

**DISEÑO DE UN MODELO MACHINE LEARNING PARA LA SELECCIÓN DEL  
MEJOR MECANISMO DE CONTROL DE ARENA ENTRE LOS MECANISMOS ZETA  
FLOW Y HIGH RATE WATER PACK EN EL CAMPO X DE LOS LLANOS  
ORIENTALES DE COLOMBIA**

**YINETH DANIELA LOPEZ LÓPEZ**

**MARIA CAMILA MARTINEZ ZAPATA**

**Proyecto integral de grado para optar al título de Ingeniero de Petróleos**

**Orientador**

**Edinson Enrique Jiménez Bracho**

**Ingeniero de Petróleos**

**FUNDACION UNIVERSIDAD DE AMERICA  
INGENIERIA  
PETROLEOS  
BOGOTA**

NOTA DE ACEPTACIÓN

---

---

---

---

---

---

---

---

Firma del Director  
Presidente Jurado

Firma del

---

Firma del Jurado

---

Firma del Jurado

Bogotá, D.C. \_\_\_\_ de 20\_\_.

**DIRECTIVOS DE LA UNIVERSIDAD**

Presidente de la Universidad y Rector del Claustro

**Dr. MARIO POSADA GARCÍA-PEÑA**

Consejero Institucional

**Dr. LUIS JAIME POSADA GARCIA-PEÑA**

Vicerrectora de Académica y de Investigaciones

**Dra. MARIA CLAUDIA APONTE GONZÁLEZ**

Vicerrector Administrativo y Financiero

**Dr. RICARDO ALFONZO PEÑARANDA CASTRO**

Secretaria General

**Dra. ALEXANDRA MEJIA GUZMAN**

Decano Facultad de Ingenierías

**Ing. JULIO CESAR FUENTES ARISMENDI**

Director del Programa de Ingeniería de Petróleos

**Ing. JUAN CARLOS RODRÍGUEZ ESPARZA**

## DEDICATORIA

Principalmente quiero dedicar este logro a Dios por permitirme vivir esta maravillosa etapa como estudiante, a mis padres Saúl López y Leticia López por brindarme esta oportunidad, quienes son pilares importantes en mi vida y en todo momento me han brindado todo su amor y su apoyo, quiero agradecerles por todo lo que me han enseñado en estos 23 años de vida y por darme fuerza cuando siempre la necesite, a ellos quienes con todo su esfuerzo hicieron esto posible.

A mis hermanos Julieth y Edier López, por ser mis confidentes, gracias por hacerme entender que siempre hay que soñar en grande y que la vida no tiene límite, gracias por darme los sobrinos más hermosos del mundo, espero algún día ser un ejemplo para ellos.

A Andrés Rodríguez quien fue testigo de todas las experiencias vividas durante esta etapa, por motivarme siempre a ser mejor en todos los aspectos de mi vida, por escucharme y aconsejarme, por ser mi polo a tierra, mi mejor amigo, mi mejor complemento, gracias por darle luz a mi vida.

A mi compañera de tesis, Camila Martínez por permitirme esta oportunidad, por todas las experiencias vividas durante esta etapa y por aguantar mi carácter.

Finalmente a todas las personas que sin interés alguno me brindaron su apoyo y que al final fueron parte de esta maravillosa experiencia, espero brindarles siempre lo mejor de mi.

**Yineth Daniela López López**

## **DEDICATORIA**

Quiero expresar mi gratitud a Dios por permitirme llegar hasta aquí de forma satisfactoria y oportuna acompañada por todos mis seres queridos; a mis padres Arquímedes Martínez Zapata y María Suley Zapata Quintero por su apoyo incondicional y sus críticas constructivas para hacerme crecer y llegar hasta donde estoy; a mi tía María Nubia Martínez Zapata por ser parte fundamental de mi vida y mi apoyo incondicional; a mi hermana Lina María García Martínez por llenarme a diario de motivos para seguir luchando, por ser mi paño de lágrimas cuando las cosas no resultaban bien y estar siempre para mí; a cada uno de mis familiares por creer en mí y alentarme cada día a seguir mis sueños hasta conseguirlos.

También quiero expresar mi gratitud a mi compañera de trabajo Yineth Daniela López López por su compañerismo, armonía y entrega para sacar adelante nuestro trabajo investigativo, por último y no menos importante a mis amigos por su apoyo, sinceridad y comprensión.

**María Camila Martínez Zapata**

## **AGRADECIMIENTO**

Agradecemos a la Fundación Universidad de América, por brindarnos siempre las herramientas para formarnos como profesionales íntegros, a el cuerpo docente por su total disposición durante nuestro desarrollo, en especial a nuestro docente y orientador el ingeniero Edinson Jiménez, por sus consejos y su guía durante la elaboración de este proyecto.

Agradecemos a la compañía por confiar en nosotras para el desarrollo de este proyecto, por suministrar la información necesaria, por el tiempo y dedicación para que en equipo sacáramos esto adelante, en especial a los ingenieros Carlos Medina y Leonel Chaparro, por su orientación profesional, su apoyo y su total disposición.

Las directivas de la Universidad de América, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores.

## TABLA DE CONTENIDO

	<b>Pág</b>
INTRODUCCION	13
1. MARCO TEORICO	15
1.1. Producción de arenas y finos	15
1.2. Causas de la producción de arenas	15
1.2.1. Grado de consolidación de la formación	15
1.2.2. Reducción de la presión de poro	17
1.2.3. Velocidad de producción	17
1.2.4. Viscosidad del fluido de reservorio	17
1.2.5. Incrementos de producción de agua	17
1.3. Consecuencias de la producción de arena	18
1.3.1. Acumulación en el equipo de superficie	18
1.3.2. Acumulación en el equipo de fondo	19
1.3.3. Erosión en equipos de fondo y superficie	19
1.3.4. Colapso de Formación	19
1.4. Presión de fractura	20
1.5. Análisis granulométrico	20
1.6. Análisis de tamaño de partículas (PSA)	20
1.7. Zeta Potential	21
1.7.1. Modelo de la doble capa	21
1.8. Zeta Flow	22
1.8.1. Sand Aid	23
1.9. Empaques con Agua a Alta Tasa (HRWP)	23
1.10. Machine Learning	24
1.10.1. Aprendizaje supervisado	24
1.10.2. Aprendizaje por refuerzo	24
1.11. Análisis de conglomerados (Cluster Analysis)	24



