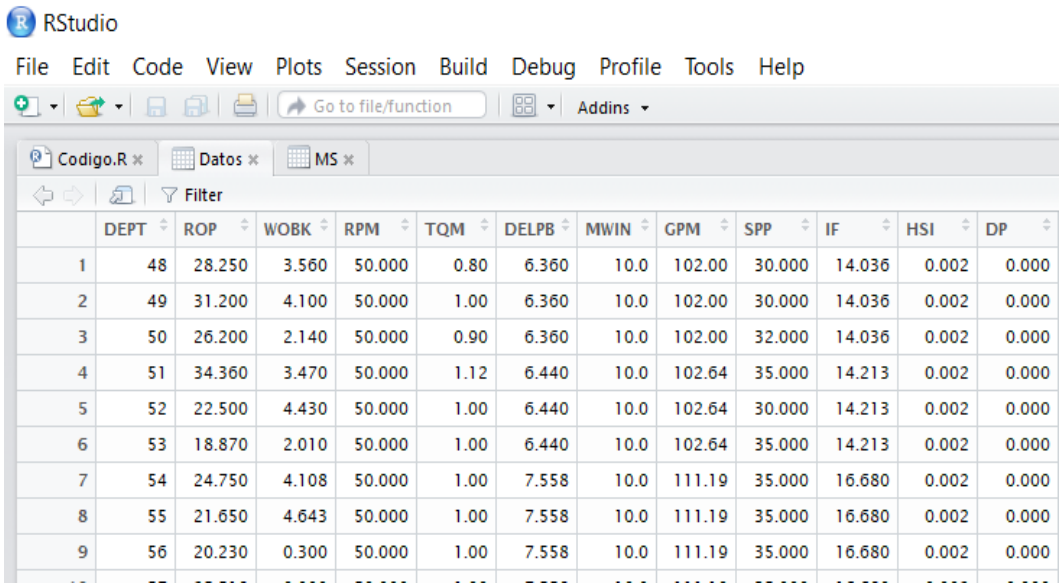
3. **Calculo de la MSE para los diferentes intervalos.**

4. **Filtrado de la MSE calculada para cada profundidad.**

5. **Selección de la MSE óptima para cada intervalo de completamiento.**

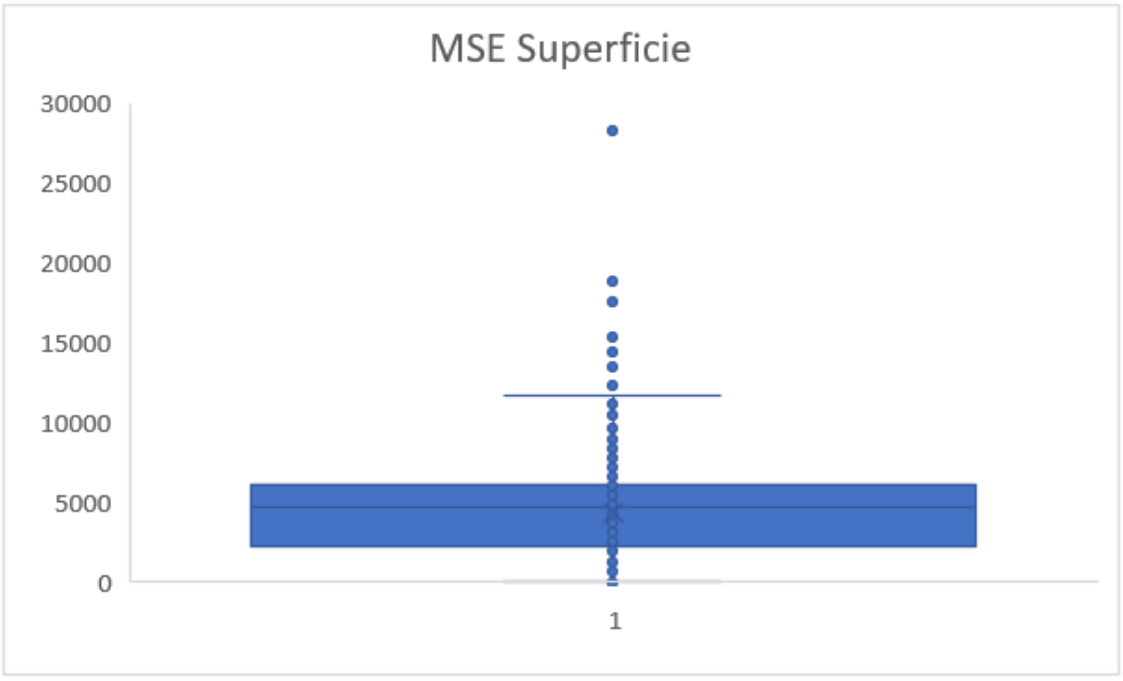
Como se puede evidenciar en la figura 37, los diagramas resultantes coinciden, con lo que se comprueba la validez del código realizado hasta la etapa de filtrado y selección de la MSE óptima.

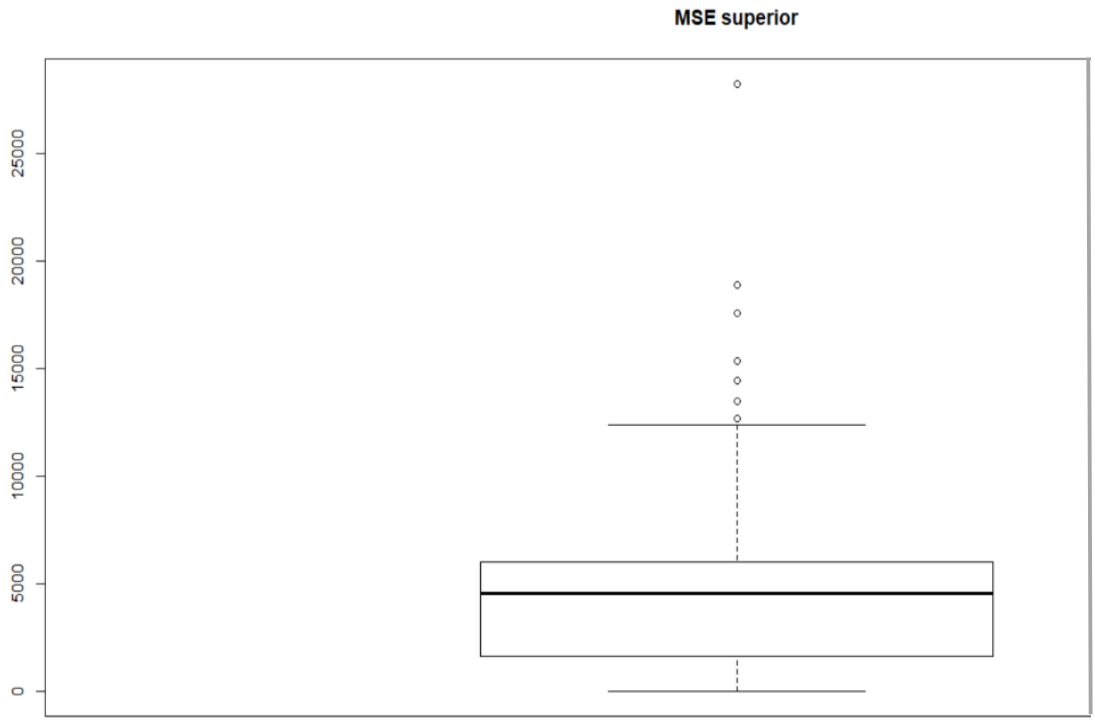
Figura 36. Data frame de entrada



	DEPT	ROP	WOBK	RPM	TQM	DELPB	MWIN	GPM	SPP	IF	HSI	DP
1	48	28.250	3.560	50.000	0.80	6.360	10.0	102.00	30.000	14.036	0.002	0.000
2	49	31.200	4.100	50.000	1.00	6.360	10.0	102.00	30.000	14.036	0.002	0.000
3	50	26.200	2.140	50.000	0.90	6.360	10.0	102.00	32.000	14.036	0.002	0.000
4	51	34.360	3.470	50.000	1.12	6.440	10.0	102.64	35.000	14.213	0.002	0.000
5	52	22.500	4.430	50.000	1.00	6.440	10.0	102.64	30.000	14.213	0.002	0.000
6	53	18.870	2.010	50.000	1.00	6.440	10.0	102.64	35.000	14.213	0.002	0.000
7	54	24.750	4.108	50.000	1.00	7.558	10.0	111.19	35.000	16.680	0.002	0.000
8	55	21.650	4.643	50.000	1.00	7.558	10.0	111.19	35.000	16.680	0.002	0.000
9	56	20.230	0.300	50.000	1.00	7.558	10.0	111.19	35.000	16.680	0.002	0.000

Figura 37. Comparación de los diagramas de cajas y bigotes para un intervalo





3.2. Python

Siguiendo con la programación de la metodología es importante notar que, a pesar de que R Studio es una gran herramienta para el análisis estadístico de datos, no ofrece la posibilidad de generar una interfaz amigable con el usuario de una manera sencilla, lo que es fundamental al ser necesaria la entrada inicial de los datos previamente filtrados en el software, por lo tanto se presenta Python como una herramienta más completa, que puede cumplir de una manera más eficaz con este objetivo, sin embargo, R Studio sigue siendo una gran opción para confirmar la veracidad de los resultados obtenidos a mayor profundidad, es decir, como herramienta complementaría de validación y análisis.

Inicialmente, antes de presentar la programación de la metodología, es necesario entender las bases generales del ambiente de trabajo de Python, para lo cual se debe entender que, al ser una herramienta de programación más robusta que R Studio, presenta diferentes formas o ambientes de trabajo de acuerdo con el objetivo del trabajo y a las preferencias del programador. Así, algunos de los más utilizados son presentados a continuación:

- **Ambiente interno de Python:** La primera opción de trabajo es el uso del ambiente interno de Python, el cual cuenta con tres componentes, presentados en las figuras 38, 39, 40 y 41, que son instalados por defecto con la aplicación, sin embargo, es importante denotar que esta no es una opción muy utilizada debido a que existen plataformas más amigables para programadores no muy familiarizados con la plataforma y el lenguaje de programación.

  La plataforma principal de este ambiente es el prompt o consola presentado en la figura 40, sin embargo, para realizar una programación más dinámica y organizada es recomendado trabajar sobre el Shell presentado en la figura 39, cuyo funcionamiento es el de un block de notas donde se realiza el código para luego correrlo directamente en el prompt, por último, el módulo presentado en la figura 41 es el complemento más

útil, ya que es necesario independientemente del ambiente en el que se vaya a trabajar, debido a que es el directorio donde se guardan y modifican las librerías que quieran ser utilizadas en el programa.

Figura 38. Ambiente interno de Python 3.6.

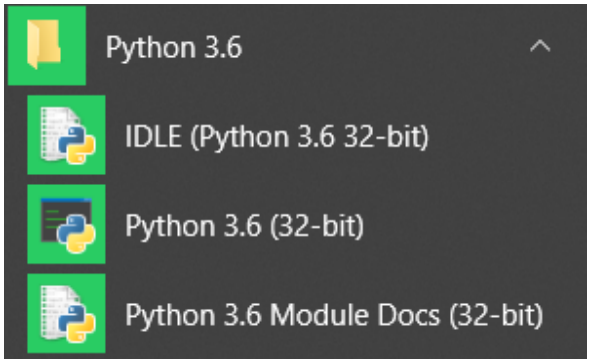


Figura 39. IDLE o Python Shell.

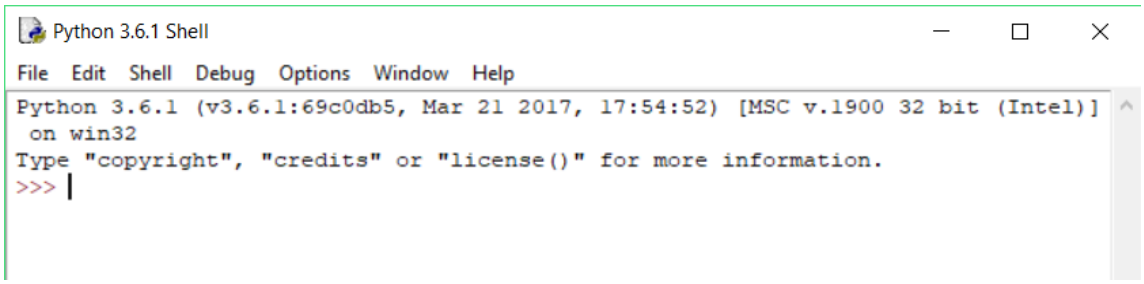


Figura 40. Python prompt.

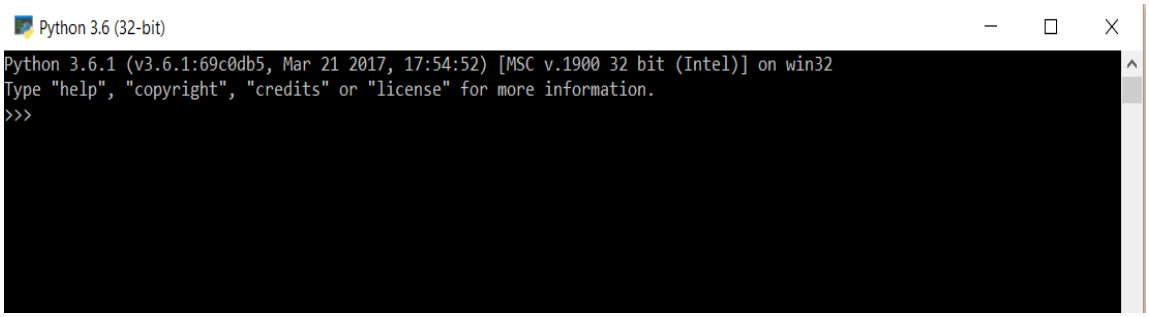
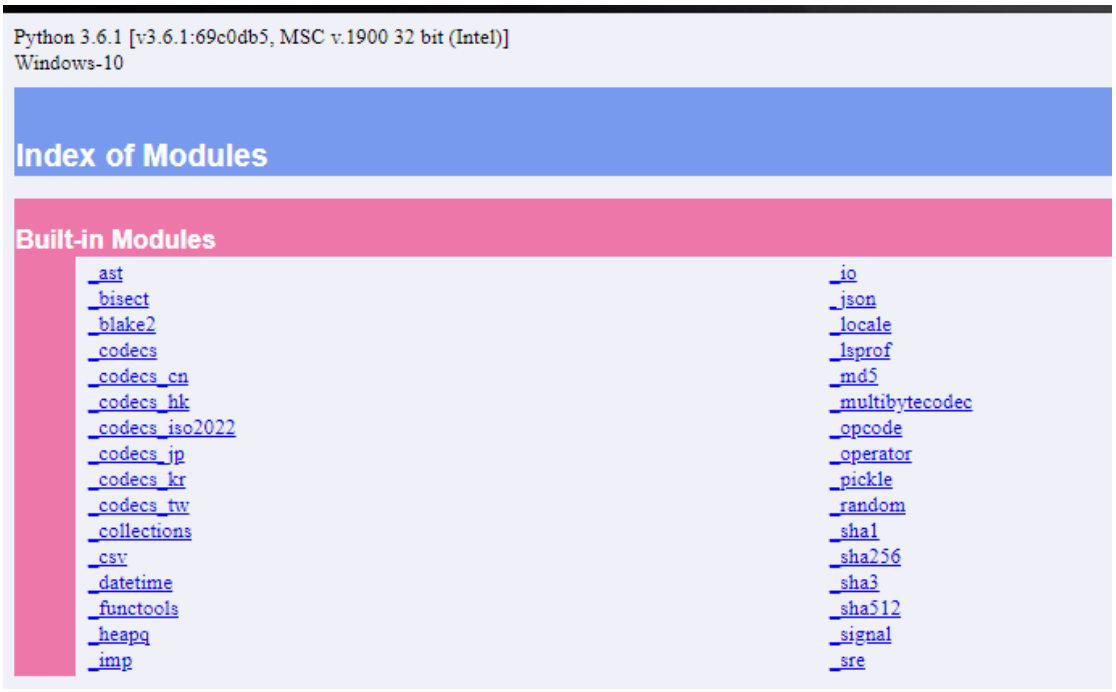


Figura 41. Python Module docs.



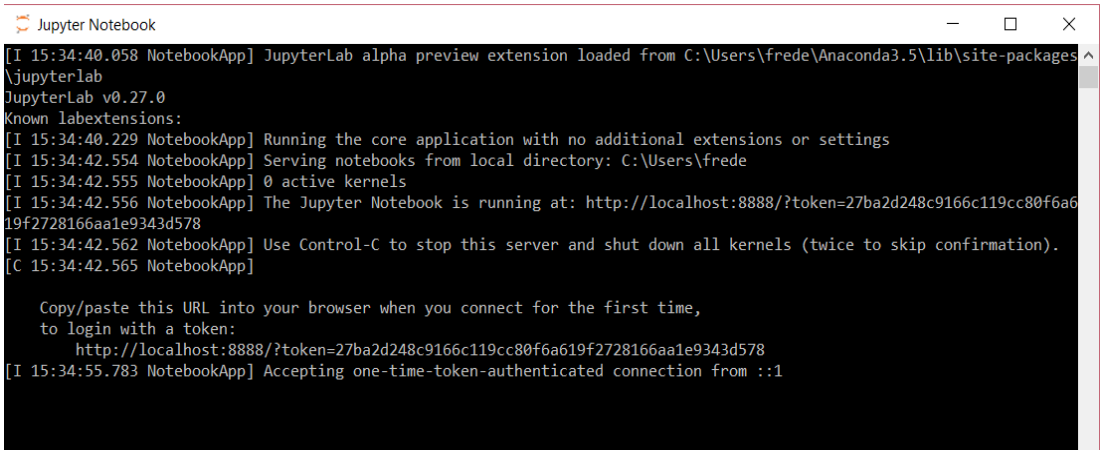
- **Anaconda:** Es una plataforma complementaria para trabajo en Python, muy utilizada debido a sus ambientes intuitivos, entre los que encontramos:

    - **Anaconda prompt:** Es una consola de trabajo directo parecida a la del ambiente interno de Python o al cmd del sistema de Windows.
    - **Jupyter:** Uno de los ambientes más utilizados de Python, ya que puede ser utilizado como consola directa, al igual que la anaconda prompt, como también es posible utilizarlo como compilador directo online, tal como se muestra en las figuras 42 y 43, sin embargo, es importante señalar que, si se quiere hacer uso de la herramienta

online, la consola debe mantenerse abierta, ya que esta es la que permite la conexión entre el equipo y el compilador.