1.12.	Python	24
1.12.1.	Operaciones aritméticas	25
1.12.2.	Arboles de decisión y regresión	26
2.	METODOLOGÍA	28
2.1.	Mapa conceptual de metodología general	28
2.2.	Elaboración de la base de datos (Data input)	29
2.3.	Desarrollo del modelo Machine Learning.	31
2.4.	Implementación del modelo de Machine Learning con una selección de pozos del campo X	34
2.5.	Evaluación económica	35
3.	RESULTADOS Y ANÁLISIS DE RESULTADOS	36
3.1.	Clúster Analysis	36
3.2.	Entrenamiento del modelo predictivo (70% de información)	43
3.3.	Calibración del modelo predictivo (30% de información)	49
3.4.	Evaluación económica	51
3.4.1.	Costos por inversión de servicio.	53
3.4.2.	Producción alcanzada	55
3.4.3.	Eficiencia de la inversión	56

## LISTA DE TABLAS

	<b>Pág.</b>
Tabla 1. Rangos de diseños de HRWP	30
Tabla 2. Rangos de aplicación ZF	30
Tabla 3. Parámetros de adición de SandAid (Zeta Flow)	31
Tabla 4 Variables críticas.	32
Tabla 5 Percentil de regresión.	41
Tabla 6 estadísticas de la regresión.	42
Tabla 7 Análisis de varianza	42
Tabla 8 Predicción del entrenamiento	49
Tabla 9 información de pozos de prueba.	52
Tabla 10 Predicción para selección de pozos	52
Tabla 11 Costo de tratamiento HRWP	53
Tabla 12 Costo de tratamiento ZF	54
Tabla 13 Ingreso estimado.	56
Tabla 14 Ganancia estimada.	56

## LISTA DE FIGURAS

<i>Figura 1</i> Columna estratigráfica cuenca de los llanos orientales	16
<i>Figura 2</i> Modelo de la doble capa	18
<i>Figura 3</i> Árbol de decisión	22
<i>Figura 4</i> Árbol de decisión aritmético	25
<i>Figura 5</i>	26
<i>Figura 6</i> Mapa conceptual de metodología general FUENTE: Elaboración propia	28
<i>Figura 7</i> Mapa conceptual para el desarrollo general del modelo machine Learning	31
<i>Figura 8</i> Información de operaciones para HRWP y ZF en el campo X	33
<i>Figura 9</i> Base de datos tomada de Python.	34
<i>Figura 10</i> Gráfico de residuales temperatura de formación.	37
<i>Figura 11</i> Gráfico de residuales permeabilidad.	37
<i>Figura 12</i> Gráfico de residuales porosidad.	38
<i>Figura 13</i> Curva de regresión ajustada de la variable TEMPERATURA.	39
<i>Figura 14</i> Curva de regresión ajustada de la variable POROSIDAD	39
<i>Figura 15</i> Curva de regresión ajustada de la variable PERMEABILIDAD.	40
<i>Figura 16</i> Gráfico de probabilidad normal.	41
<i>Figura 17</i> Gráfico de dispersión de variables	43
<i>Figura 18</i> Dispersión de variables	44
<i>Figura 19</i> Matriz de correlación	45
<i>Figura 20</i> Árbol de clasificación de 2 ramas.	46
<i>Figura 21</i> Importancia de variables	46
<i>Figura 22</i> Árbol de clasificación de 4 ramas.	47
<i>Figura 23</i> Importancia de variables	47
<i>Figura 24</i> Árbol de clasificación de 5 ramas.	48
<i>Figura 25</i> Importancia de variables	49
<i>Figura 26</i> Ejemplo de predicción FUENTE: Tomado de Python	53
<i>Figura 27</i> Costos por inversión de operación.	54
<i>Figura 28</i> Producción alcanzada en el pozo seleccionado	55

## RESUMEN

Una formación con una alta presencia de finos representa un problema en el plan de producción de un campo, acarreando sobrecostos por daños en líneas de producción y equipos. El método de control de arena seleccionado depende de las condiciones específicas del área a trabajar, las prácticas operativas y las consideraciones económicas.

La correcta selección del método de sand control permite moderar o mitigar los efectos negativos de un alto BSW. En este sentido, el objetivo de esta investigación es diseñar un modelo predictivo mediante Machine Learning para la selección del mejor mecanismo de control de arena entre los mecanismos Zeta Flow y High Rate Water Pack en el campo X, logrando con este modelo la reducción de producción de arena. Para llevar a cabo este propósito se desarrolló un modelo predictivo basado en el lenguaje de programación Python, que interpreta los parámetros operativos y antecedentes de trabajo y los asimila como entrenamiento para llegar a la predicción más acertada.

El Machine Learning desarrollado, fue entrenado con 7 variables de 221 pozos operados por la empresa X con los métodos HRWP y ZF, generando un modelo con índice de efectividad del 100% con asunción de datos ( estimado en 78% real) lo que permite en el caso ideal, la disminución de la producción de arena por la correcta selección de un método adecuado de control de arena, El uso de tecnologías como el aprendizaje automático permite la fusión del criterio humano y el criterio tecnológico que posibilita la disminución de errores operativos presentes día a día en la industria petrolera.

Palabras claves: Machine Learning, Sand control, BSW, Zeta Flow, High Rate Water Pack

## INTRODUCCION

En la industria petrolera se ha evidenciado que la extracción de crudo acarrea varias dificultades que afectan directamente la producción, las condiciones estructurales del yacimiento, las condiciones de seguridad y rentabilidad económica. Una de las situaciones más frecuentes en el campo operacional es la alta tasa de producción de arenas, dada por viscosidades de crudo altas, producción de agua, tasa crítica de migración de finos, parámetros de producción y formaciones poco consolidadas. Una mala selección del método puede llevar al fracaso de la operación de Sand control.

El Campo X se encuentra ubicado en la cuenca de los llanos orientales en el departamento de Arauca, siendo unos de los grandes representantes en Colombia de los campos maduros con grandes problemas de producción de arenas. Presenta características con valores promedio de porosidad 25%, permeabilidad 5.000MD, temperatura 207°F y presión de fractura 0,7PSI/ft en unidades como C5 y subunidades K1, M1 de la formación carbonera, pertenecientes a la edad de Oligoceno. Estas unidades se caracterizan por la alternancia de areniscas, areniscas con arcillolitas, limolitas y lutitas, con depositaciones correspondientes a ambientes continentales y marinos someros, granos con tendencia a la migración. Su espesor total varía desde 50 hasta 450 pies y su ambiente de depositación es continental. La formación Carbonera C5 se encuentra en contacto concordante con la Formación Carbonera C6, de igual manera con la Formación Carbonera C4. Posee crudo promedio de 30°API, porcentajes de BSW entre 52% y el 92% ocasionando un aumento de producción de arena en el campo, una disminución en la producción de crudo y deterioro en equipos.