- **Spyder:** Uno de los ambientes más amigables con el programador, ya que tiene múltiples ventanas desde donde se pueden controlar las variables o ejecutar y probar los códigos, de forma similar a la mesa de trabajo de R Studio o de Matlab, razón por la que es seleccionado como el ambiente a utilizar para este caso de estudio, los componentes de la mesa de trabajo de spyder son mostrados en la figura 44.

Figura 42. Consola directa de Jupyter.



En la figura 44 se puede observar que, al igual que R Studio, Spyder cuenta con las mismas 3 zonas principales, el editor de código, con el número uno, el multipropósito con el 2 y la consola en la parte inferior con el 3, los cuales cumplen los mismos propósitos, es importante, sin embargo, señalar que para compilar el código para su uso general se den utilizar herramientas complementarias como cx_freeze o pyinstaller.

Figura 43. Compilador online de Jupyter.

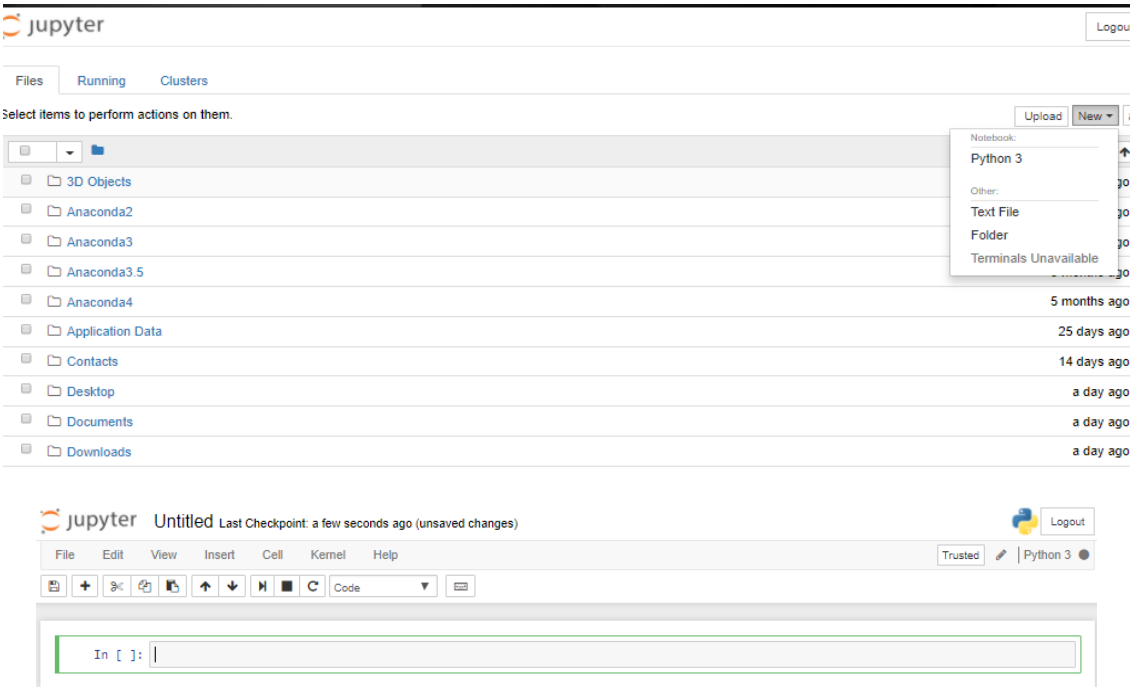
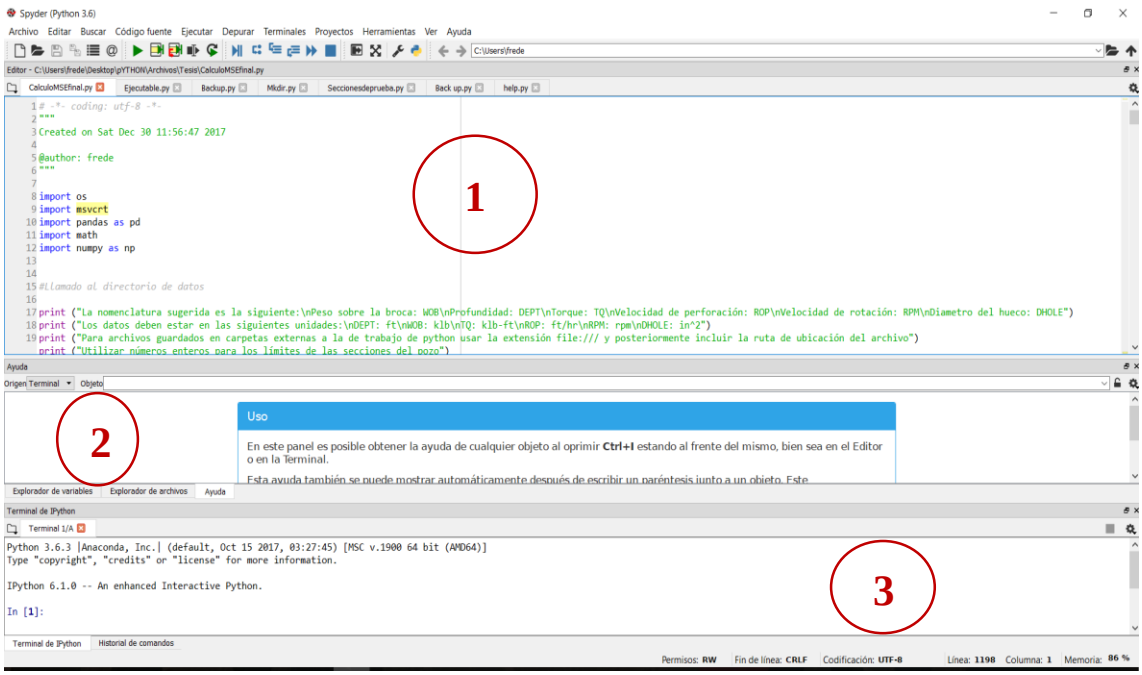


Figura 44. Mesa de trabajo de Spyder.

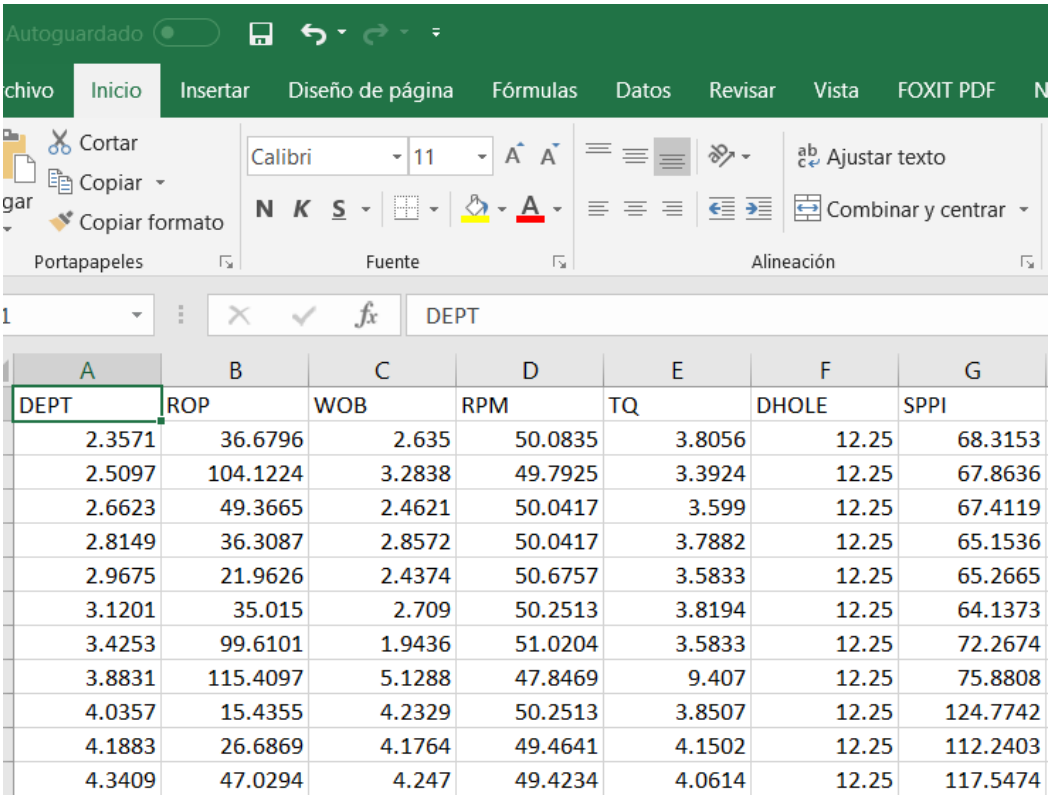


De esta manera, para dar un entendimiento más completo de la metodología planteada a ser programada, es posible dividirla en las siguientes etapas:

3.2.1. Filtrado de los archivos de entrada

La etapa inicial del proceso se centra en el filtrado de los archivos de entrada, ya que la calidad de los resultados dependerá de la confiabilidad que presenten los inputs del programa. En el software se especifican las unidades en las que se deben manejar los datos de la ecuación de MSE y que los archivos de entrada deben estar en formato .csv, también es importante que se maneje una nomenclatura única, sugerida igualmente por el programa. Esto se ilustra en la figura 46. Además, se deben eliminar todos los datos que sean negativos o iguales a cero ya que estos son considerados como fuentes potenciales de error y pueden ser producto de malas lecturas o registros.

Figura 45. Datos de entrada del pozo x para la validación de la metodología.



	A	B	C	D	E	F	G
	DEPT	ROP	WOB	RPM	TQ	DHOLE	SPPI
	2.3571	36.6796	2.635	50.0835	3.8056	12.25	68.3153
	2.5097	104.1224	3.2838	49.7925	3.3924	12.25	67.8636
	2.6623	49.3665	2.4621	50.0417	3.599	12.25	67.4119
	2.8149	36.3087	2.8572	50.0417	3.7882	12.25	65.1536
	2.9675	21.9626	2.4374	50.6757	3.5833	12.25	65.2665
	3.1201	35.015	2.709	50.2513	3.8194	12.25	64.1373
	3.4253	99.6101	1.9436	51.0204	3.5833	12.25	72.2674
	3.8831	115.4097	5.1288	47.8469	9.407	12.25	75.8808
	4.0357	15.4355	4.2329	50.2513	3.8507	12.25	124.7742
	4.1883	26.6869	4.1764	49.4641	4.1502	12.25	112.2403
	4.3409	47.0294	4.247	49.4234	4.0614	12.25	117.5474

En la figura 45 se presentan los datos de entrada, previamente tratados mediante la unificación de la nomenclatura y la conversión de los datos a las unidades

sugeridas en el software, que se presentan en la figura 46, de uno de los pozos que serán utilizados para la validación de la metodología, el cual por motivos de confidencialidad será nombrado pozo x.

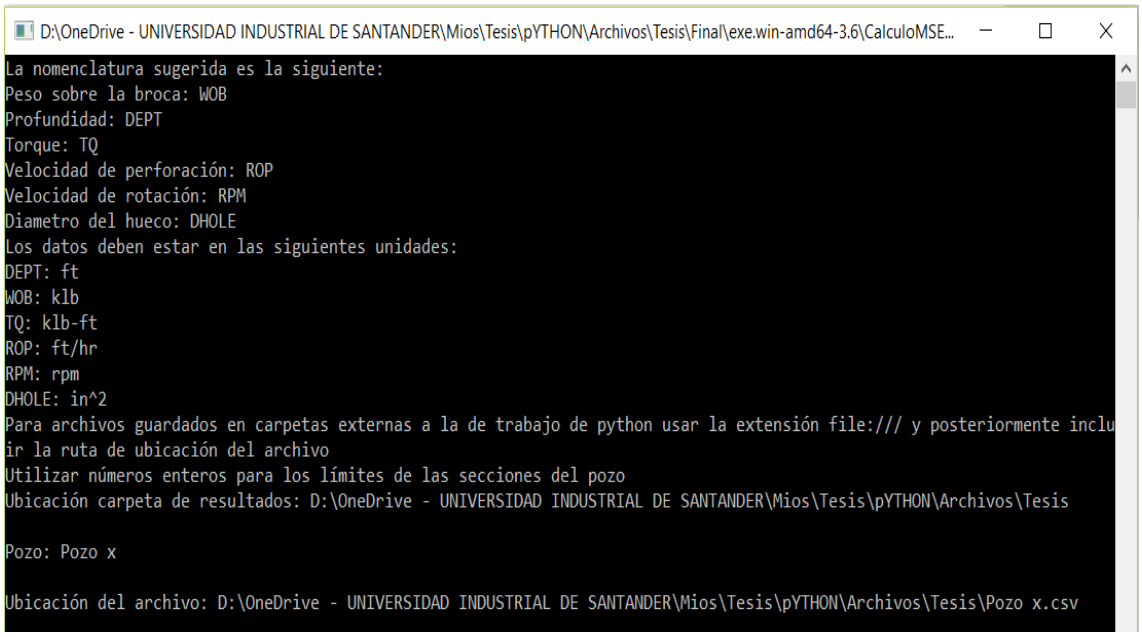
3.2.2. Calculo de MSE óptimos por intervalos

Posterior al filtrado inicial de los datos de entrada, se solicitará el nombre del pozo y la ubicación del archivo de entrada, junto con la ubicación en la que se quiere guardar la carpeta de resultados final, para, posteriormente, realizar un segundo filtrado de estos datos mediante el método de cajas y bigote o rangos intercuartílicos para cada variable, calculando finalmente el MSE en toda la profundidad del pozo, tal como se muestra en la figura 47, es importante tener en cuenta que estos cálculos fueron validados mediante la comparación con datos de pozos ya conocidos de cierto campo colombiano.

Lo siguiente es especificar el número y la profundidad de las secciones representativas en la que fue completada el pozo, es decir, sin tener en cuenta el casing conductor, ya que este alcanza, normalmente, profundidades muy someras. Con estos datos el software se encarga de calcular el MSE por secciones y de realizar un tercer filtrado sobre el resultado de estas, para asegurar la mayor confiabilidad en los resultados, ya que, con base en estas mediciones, se encuentra el dato óptimo de cada sección, el cual corresponde al resultado mínimo con la mayor moda, asegurando, no solo el menor dato de MSE, que por sí solo puede ser muy difícil de alcanzar en la práctica, sino que también el que es más fácil de alcanzar y mantener, esto se muestra en las figuras 48 y 49.

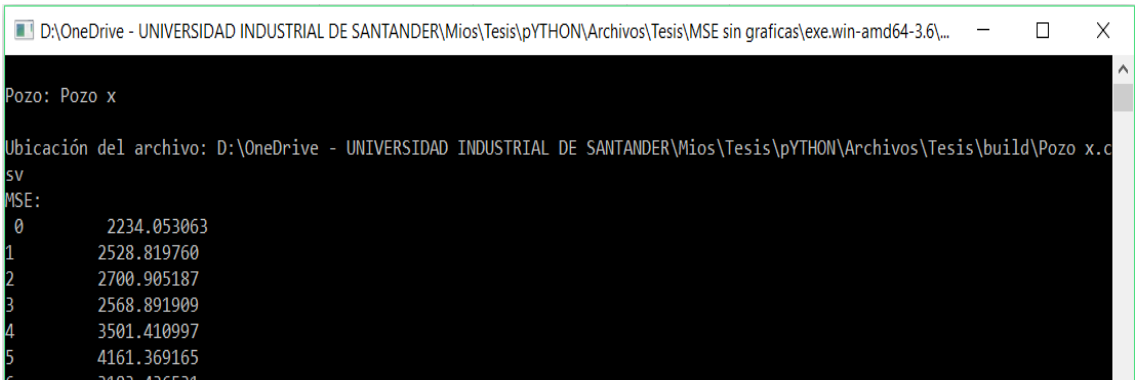
Cabe denotar que, como se aprecia en la figura 49, el software calcula y registra los datos estadísticos y de filtrado principales, en caso de que el usuario quiera revisarlos inmediatamente y da la opción final de mantener un trabajo continuo con varios pozos o salir de él.

Figura 46. Requerimientos para los archivos de entrada.



Cabe denotar que, como se aprecia en la figura 49, el software calcula y registra los datos estadísticos y de filtrado principales, en caso de que el usuario quiera revisarlos inmediatamente y da la opción final de mantener un trabajo continuo con varios pozos o salir de él.

Figura 47. MSE para la profundidad total del pozo.



Finalmente, el software genera la carpeta de resultados en el directorio o ubicación solicitada y, dentro de esta, guarda las carpetas de resultados individuales de cada pozo, donde se pueden encontrar archivos .csv con los datos de MSE para toda la profundidad, para cada intervalo y los objetivos

finales, para análisis posteriores que quiera realizar el usuario, tal y como se presenta en las figuras 50 y 51.

Figura 48. Determinación de intervalos y cálculo de MSE para el primero.

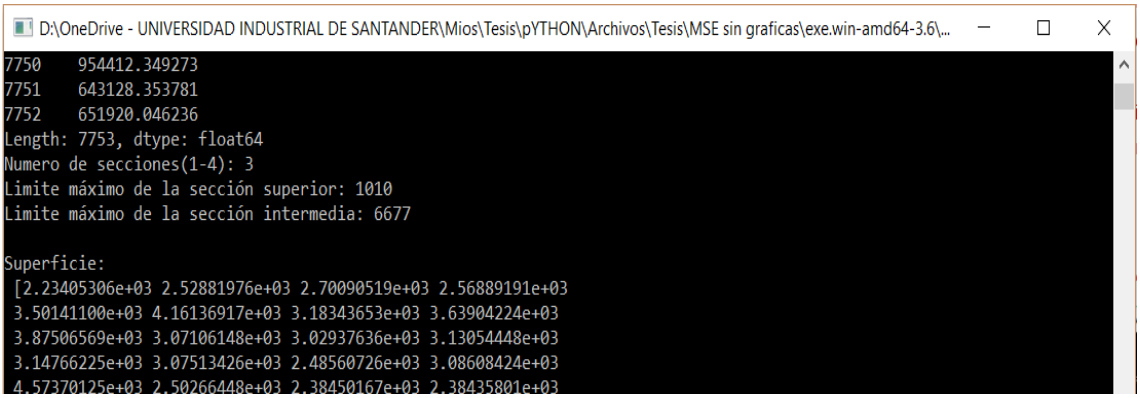


Figura 49. MSE óptimo u objetivo para cada intervalo.