Con el fin de mitigar la alta producción de arena se decidió diseñar un modelo machine Learning para la selección del mejor método de control de arena entre los métodos Zeta Flow y High Rate Water Pack en el campo x de los llanos orientales de Colombia mediante la elaboración de una base de datos con variables operacionales e históricas de producción, implementando el lenguaje de programación Python.

Desde el 2013 se han realizado operaciones de sand control en el campo X con las metodologías a evaluar en esta investigación, registrando aproximadamente 29 pozos con el tratamiento químico de control de arena denominado Zeta Flow de los cuales 9 pozos no tuvieron la respuesta esperada, se realizaron tratamientos adicionales mediante la adición de ácido fórmico y/o fluoruro de hidrogeno según corresponda, por otro lado se registraron 5 operaciones con tratamiento mecánico denominado High Rate Water Pack de los cuales 2 no arrojaron los resultados esperados, también se les realizo tratamiento adicional con ácido clorhídrico (HCl) y/o fluoruro de hidrogeno (HF) generando así costos adicionales, la implementación de la herramienta permite reducir la incertidumbre evitando el uso de tratamientos adicionales o un futuro fracaso operacional

La metodología planteada tiene en cuenta diversos datos que van desde la recolección de información de pozos del campo X, desarrollar un sistema automático Machine Learning, determinar la eficiencia del modelo evaluando los métodos High Rate Water Pack y Zeta flow, determinar el método de control de arena más adecuado mediante la predicción del Machine Learning, definir los criterios y consideraciones para las recomendaciones sobre operaciones futuras basadas en la predicción arrojada por el Machine Learning, teniendo en cuenta las variables petrofísicas, litológicas y geológicas que tienen como consecuencia el aumento de la producción de arena, por lo tanto el estudio y correcta selección del mejor mecanismo de control de arena entre Zeta Flow (ZF) y High Rate Water Pack(HRWP) permitirá disminuir el %BSW y aumentar el Qo en cada pozo, optimizando los problemas de alta producción de arena existente.

## **1. MARCO TEORICO**

### **1.1. Producción de arenas y finos**

La producción de arenas y finos se define como el arrastre de partículas sólidas provenientes del yacimiento hacia el pozo y hacia la superficie, el material sólido producido de un pozo puede ser fino (todo material con diámetro menor a 44 micrones) o material de carga (todo material que al ser movilizado afecta los esfuerzos en la roca) dentro del cual se incluye todo material desprendido de la roca con diámetro superior a 44 micrones.

### **1.2. Causas de la producción de arenas**

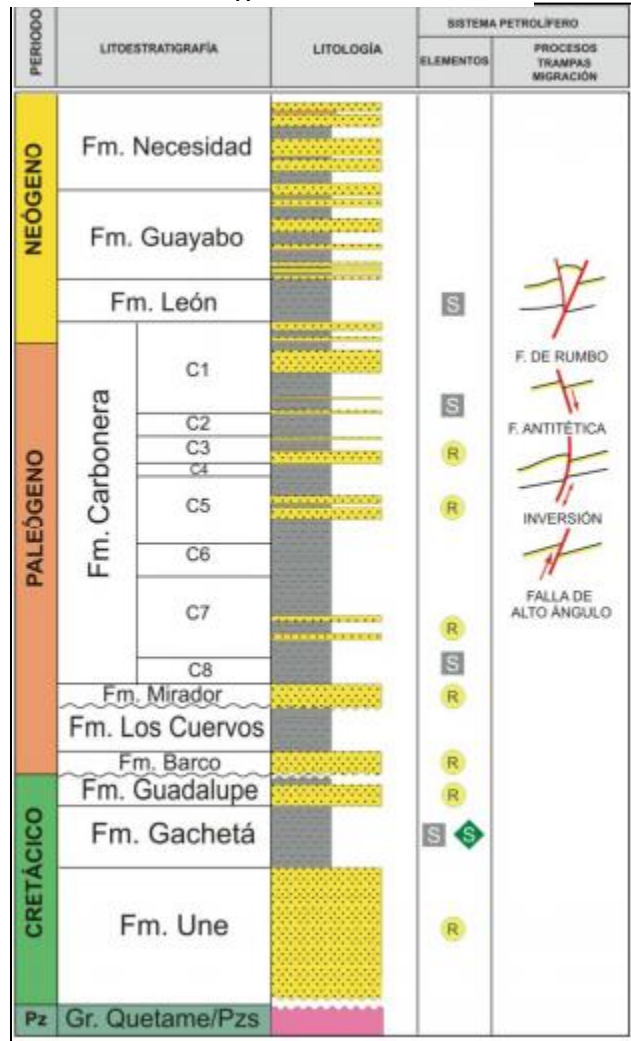
#### ***1.2.1. Grado de consolidación de la formación***

La habilidad de mantener abiertos los túneles de perforación está relacionada con la cementación de los granos de arena alrededor de dicho túnel, la cementación de las areniscas es típicamente un proceso geológico secundario y es una regla general, antiguos sedimentos tienden a ser más consolidados que los nuevos, esto indica que la producción de arena es un problema normal cuando se producen por formaciones poco profundas, geológicamente jóvenes de la era terciaria. Las formaciones jóvenes de la era terciaria suelen tener una matriz de material cementante pequeña, juntando los granos y manteniéndolos juntos, estas formaciones generalmente tienden a ser poco consolidadas o no consolidadas.

La formación carbonera es una formación que se encuentra constituida por secuencias arenosas (C1, C3, C5 y C7) depositadas en una llanura costera inferior con fuerte influencia marina, separadas por unidades arcillosas (C2, C4, C6 y C8) que representan episodios de máxima inundación. En la formación Carbonera se incrementan las facies arenosas, hacia el oriente, debido a la proximidad al escudo de Guayana. Por otra parte las facies arcillosas, constituyen sellos verticales y laterales para los yacimientos de la Formación Mirador [20].