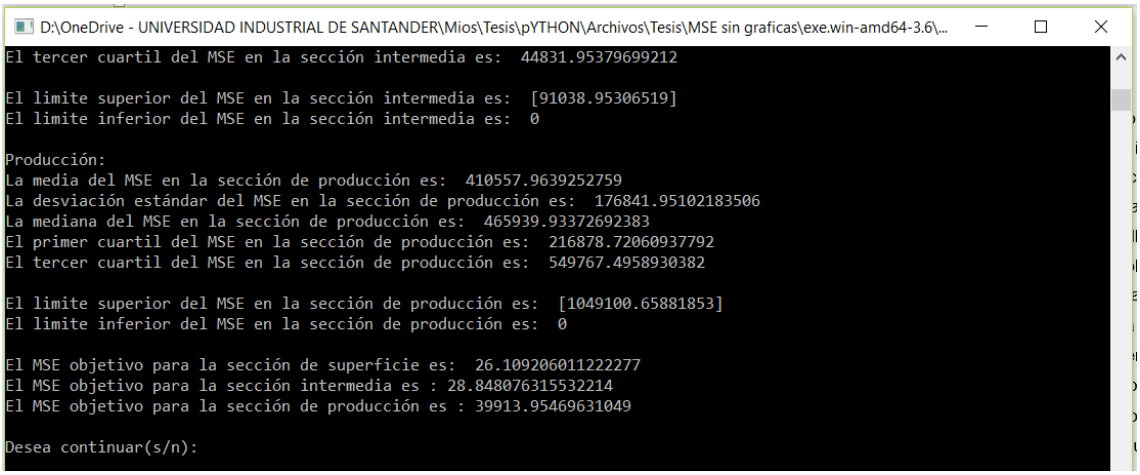
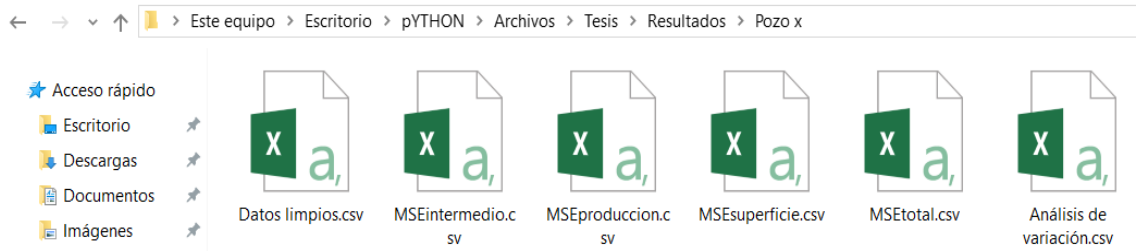


Figura 50. Carpeta de resultados.

Figura 51. Documentos finales de MSE.

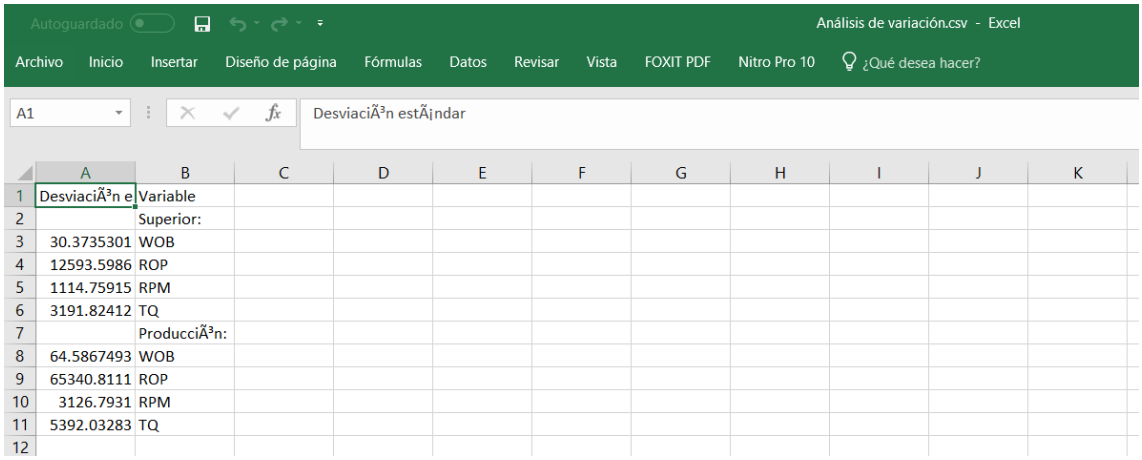


3.2.3. Optimización de las variables

La parte final del proceso es optimizar las variables de interés, las cuales, para este caso, son las incluidas en la ecuación de MSE, para lo que, inicialmente, es necesario determinar el impacto de cada una sobre el resultado, por lo tanto, se realiza un análisis de varianza, es decir, se varían individualmente los parámetros de la ecuación, suponiendo que el diámetro del hueco se toma como un dato definido, ya que este es determinado al hacerse el plan de perforación, y dejando 3 de las 4 variables restantes como constantes iguales al valor promedio de cada una en la lectura del pozo, procedimiento que se repite para cada una de las 4 variables.

Con base en el análisis realizado siguiendo este proceso, aplicándolo para varios pozos de un campo colombiano, se llega a la conclusión que los parámetros que más afectan a la MSE son la velocidad de penetración y el torque respectivamente, seguidos de las revoluciones de la broca y finalmente del peso sobre la misma, donde esta última exhibe una desviación significativamente menor a los otros parámetros, generando un error porcentual sobre el resultado menor al 1% en promedio, esto puede ser verificado por el usuario para cada pozo en estudio mediante la evaluación de los resultados de la varianza guardados por el software en la carpeta de resultados, tal como se presenta en la figura 51, es por este motivo que se desprecia la variación del peso sobre la broca en la optimización de los valores, un ejemplo de esto puede ser observado en la figura 52.

FIGURA 52. Análisis de variación del pozo x



	A	B	C	D	E	F	G	H	I	J	K
1	DesviaciÃ³n e	Variable									
2		Superior:									
3	30.3735301	WOB									
4	12593.5986	ROP									
5	1114.75915	RPM									
6	3191.82412	TQ									
7		ProducciÃ³n:									
8	64.5867493	WOB									
9	65340.8111	ROP									
10	3126.7931	RPM									
11	5392.03283	TQ									
12											

Así, el problema queda discretizado a un sistema de tres incógnitas o variables, para lo cual se pasa a determinar los posibles conjuntos de este trio de variables que den como resultado la MSE objetivo, lo que se logra mediante la aplicación de una red neuronal artificial programada en Python mediante cuatro complementos de machine learning y Deep learning previamente instalados, estos son:

- Theano
- TensorFlow
- Scikit-learn
- Keras

Para programar la red es necesario determinar la cantidad de capas ocultas que la conformarán, el número de neuronas presentes en cada capa, las funciones de activación, los métodos de optimización, la cantidad de iteraciones por capa y, en general, la topología de la red, además del porcentaje de datos para entrenamiento y prueba, posteriormente se evalúan los resultados obtenidos, buscando el mayor ajuste posible a la tendencia de estos.

De acuerdo con esto se decide utilizar 4 capas ocultas, con 16 neuronas para la primera, 10 para la segunda, 6 para la tercera y 3 para la cuarta, el método de entrenamiento utilizado es el de backpropagation, se utilizó la función de activación de Rectified Linear Units (Relu), Adam como función de optimización,

200 iteraciones por capa, con un batch_size de 4 valores por iteración, la función a minimizar es el error medio cuadrático, debido a que con estos atributos se obtiene el mayor ajuste para la tendencia real; parte del procedimiento de entrenamiento y ajuste realizado por la red, para un pozo, puede ser observado en la figura 53, un ejemplo de los resultados obtenidos en el pozo x es mostrado en la figura 54.

Finalmente, los resultados del software NPT, New Parameters Tool, son evaluados mediante el uso de crossplots en Excel, obteniendo una exactitud mayor al 80%, tal como se muestra en la figura 55.

Es importante denotar que los datos deben tener un orden aleatorio a la entrada para evitar que se genere una tendencia parcial no adecuada para todo el intervalo de perforación y que, para el caso planteado, se realiza el entrenamiento y ajuste con el 80% de los datos reales obtenidos mediante los cálculos internos del software y el 20% restante se usa para probar el resultado de la red y determinar si el entrenamiento se ajusta o no a la tendencia, además, debido a que las variables se correlacionan mediante una ecuación conocida, para reducir los grados de libertad en las entradas de la red y obtener un mejor resultado, se toman como variables de salida el torque y las revoluciones de la broca, y con ellas se calculan la rata de perforación, ya que esta última es la que presenta la mayor desviación, posteriormente, tras verificar un ajuste adecuado de la red a la tendencia, se realiza la predicción de los valores, ahora ajustados a la energía mecánica especifica objetivo, la cual es variada en más o menos un 5% del resultado obtenido previamente por el programa, para permitir un intervalo de error aceptable.

FIGURA 53. Proceso de entrenamiento y ajuste de la red neuronal.

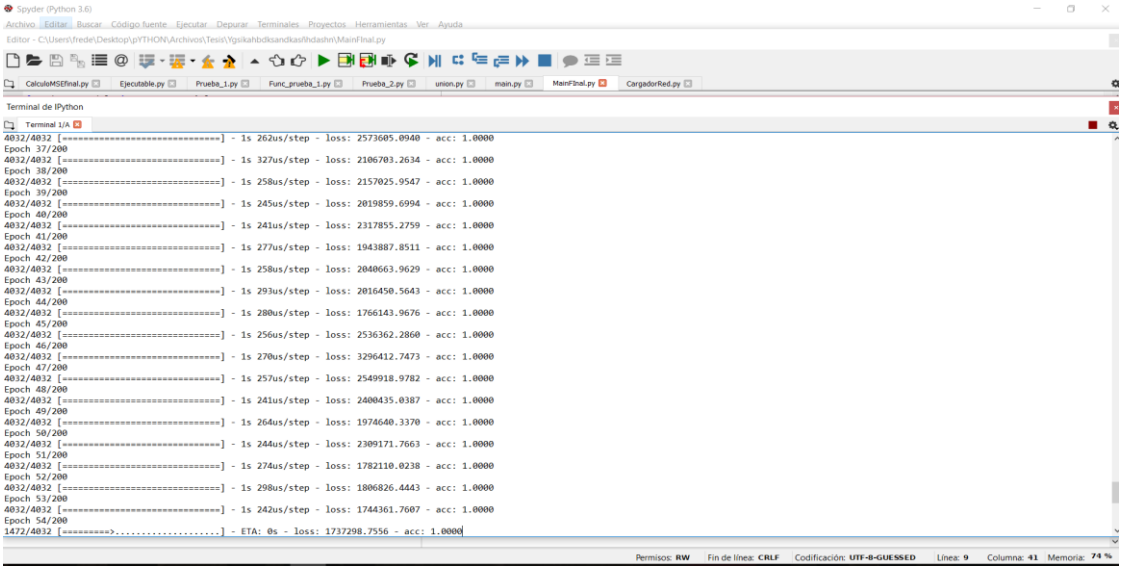
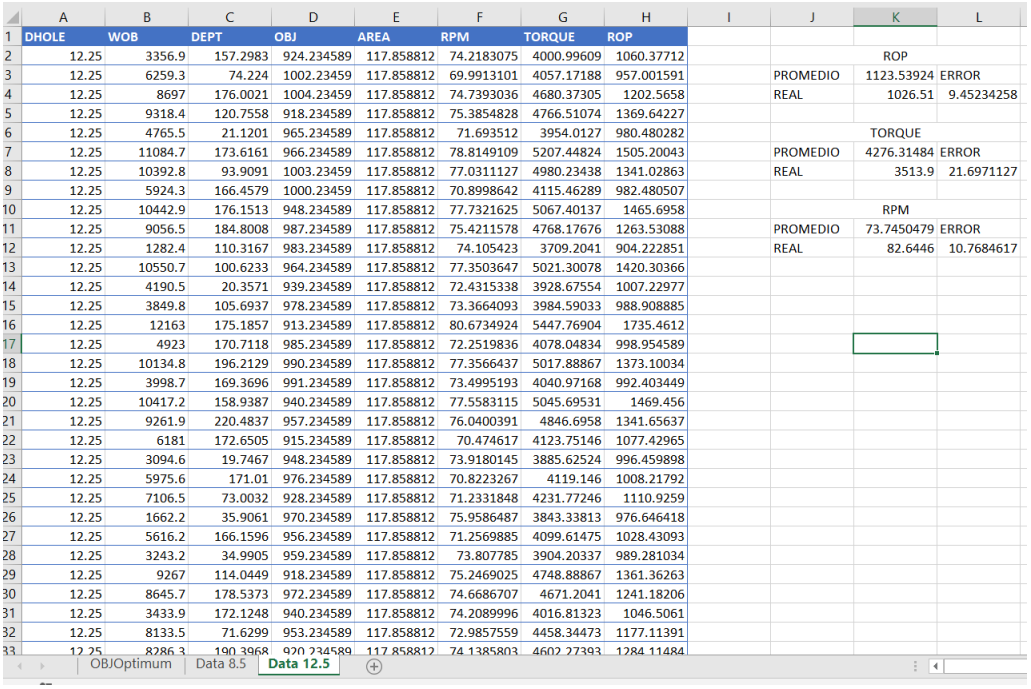


FIGURA 54. Resultados obtenidos para un pozo x de un campo colombiano de crudo pesado de los llanos orientales.

	A	B	C	D	E	F	G	H	I	J	K	L
1	DHOLE	WOB	DEPT	OBJ	AREA	RPM	TORQUE	ROP				
2	8.5	1275	811.7602	1893.25841	56.7450173	62.7603951	3443.21265	767.410932				
3	8.5	2431.2	1730.3346	1887.25841	56.7450173	64.4884262	4049.52905	940.6569			ROP	
4	8.5	1576.5	1535.0922	1893.25841	56.7450173	60.308712	3469.39722	745.15693		PROMEDIO	960.167156	ERROR
5	8.5	351.7	2450.2176	2037.25841	56.7450173	65.4520111	4397.23047	941.418543		REAL	984.8675	2.50798647
6	8.5	1946.6	755.9492	2054.25841	56.7450173	63.1869278	3517.28589	730.965366				
7	8.5	4855	2776.6736	1945.25841	56.7450173	68.0185318	4705.74072	1143.44607			TORQUE	
8	8.5	2110.7	1716.4878	1891.25841	56.7450173	62.6366768	3830.51147	859.734914		PROMEDIO	4150.52447	ERROR
9	8.5	1473.4	1527.0385	1906.25841	56.7450173	60.3250961	3484.65845	742.738721		REAL	3612.8	14.8838705
10	8.5	4931.3	2719.8398	2016.25841	56.7450173	68.0271149	4706.16602	1102.40303				
11	8.5	5324.3	2739.621	2039.25841	56.7450173	67.9206696	4653.97119	1079.47713			RPM	
12	8.5	2255.6	1830.1026	1886.25841	56.7450173	63.3713112	3948.552	900.290343		PROMEDIO	65.8308397	ERROR
13	8.5	2660.1	1173.6436	2068.25841	56.7450173	62.9027252	3697.12671	764.344522		REAL	80.429	18.150369
14	8.5	1607	1564.0319	1953.25841	56.7450173	60.7927704	3544.46094	743.684282				
15	8.5	1782.5	2324.7488	1944.25841	56.7450173	64.3215637	4125.16992	921.556199				
16	8.5	1248.3	860.0827	2072.25841	56.7450173	62.5544777	3499.75854	709.399588				
17	8.5	2011.4	626.6494	1995.25841	56.7450173	62.7063446	3362.56982	714.77893				
18	8.5	2648.7	2297.4791	1998.25841	56.7450173	66.5626984	4444.13965	1007.0137				
19	8.5	5148.8	2766.3255	1928.25841	56.7450173	67.5268631	4596.625	1122.2415				
20	8.5	5977.7	2916.0856	2017.25841	56.7450173	69.5407562	4767.8877	1152.12653				
21	8.5	2911.9	1283.5882	1916.25841	56.7450173	62.2565575	3604.23926	799.349236				
22	8.5	5505.9	2845.2674	2000.25841	56.7450173	68.5112305	4707.99609	1125.92655				
23	8.5	2106.8	883.9613	1926.25841	56.7450173	62.7720757	3474.60474	767.031548				
24	8.5	900.8	2498.6814	1999.25841	56.7450173	65.1608124	4359.13379	951.444136				
25	8.5	1904.9	2247.4612	2000.25841	56.7450173	64.3115234	4104.41016	891.676377				
26	8.5	1095.7	1924.1812	2050.25841	56.7450173	62.8821411	3959.99731	814.566476				
27	8.5	2606.7	562.9282	1912.25841	56.7450173	61.5150375	3135.33521	686.565913				
28	8.5	5064.9	2775.6846	2070.25841	56.7450173	68.5384216	4790.01367	1101.00551				
29	8.5	1931.3	1719.8788	2034.25841	56.7450173	62.7583351	3852.68652	803.082023				
30	8.5	1088.1	1917.9642	2038.25841	56.7450173	62.7892532	3946.01465	815.254475				
31	8.5	1396.8	2481.0197	1965.25841	56.7450173	64.9859161	4284.08008	953.091409				
32	8.5	2824.2	572.2603	1930.25841	56.7450173	61.2816734	3115.76514	674.571606				
33	8.5	2763.1	578.0929	2055.25841	56.7450173	61.7633591	3234.42749	661.421715				