**Figura 1**

Columna estratigráfica cuenca Llanos



La columna estratigráfica muestra la composición geológica de la cuenca de los llanos orientales. Tomado de ANH. Colombia. 2009. P.31 (consultado: 11 de junio de 2020) <https://www.anh.gov.co/Informacion>



### **1.2.2. Reducción de la presión de poro**

La presión de un reservorio es disminuida a lo largo de la vida productiva de un pozo, se elimina parte del soporte para la roca suprayacente, bajando la presión del yacimiento se crea un incremento de la cantidad de tensión en la formación de arena, a este punto los granos de arena pueden romperse de la matriz o pueden deformarse creando finos que serán producidos después con el fluido. La Compactación de la roca reservorio debido a la reducción de la presión de poro puede resultar en un hundimiento de la cara de la formación

### **1.2.3. Velocidad de producción**

La producción de los fluidos de reservorio crean diferenciales de presión y fricción por fuerzas de arrastre que combinadas exceden las fuerzas compresivas de la formación, esto indica que la velocidad de producción para muchos pozos no es buena para las fuerzas compresivas y causa producción de arena, la alta velocidad de producción de un pozo puede ser determinante por el lento incremento de producción de arena

### **1.2.4. Viscosidad del fluido de reservorio**

Las fuerzas de arrastre y fricción ejercidas en los granos de la formación de arena son creadas por el flujo del fluido de reservorio, estas fuerzas de arrastre y fricción son directamente proporcionales a la velocidad del fluido y la viscosidad del fluido del reservorio, una alta viscosidad puede ser aplicada a una mayor fuerza de fricción a los granos de arena que si fluyera a una baja viscosidad- La influencia del arrastre viscoso causa que las arenas sean producidas por reservorios de crudo pesado que contienen baja API, Alta viscosidad y Baja velocidad de flujo

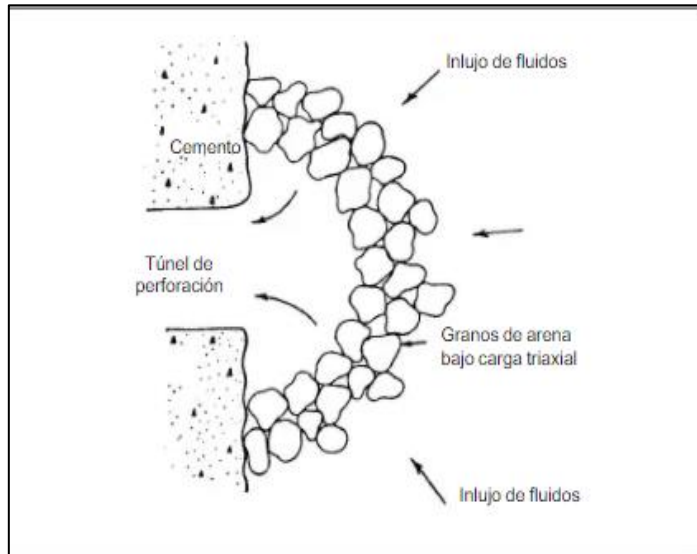
### **1.2.5. Incrementos de producción de agua**

Es el incremento total del fluido de producción para mantener tasas de producción de petróleo o gas con fuerzas de arrastre a la arena dado por disolución de material cementante y/o Perturbación de las fuerzas cohesivas que mantienen los granos juntos y que con el agua comienzan a moverse

Adicionalmente la producción de agua tiene un límite para estabilizar el arco de arena alrededor a una perforación resultando, si se sobrepasa hay producción de arena

### **Figura 2**

Incremento en producción de agua



*Nota: Comportamiento el influjo de fluidos en un túnel de perforación Tomado de: Ott William k. and Woods Joe D. 1*

## **1.3. Consecuencias de la producción de arena**

### **1.3.1. Acumulación en el equipo de superficie**

Si la velocidad de la producción de arena es lo suficientemente alta para llegar al tubing de producción, la arena comenzará a quedarse atrapada en el separador, calentador o en la línea de flujo de superficie, si es un volumen suficientemente grande la arena comenzará a quedarse atrapada en una de estas áreas, por lo cual será requerida la limpieza del equipo para permitir una producción eficiente del pozo.

### ***1.3.2. Acumulación en el equipo de fondo***

Si la velocidad de producción de arena no es suficientemente alta para llevar la arena a superficie, comenzará a llenar el interior de la tubing. Finalmente, el intervalo de producción puede estar completamente cubierto de arena en este caso la velocidad de producción disminuirá porque el pozo comenzará a arenarse y la producción cesará, en situaciones como esta las operaciones remediales son requeridas para limpiar fuera del pozo y restaurar la producción.

### ***1.3.3. Erosión en equipos de fondo y superficie***

En pozos de alta productividad, los fluidos contenidos con arena a alta velocidad pueden erosionar excesivamente los equipos en fondo y los equipos en superficie conduciendo a un mantenimiento frecuente y un reemplazo del equipo dañado, si la erosión es intolerante y ocurre en un tiempo prolongado, hay una falla completa en el equipo de fondo o superficie, resultando un riesgo para la seguridad y un daño ambiental en el pozo, lo que lleva a la suspensión de la producción.

### ***1.3.4. Colapso de Formación***

El colapso de la formación alrededor del pozo ocurre cuando se producen grandes volúmenes de arena. Aparentemente, cuando se forma un vacío y se vuelve lo suficientemente grande como para soportar adecuadamente las formaciones suprayacentes, el colapso ocurre debido a la falta de material para proporcionar soporte. Cuando ocurre este colapso, los granos de arena pueden reorganizarse para crear una menor permeabilidad que la que ha existido El colapso de la formación es particularmente importante si los materiales llenan o llenan parcialmente los túneles de perforación. Cuando una pequeña cantidad de material de la formación llena los túneles de perforación dirigiendo a un significativo incremento en la presión cayendo a través de la formación cerca del pozo para un caudal dado.

#### **1.4. Presión de fractura**

Fuerza por unidad de área, necesaria para vencer la presión de formación y la resistencia matricial de la roca, esta resistencia que se opone a que la roca se fracture hidráulicamente, dependiendo principalmente de la solides, cohesión de la misma, y de los esfuerzos compresivos a los que se someta, a medida que aumenta la profundidad de la formación, aumentan los esfuerzos compresivos por sobrecarga de las demás formaciones, dado esto se puede afirmar que las fracturas originadas en las formaciones someras tienden a ser horizontales.