OBJOptimum | Data 8.5 | Data 12

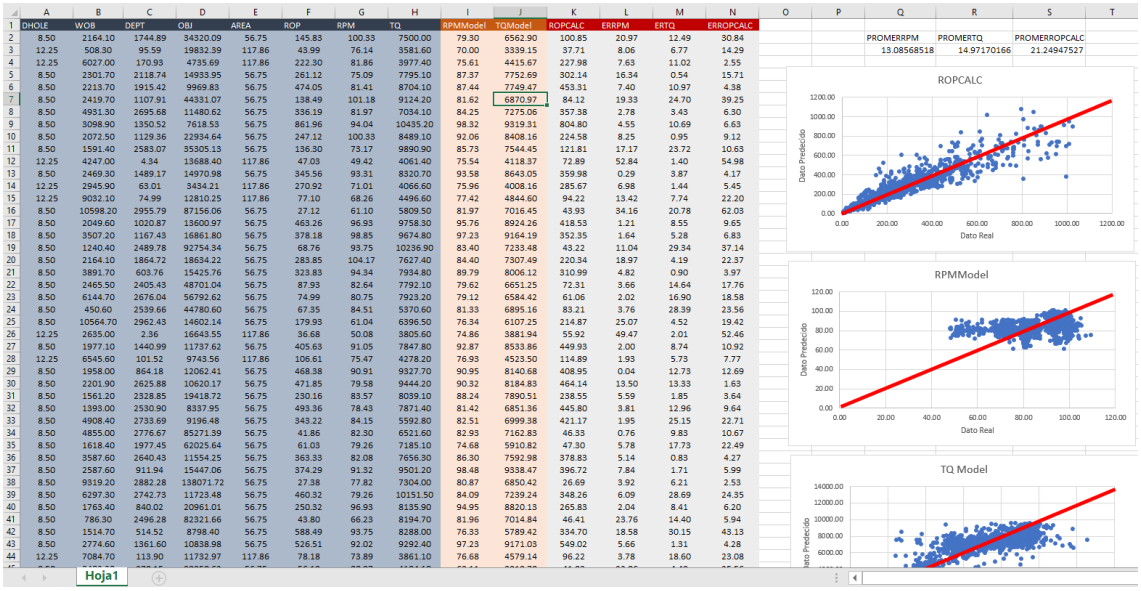


DHOLE	WOB	DEPT	OBJ	AREA	RPM	TORQUE	ROP
12.25	3356.9	157.2983	924.234589	117.858812	74.2183075	4000.99609	1060.37712
12.25	6259.3	74.224	1002.23459	117.858812	69.9913101	4057.17188	957.001591
12.25	8697	176.0021	1004.23459	117.858812	74.7393036	4680.37305	1202.5658
12.25	9318.4	120.7558	918.234589	117.858812	75.3854828	4766.51074	1369.64227
12.25	4765.5	21.1201	965.234589	117.858812	71.693512	3954.0127	980.480282
12.25	11084.7	173.6161	966.234589	117.858812	78.8149109	5207.44824	1505.20043
12.25	10392.8	93.9091	1003.23459	117.858812	77.0311127	4980.23438	1341.02863
12.25	5924.3	166.4579	1000.23459	117.858812	70.8998642	4115.46289	982.480507
12.25	10442.9	176.1513	948.234589	117.858812	77.7321625	5067.40137	1465.6958
12.25	9056.5	184.8008	987.234589	117.858812	75.4211578	4768.17676	1263.53088
12.25	1282.4	110.3167	983.234589	117.858812	74.105423	3709.2041	904.222851
12.25	10550.7	100.6233	964.234589	117.858812	77.3503647	5021.30078	1420.30366
12.25	4190.5	20.3571	939.234589	117.858812	72.4315338	3928.67554	1007.22977
12.25	3849.8	105.6937	978.234589	117.858812	73.3664093	3984.59033	988.908885
12.25	12163	175.1857	913.234589	117.858812	80.6734924	5447.76904	1735.4612
12.25	4923	170.7118	985.234589	117.858812	72.2519836	4078.04834	998.954589
12.25	10134.8	196.2129	990.234589	117.858812	77.3566437	5017.88867	1373.10034
12.25	3998.7	169.3696	991.234589	117.858812	73.4995193	4040.97168	992.403449
12.25	10417.2	158.9387	940.234589	117.858812	77.5583115	5045.69531	1469.456
12.25	9261.9	220.4837	957.234589	117.858812	76.0400391	4846.6958	1341.65637
12.25	6181	172.6505	915.234589	117.858812	70.474617	4123.75146	1077.42965
12.25	3094.6	19.7467	948.234589	117.858812	73.9180145	3885.62524	996.459898
12.25	5975.6	171.01	976.234589	117.858812	70.8223267	4119.146	1008.21792
12.25	7106.5	73.0032	928.234589	117.858812	71.2331848	4231.77246	1110.9259
12.25	1662.2	35.9061	970.234589	117.858812	75.9586487	3843.33813	976.646418
12.25	5616.2	166.1596	956.234589	117.858812	71.2569885	4099.61475	1028.43093
12.25	3243.2	34.9905	959.234589	117.858812	73.807785	3904.20337	989.281034
12.25	9267	114.0449	918.234589	117.858812	75.2469025	4748.88867	1361.36263
12.25	8645.7	178.5373	972.234589	117.858812	74.6686707	4671.2041	1241.18206
12.25	3433.9	172.1248	940.234589	117.858812	74.2089996	4016.81323	1046.5061
12.25	8133.5	71.6299	953.234589	117.858812	72.9857559	4458.34473	1177.11391
12.25	8286.3	190.3968	920.234589	117.858812	74.1385803	4602.27393	1284.11484

	ROP	
PROMEDIO	1123.53924	ERROR
REAL	1026.51	9.45234258
	TORQUE	
PROMEDIO	4276.31484	ERROR
REAL	3513.9	21.6971127
	RPM	
PROMEDIO	73.7450479	ERROR
REAL	82.6446	10.7684617

OBJOptimum | Data 8.5 | Data 12.5

FIGURA 55. Resultados finales y errores porcentuales de los mismos.



DHOLE	WOB	DEPT	OBJ	AREA	ROP	RPM	TQ	RPMModel	TQModel	ROPCALC	ERRPM	ERTQ	ERROPCALC
8.50	2164.10	1744.89	34320.09	56.75	145.83	100.33	7500.00	79.30	6562.90	100.85	20.97	12.49	30.84
12.25	508.30	95.59	19832.39	117.86	43.99	76.14	3581.60	70.00	3339.15	37.71	8.06	6.77	14.29
12.25	6027.00	170.93	4735.69	117.86	222.30	81.86	3977.40	75.61	4415.67	227.98	7.63	11.02	2.55
8.50	2301.70	2118.74	14933.95	56.75	261.12	75.09	7795.10	87.37	7752.69	302.14	16.34	0.54	15.71
8.50	2213.70	1915.42	9969.83	56.75	474.05	81.41	8704.10	87.44	7749.47	453.31	7.40	10.97	4.38
8.50	2419.70	1107.91	44331.07	56.75	138.49	101.18	9124.20	81.62	6870.97	84.12	19.33	24.70	39.25
8.50	4931.30	2695.68	11480.62	56.75	336.19	81.97	7034.10	84.25	7275.06	357.38	2.78	3.43	6.30
8.50	3098.90	1350.52	7618.53	56.75	861.96	94.04	10435.20	98.32	9319.31	804.80	4.55	10.69	6.63
8.50	2072.50	1129.36	22934.64	56.75	247.12	100.33	8489.10	92.06	8408.16	224.58	8.25	0.95	9.12
8.50	1591.40	2583.07	35305.13	56.75	136.30	73.17	9890.90	85.73	7544.45	121.81	17.17	23.72	10.63
12.25	4247.00	4.34	13688.40	117.86	47.03	49.42	4061.40	75.54	4118.37	72.89	52.84	1.40	54.98
8.50	2469.30	1489.17	14970.98	56.75	345.56	93.31	8320.70	93.58	8643.05	359.98	0.29	3.87	4.17
12.25	2945.90	63.01	3434.21	117.86	270.92	71.01	4066.60	75.96	4008.16	285.67	6.98	1.44	5.45
12.25	9032.10	74.99	12810.25	117.86	77.10	68.26	4496.60	77.42	4844.60	94.22	13.42	7.74	22.20
8.50	10598.20	2955.79	87156.06	56.75	27.12	61.10	5809.50	81.97	7016.45	43.93	34.16	20.78	62.03
8.50	2049.60	1020.87	13600.97	56.75	463.26	96.93	9758.30	95.76	8924.26	418.53	1.21	8.55	9.65
8.50	3507.20	1167.43	16861.80	56.75	378.18	98.85	9674.80	97.23	9164.19	352.35	1.64	5.28	6.83
8.50	1240.40	2489.78	92754.34	56.75	68.76	93.75	10236.90	83.40	7233.48	43.22	11.04	29.34	37.14
8.50	2164.10	1864.72	18634.22	56.75	283.85	104.17	7627.40	84.40	7307.49	220.34	18.97	4.19	22.37
8.50	3891.70	603.76	15425.76	56.75	323.83	94.34	7934.80	89.79	8006.12	310.99	4.82	0.90	3.97
8.50	2465.50	2405.43	48701.04	56.75	87.93	82.64	7792.10	79.62	6651.25	72.31	3.66	14.64	17.76
8.50	6144.70	2676.04	56792.62	56.75	74.99	80.75	7923.20	79.12	6584.42	61.06	2.02	16.90	18.58
8.50	450.60	2539.66	44780.60	56.75	67.35	84.51	5370.60	81.33	6895.16	83.21	3.76	28.39	23.56
8.50	10564.70	2962.43	14602.14	56.75	179.93	61.04	6396.50	76.34	6107.25	214.87	25.07	4.52	19.42
12.25	2635.00	2.36	16643.55	117.86	36.68	50.08	3805.60	74.86	3881.94	55.92	49.47	2.01	52.46
8.50	1977.10	1440.99	11737.62	56.75	405.63	91.05	7847.80	92.87	8533.86	449.93	2.00	8.74	10.92
12.25	6545.60	101.52	9743.56	117.86	106.61	75.47	4278.20	76.93	4523.50	114.89	1.93	5.73	7.77
8.50	1958.00	864.18	12062.41	56.75	468.38	90.91	9327.70	90.95	8140.68	408.95	0.04	12.73	12.69
8.50	2201.90	2625.88	10620.17	56.75	471.85	79.58	9444.20	90.32	8184.83	464.14	13.50	13.33	1.63
8.50	1561.20	2328.85	19418.72	56.75	230.16	83.57	8039.10	88.24	7890.51	238.55	5.59	1.85	3.64
8.50	1393.00	2530.90	8337.95	56.75	493.36	78.43	7871.40	81.42	6851.36	445.80	3.81	12.96	9.64
8.50	4908.40	2733.69	9196.48	56.75	343.22	84.15	5592.80	82.51	6999.38	421.17	1.95	25.15	22.71
8.50	4855.00	2776.67	85271.39	56.75	41.86	82.30	6521.60	82.93	7162.83	46.33	0.76	9.83	10.67
8.50	1618.40	1977.45	62025.64	56.75	61.03	79.26	7185.10	74.68	5910.82	47.30	5.78	17.73	22.49
8.50	3587.60	2640.43	11554.25	56.75	363.33	82.08	7656.30	86.30	7592.98	378.83	5.14	0.83	4.27
8.50	2587.60	911.94	15447.06	56.75	374.29	91.32	9501.20	98.48	9338.47	396.72	7.84	1.71	5.99
8.50	9319.20	2882.28	138071.72	56.75	27.38	77.82	7304.00	80.87	6850.42	26.69	3.92	6.21	2.53
8.50	6297.30	2742.73	11723.48	56.75	460.32	79.26	10151.50	84.09	7239.24	348.26	6.09	28.69	24.35
8.50	1763.40	840.02	20961.01	56.75	250.32	96.93	8135.90	94.95	8820.13	265.83	2.04	8.41	6.20
8.50	786.30	2496.28	82321.66	56.75	43.80	66.23	8194.70	81.96	7014.84	46.41	23.76	14.40	5.94
8.50	1514.70	514.52	8798.40	56.75	588.49	93.75	8288.00	76.33	5789.42	334.70	18.58	30.15	43.13
8.50	2774.60	1361.60	10838.98	56.75	526.51	92.02	9292.40	97.23	9171.03	549.02	5.66	1.31	4.28
12.25	7084.70	113.90	11732.97	117.86	78.18	73.89	3861.10	76.68	4579.14	96.22	3.78	18.60	23.08