#### **1.5. Análisis granulométrico**

El análisis granulométrico es un proceso mediante el cual se analiza la muestra de roca de una formación; consiste en una serie de tamices, donde en el tope se coloca la muestra de arena de formación, cada tamiz tiene una malla de tamaño menor al anterior ubicadas en la plataforma vibratoria que provoca un movimiento y a su vez la caída de los granos de arena hasta llegar a un tamaño de malla que no permite su paso hacia la malla siguiente de la plataforma. Después del tamizaje se realiza la toma del peso de las muestras retenidas por cada una de las mallas.

#### **1.6. Análisis de tamaño de partículas (PSA)**

También denominado medición de tamaño de partículas o dimensionamiento de partículas, es el nombre de las técnicas de laboratorio que determinan el rango de tamaño y el tamaño medio de las partículas en una muestra. El análisis de tamaño de partícula es parte de la ciencia de partículas, y su determinación se lleva a cabo generalmente en laboratorios de tecnología de partículas. En la industria petrolera los métodos más comunes son el de análisis de tamiz y el de difracción de luz láser (LPSA).

## **1.7. Zeta Potential**

El potencial Zeta es una medida de la carga eléctrica adquirida en el corte interfaz del límite de corte hidrodinámico entre superficies sólidas como producto de la repulsión electrostática y las fuerzas de atracción relacionadas con las fuerzas de Van der Waals, se considera una función de la carga de cualquiera de las partículas absorbidas en la interfaz y la naturaleza de la composición del medio de suspensión circundante, puede ser afectado por cambios de PH, conductividad del medio (salinidad) y la concentración de aditivos específicos (polímeros, surfactantes, etc.) Dichas partículas con potencial Zeta tienen valores entre -20 y 20 mV teniendo una carga efectiva lenta suficiente para que la repulsión entre ellos se reduzca a un punto en que puede ocurrir conglomeración (floculación y agregación).

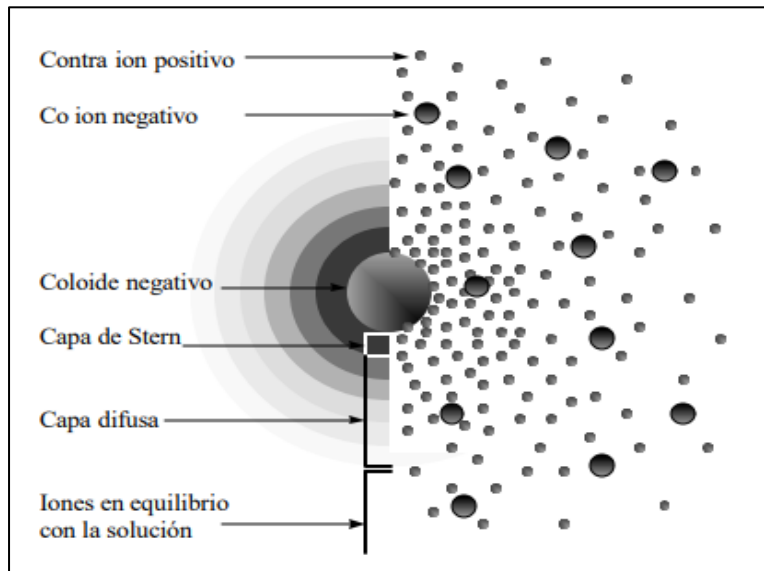
### **1.7.1. Modelo de la doble capa**

Según la hipótesis de Helmholtz (1879) [1] El origen de la doble capa en la interfaz sólido-líquido se debe a que los iones cargados opuestos a la superficie se distribuyen en paralelo frente a ella, dando como resultado la aparición de dos signos cargados y opuestos. El modelo de dos capas se utiliza para visualizar la atmósfera de iones cerca del coloide cargado y explicar la forma de la fuerza repulsiva. Si el ion que neutraliza su carga se extrae repentinamente, el modelo puede entenderse como una serie de etapas que ocurren alrededor de un solo coloide negativo.

La figura mostrada a continuación representa el cambio en la densidad de carga alrededor del coloide. El lado derecho muestra la distribución de iones positivos y negativos alrededor del coloide cargado.

### Figura 3

#### Modelo doble capa



*Nota: la figura representa el Comportamiento de iones en una capa difusa tomado de Zeta-Meter.inc, pag.2, Coagulation & Flocculation. 1993, Comportamiento de iones en una capa difusa. <http://www.zeta-meter.com/coag.pdf>*

### 1.8. Zeta Flow

Zeta Flow es un método químico utilizado para aumentar la tasa máxima libre de arena, eliminar o reducir significativamente la migración de finos y puede reducir el corte de agua. No es una solución absoluta para el control de arena como un paquete de grava, un paquete de Frack o una pantalla expandible, pero permite aumentar las fuerzas de conglomeración entre la arena y los granos finos, modificando la permeabilidad relativa de la formación; Su función principal es conglomerar granos finos de arena y es por lo tanto, un agente de conglomeración de matrices. Funciona modificando el Potencial Zeta del grano de arena a valores entre 0 y -20 mV, aumentando las fuerzas de conglomeración entre los granos de arena

Una de las principales características de Zeta Flow es la adaptación a las condiciones cambiantes como las variaciones en la tensión de formación dada por el agotamiento del reservorio y sus Objetivos de la aplicación son:

- Ampliar el número de días de producción de un pozo sin que éste se detenga por problemas de arenamiento
- Maximizar la tasa de producción de un pozo libre de arena

### **1.8.1. Sand Aid**

Es una solución química utilizada en los trabajos de control de arena con Zeta Flow, aumentando significativamente la consolidación de la arena que conlleva a un aumento en la producción, este proceso químico altera y aumenta la atracción entre las partículas de arena no consolidada sin causar daño en la matriz [2], maximizando las tasas de flujo e incrementando el caudal de producción.

### **1.9. Empaques con Agua a Alta Tasa (HRWP)**

Como su nombre lo sugiere corresponde a un paquete de grava colocado por encima de la presión de fractura que usa generalmente una salmuera o geles lineales ligeros como fluido apuntalante, y la presión hidráulica se usa para generar el empaque de presión en el lugar a perforar minimizando el riesgo de fracturar zonas productoras de gas o agua en cercanía, La cantidad de grava colocada es de aproximadamente 40 a 75 lbs/ft perforado, Este método solo es adecuado para pozos petroleros con alta resistencia a las fracturas. Debido a la alta velocidad y presión de bombeo, la tubería inferior y el equipo deben poder soportar estas duras condiciones. Presenta Pérdidas de fluido de por lo menos 0.2 gpm/perforación Aproximadamente 0.5 bpm por cada 10 pies En caso de arenas múltiples o zonas largas se requiere cierto criterio, Tratamiento ejecutado por debajo de la presión de fractura, Salmuera como fluido de transporte, Concentración de grava entre 1.0 y 2.0 ppa.

## **1.10. Machine Learning**

Aprendizaje automático (machine Learning) es el estudio computacional de algoritmos que mejoran el rendimiento en función de la experiencia, consiste en programar un equipo para optimizar un criterio de desempeño utilizando datos de experiencia pasada o ejemplos si no existe experiencia humana, desarrollando así dos tipos de aprendizaje:

### **1.10.1. Aprendizaje supervisado**

El objetivo es realizar un mapeo con datos de entrada y de salida cuyos valores correctos son proporcionados por un supervisor (programador), La finalidad es encontrar regularidades en la data input, seleccionar los patrones que ocurren con mayor frecuencia que otros, generalidades y realizar una estimación de densidad de datos.

### **1.10.2. Aprendizaje por refuerzo**

En algunas aplicaciones, la data input del sistema es una secuencia de acciones, en tal caso, una sola acción no es relevante, lo importante es la secuencia de acciones correctas para alcanzar el objetivo. El sistema de aprendizaje responde a las valoraciones acerca de la competencia de la respuesta obtenida, cuando la respuesta es correcta el aprendizaje por refuerzo se convierte en aprendizaje supervisado

## **1.11. Análisis de conglomerados (Cluster Analysis)**

Es una técnica multivalente que consiste en definir grupos mediante el cálculo de similitudes a partir de los valores de las variables que se consideren adecuadas para ello, en este análisis todas las variables incluidas contribuyen en la formación de grupos en igualdad de condiciones, la ocurrencia de esta agrupación se debe a la supuesta existencia de una relación natural entre los casos de cada conglomerado.

## **1.12. Python**

Es un lenguaje de programación multiparadigma interpretado, cuya filosofía hace hincapié en la legibilidad de su código, cuya programación está dirigida a objetos siendo así imperativa y funcional, Debido a sus características aritméticas y lógicas,



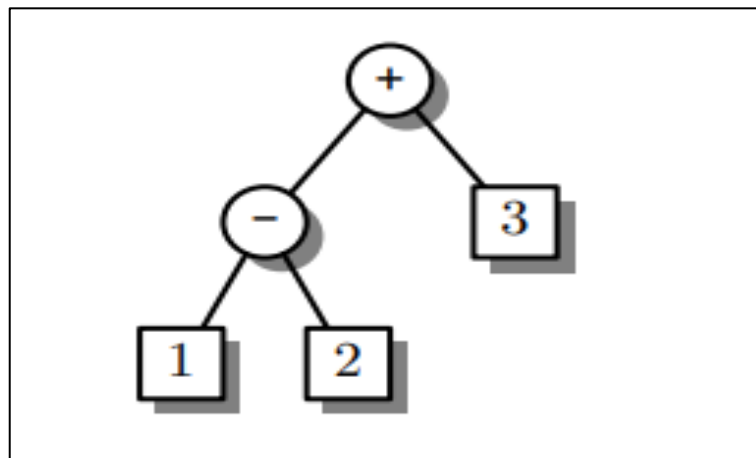
puede ejecutar instrucciones utilizando un lenguaje de programación para hacer un software apropiado que puede usarse de forma independiente o en combinación.

### 1.12.1. Operaciones aritméticas

Las operaciones aritméticas como suma, resta multiplicación y división se denotan con los símbolos operadores +, -, x y ÷ respectivamente; Python codifica los datos ingresados en el orden en la cual se digitan las operaciones Un árbol sintáctico es una representación gráfica en la que acondicionamos los operadores y los operandos como nodos y en los que cada operador está conectado a sus operandos. El árbol sintáctico de la expresión <<1 - 2 + 3>> es:

**Figura 4**

Árbol sintáctico



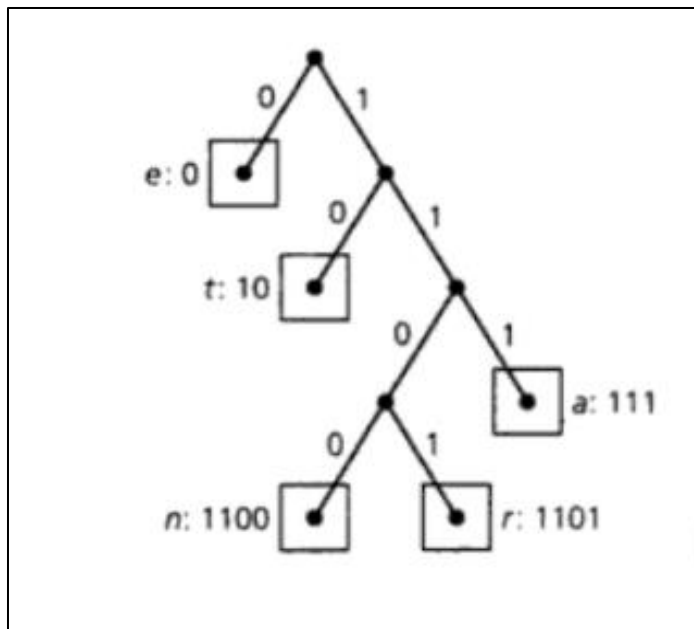
*Nota: Operaciones en un árbol sintáctico, tomado de GARCIA, Isabel. MARZAL, Andrés. Introducción a la programación con Python. Castellón.: Departamento de lenguajes y sistemas informáticos Universitat Jaume I. 2003.*

### 1.12.2. Árboles de decisión y regresión

Los árboles de decisión y regresión (CART= Classification and Regression Trees) son una excelente opción dentro del campo analítico para la clasificación/discriminación o a la predicción tradicional (regresión), son árboles robustos, de fácil interpretación que permite la transformación de variables monótonas independientes [3]; Un árbol de decisión se convierte en un árbol de regresión cuando la variable dependiente es continua y es de tipo cualitativa, sus bifurcaciones o “ramas” permiten que al final de estas obtengamos una predicción para la clase de pertenencia o para el valor que tome.