PROMERRPM	PROMERTQ	PROMERROPCALC
13.08568518	14.97170166	21.24947527

Hoja1

4. CONCLUSIONES

- Debido a la masiva cantidad de información que es manejada en la actualidad, producto de diversas fuentes como sensores, reportes históricos, informes internos, entre otros; Big Data Analytics aparece como una buena alternativa, no convencional, para, por ejemplo, la optimización de procesos, ya que es una técnica que se centra en el manejo de los datos, buscando obtener a partir de ellos, información relevante y de valor, que no puede ser apreciada a simple vista, y que puede representar beneficios para la empresa que la utiliza.
- Mediante la aplicación del análisis de varianza presentado en el software, aplicándolo a diversos pozos, se puede concluir que, siempre que la ROP se mantenga alta, el peso sobre la broca no es un parámetro decisivo dentro de la ecuación de la energía mecánica específica, ya que representa menos de un 1% de error en los resultados, bajo estas condiciones.
- La energía mecánica especifica se presenta como una variable base o guía efectiva para la optimización de la perforación y reducción en el tiempo de la misma, ya que envuelve de manera intrínseca los parámetros principales que involucra la operación.
- El uso de redes neuronales artificiales se presenta como una buena opción para la obtención de soluciones dinámicas, que se ajusten de acuerdo con los datos de entrada suministrados, siempre que se cuente con una base de datos amplia que permita suministrar la cantidad de datos suficientes para el entrenamiento óptimo de la red.
- Mediante la validación realizada con base en datos de pozos colombianos ya estudiados, se pudo determinar que NPT, New Parameters Tool, es una herramienta nueva, que permite obtener una predicción acerca del rango en el que deben mantenerse parámetros de perforación como el torque o las revoluciones por minuto de la broca, de tal manera que se pueda alcanzar un resultado óptimo, que reduzca el tiempo que se invierte en la operación.

- Al analizar grandes volúmenes de datos, es fundamental realizar un filtrado correcto de los datos para obtener resultados precisos, con la menor cantidad de error posible.
- Aplicando la herramienta en diversos pozos se puede observar que los principales parámetros que inciden sobre la eficiencia de la operación son el WOB y las RPM de la broca, asimismo, la variable que presenta el mayor impacto sobre los resultados es la ROP.
- La herramienta NPT, New Parameters Tool, resulta útil a la hora de manejar grandes volúmenes de datos o Big Data, ya que permita, en unos pocos minutos, analizar volúmenes de información que superan los 7000 u 8000 datos por pozo.
- El método de entrenamiento con una distribución de prueba 80/20, como la presentada, permite que la red neuronal alcance un buen grado de ajuste respecto a la tendencia original de los datos, lo que genera que se disminuya la cantidad de error a la salida de esta.

5. RECOMENDACIONES

- Adicionar una mayor cantidad de variables a la hora de efectuar la optimización, para obtener resultados más robustos, al evaluar conjuntos de datos cada vez más amplios.
- Realizar la programación de diferentes tipos de redes neuronales artificiales, con diferentes funciones de transferencia y métodos de entrenamiento a la hora de ejecutar la optimización de variables, comparando los resultados para determinar la opción más eficiente y posible de alcanzar en la práctica.
- Incluir una interfaz gráfica más didáctica, que permita una interacción más dinámica del software con el usuario y, a su vez, genere mayor interés e impacto a simple vista.
- Basados en el método estadístico y las herramientas computacionales evaluadas, es posible diseñar otro tipo de metodología, para realizar la comparación y determinar el mejor curso de acción.
- Ampliar las bases de datos disponibles o el acceso a la información de tal manera que el desarrollo de redes neuronales se vuelva factible y eficiente.
- Utilizar diferentes tipos de códigos y herramientas softwares para determinar en cuál de ellas se puede obtener un flujo de trabajo, con la nueva metodología propuesta, más exacto, rápido y efectivo.

BIBLIOGRAFÍA

ANAND, Pradeep. Big Data Is a Big Deal. Revista JPT. Abril de 2013. 1-3 p.

BECKWITH, Robin. Managing Big Data: Cloud Computing and Co-Location Centers. Revista JPT. Octubre de 2011. 1 p.

BRULÉ, Michael R. Big Data in E&P: Real-Time Adaptive Analytics and Data-flow Architecture. 2013 SPE digital energy conference. Marzo de 2013. 1-2 p.

BRULÉ, Michael R. The Data Reservoir: How Big Data Technologies Advance Data Management and Analytics in E&P. 2015 SPE digital energy conference. Marzo de 2015. 1-4 p.

Data Using Big Data Mining Technique: Heavy Oils Fields Example. International Petroleum Technology Conference. Diciembre de 2014.

DAVYDOVA, Olga. 7 types of Artificial Neural Networks for Natural Language Processing. Septiembre de 2017. [En línea]. Disponible en: <https://medium.com/@datamonsters/artificial-neural-networks-for-natural-language-processing-part-1-64ca9ebfa3b2>

FEBLOWITZ, Jill. Analytics in Oil and Gas: The Big Deal About Big Data. 2013 SPE digital energy conference. Marzo de 2013. 1-2 p.

GALVIS C, Laura V; *et al*. Estimación de propiedades mecánicas de roca utilizando inteligencia artificial. Diciembre 2011.

HAGAN, Martin; DEMUTH, Howard; HUDSON, Mark; DE JESÚS, Mark. Neuronal network design.Septiembre de 2014.

HOLDAWAY, Keith. Harness oil and gas big data with analytics. 2014.

HOWLETT, Robert; JAIN, Lakhmi. Radial basis function networks 1. 2001

INSTITUTE FOR BUSINESS VALUE. IBM. Analytics: el uso de big data en el mundo real Cómo las empresas más innovadoras extraen valor de datos inciertos. 2012.

INSTITUTE FOR BUSINESS VALUE. IBM. Maximo for oil and gas [En línea]. Disponible en: <http://www-03.ibm.com/software/products/es/maximo-for-oil-and-gas>

JONHNSTON, J; GUICHARD, A. New Findings in Drilling and Wells using Big Data Analytics. Offshore technology conference. Mayo de 2015.

KALNISHKAN, Yuri. An introduction to kernel methods. Mayo de 2009.

KOEDERITZ, William. A real-time implementation of MSE. 2005.

KONOVALOV, Serhii. Addressing O&G Data Challenges at the Remote Edge. 2015 SPE digital energy conference. Marzo de 2015.

KOZA, Jhon R. Genetic programming: on the programming of computers by means of natural selection. Enero 1992.

LOGAN, W. D. Engineered Shale Completions Based on Common Drilling Data. SPE. 2015

MICROSOFT AZURE. HDInsight [En línea]. Disponible en: <https://azure.microsoft.com/es-es/services/hdinsight/>

Neuronal Networks. Python course. 2018. [En línea]. Disponible en: <https://www.python-course.eu/neural_networks.php>

ROJAS, Raúl. Neuronal networks: A systematic introduction. Marzo de 1996.

SANTOS, I. H; MACHADO M.M; RUSSO E. E. Big-Data Analytics for Predictive Maintenance Modeling: Challenges and Opportunities. Offshore Technology Conference. Octubre de 2015.

SAS. Company information [En línea]. Disponible en: <http://www.sas.com/es_co/company-information.html#>

SPATH, Jeff. Big Data!. Revista JPT. Enero de 2014. 1 p.

TÓTH, Lázló; GRÓSZ, Tamáz. A Comparison of Deep Neural Network Training Methods for Large Vocabulary Speech Recognition. 2013.

VAN OORT, Eric. Drilling Optimization, Risk and Uncertainty Reduction, and Future Workforce Education Using Big Data Analysis. SPE. Febrero de 2016

VON PLATE, Moritz. Big Data Analytics for Prognostic Foresight. SPE. 2016

WU, Wenkuang; et al. Retrieving Information and Discovering Knowledge from Unstructured Data Using Big Data Mining Technique: Heavy Oils Fields Example. 2014. 1-2 p.