Este árbol se construye mediante el algoritmo de segmentación recursiva mediante los procedimientos de CHAID (Chi-square Automatic Interaction Detector) QUEST (Quick Unbiased Efficient Statical Tree) y CART

*Figura 5 Ejemplo de árbol de decisión y regresión*



*Nota: Jerarquización de elementos en un árbol de decisión y regresión tomado de Grimaldi Ralph P. Matemáticas discreta y combinatoria, Pearson 1998*

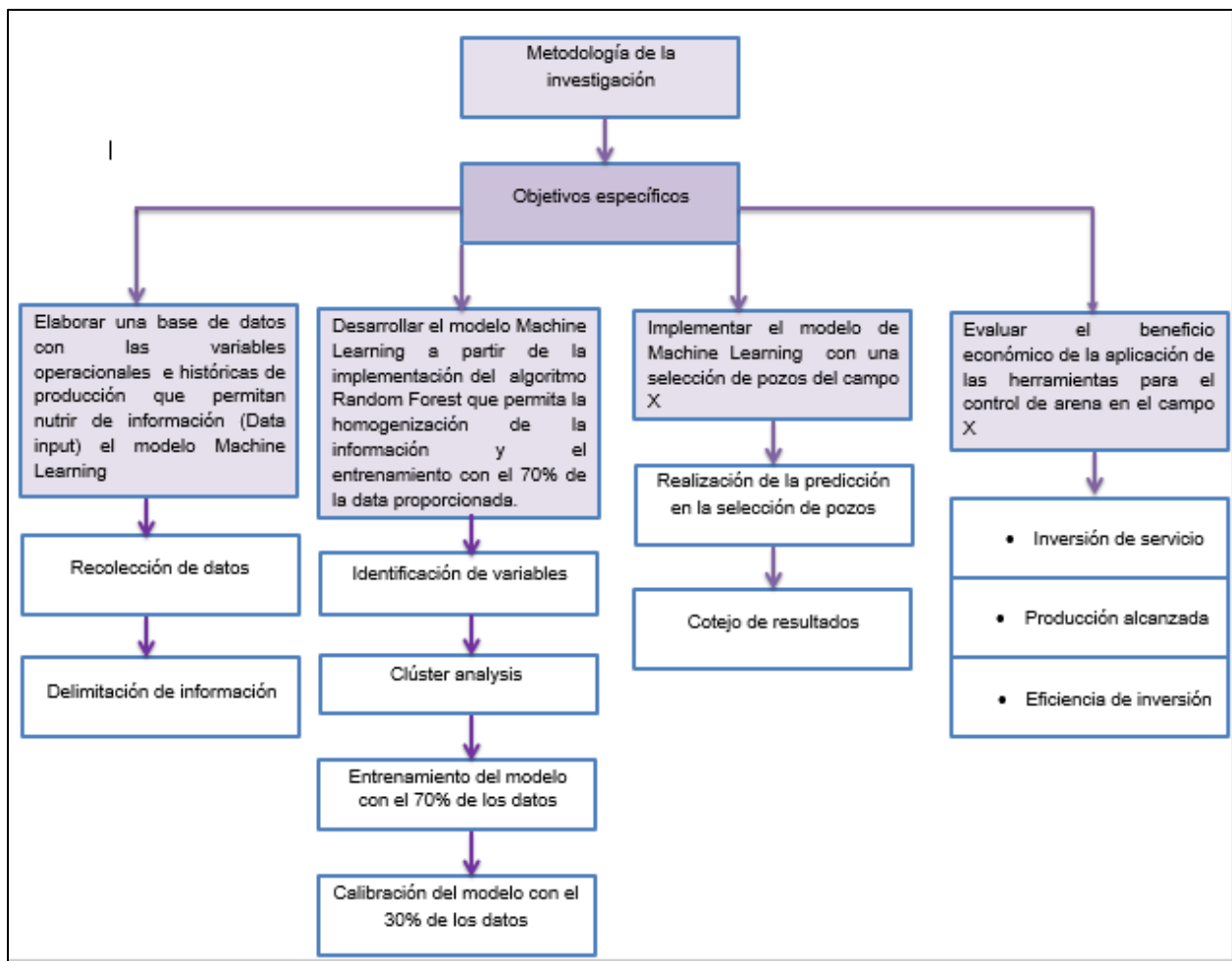


## 2. METODOLOGÍA

### 2.1. Mapa conceptual de metodología general

Para llevar a cabo el objetivo general de este proyecto de investigación se llevara a cabo la siguiente metodología en su respectivo orden.

Figura 6 Mapa conceptual de metodología general



Nota: Descripción de la metodología a trabajar

## **2.2. Elaboración de la base de datos (Data input)**

En la elaboración de la base de datos necesaria para el desarrollo del modelo machine Learning se necesitó el apoyo de la empresa X, la cual suministro toda la información necesaria sobre sus operaciones realizadas en la cuenca de los llanos orientales con los dos métodos de control de arena planteados, para un total de 221 pozos. La data a utilizar en el presente estudio se obtiene de los datos de operaciones con High Rate Water Pack (HRWP) y Zeta Flow desde el 14 de agosto de 2011 hasta el 19 de noviembre de 2019, de esta base de datos se realizó una delimitación adicional para la selección de las operaciones realizadas por la empresa X en el campo X entre el 8 de abril de 2013 hasta el 25 de octubre de 2019, esta base de datos se desarrolló en Excel y es de carácter confidencial; Se realizó una selección adicional de pozos que tuvieran la mayor cantidad de información, dado que esta pertenece a terceros y su recolección en algunos casos es inaccesible. Se seleccionaron 149 pozos con la siguiente información:

- Nombre del pozo
- Campo/bloque
- Etapa
- Tratamiento de control de arena
- Permeabilidad
- Porosidad
- Temperatura de formación
- Gradiente de fractura
- Gravedad API
- Presión de yacimiento
- Caudal Máximo
- WHP promedio
- Caudal Promedio

- Volumen de preflujo
- Volumen de tratamiento
- Volumen Postflujo
- BSW antes del tratamiento

Adicional a esto, se debe tener en cuenta que la implementación de las herramientas debe cumplir con los siguientes parámetros de aplicación y/o diseño.

**Tabla 1. Rangos de diseños de HRWP**

<i>Variables</i>	<i>Límite inferior</i>	<i>Límite superior</i>
Concentración de grava (ppa)	1	2
Salmuera (BPM)	8	12
Longitud de fractura (ft)	5	10

*Nota: La presente tabla muestra los rangos de implementación del HRWP*

**Tabla 2. Rangos de aplicación ZF**

<i>Variable</i>	<i>Límite inferior</i>	<i>Límite superior</i>
Espesor (ft)	2	62
Temperatura °F	100	228
Permeabilidad (md)	3,6	8,000
Porosidad (%)	15	38
Tratamiento por etapa (bbl)	12	373
Tasa de bombeo	0,005	5
Presiones de superficie	0	3,400

*Nota: La presente tabla muestra los rangos de implementación del ZF*

**Tabla 3. Parámetros de adición de SandAid (Zeta Flow)**

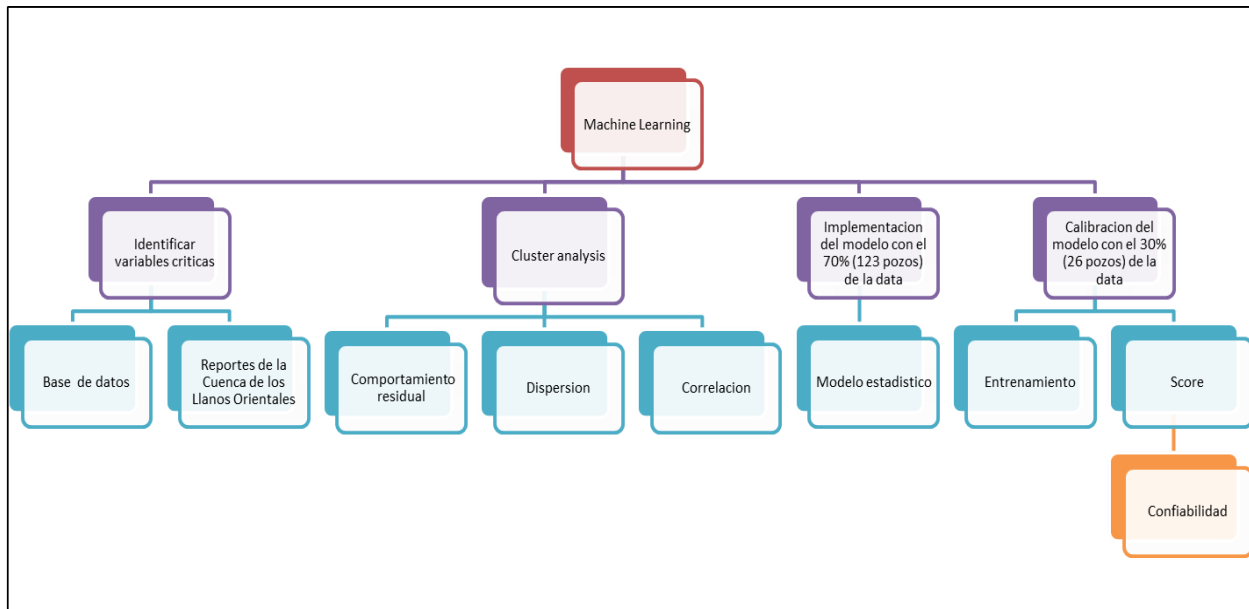
Permeabilidad (Md)	SandAid concentration %
200000	8.0%
2000	7.5%
1000	7.0%
400	6.5%
100	6.0%
50	5.5%
0	5.0%

*Nota: La presente tabla muestra los rangos de implementación del aditivo*

### 2.3. Desarrollo del modelo Machine Learning.

Para el adecuado desarrollo del modelo predictivo Machine Learning se tendrá en cuenta el siguiente mapa conceptual.

*Figura 7 Mapa conceptual para el desarrollo general del modelo machine Learning*



*Nota: Pasos para el Desarrollo del Machine Learning*

Para iniciar con el desarrollo del segundo objetivo se deben identificar las variables críticas en la producción de arena en el campo X de manera teórica, teniendo en cuenta variables generales como temperatura de formación, permeabilidad, porosidad, gradiente de fractura, Gravedad API, presión de yacimiento, Qo antes del tratamiento con High Rate Water Pack o Zeta Flow y porcentaje de agua y sedimentos (BSW) antes del tratamiento con High Rate *Water Pack* o *Zeta Flow* presentadas en la tabla 4.

*Tabla 4 Variables críticas.*

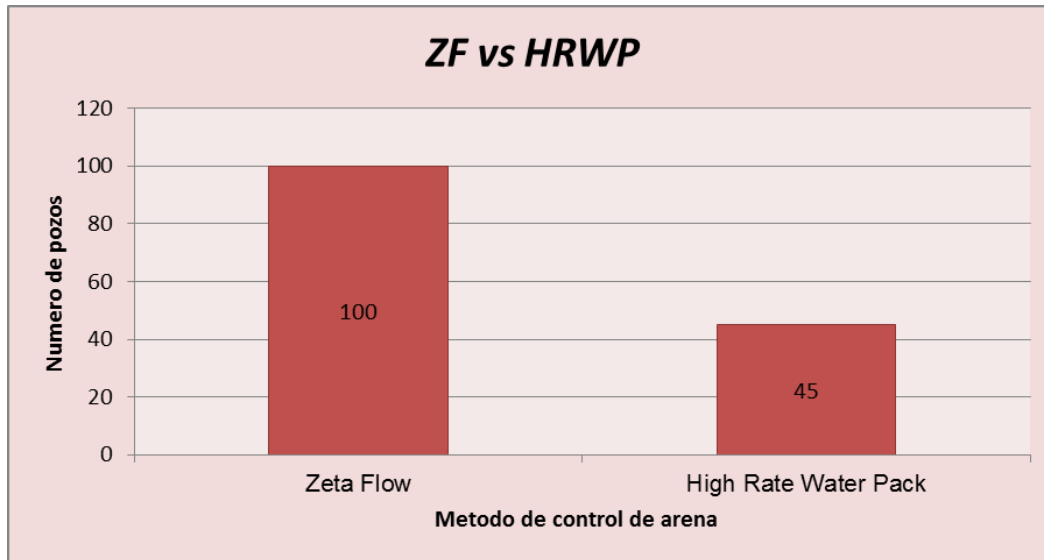
<b>Variable</b>	<b>Valor promedio</b>
Temperatura de formación °F	158,76
Permeabilidad Md	1499,36
Porosidad %	23,82
Gradiente de fractura PSI/ft	0,74
Gravedad °API	22,01
Presión de yacimiento PSI	1555,53
% BSW	60,36
Q promedio BPM	2,21

*Nota: La presente tabla determina los valores promedios de la base de datos*

Se debe tener en cuenta que la densidad de datos para operaciones de ZF es mucho mayor a la de HRWP como se muestra en la figura 8, dado que los tratamientos mecánicos suponen un funcionamiento óptimo a largo plazo y además de esto ha significado valores de inversión mucho mayores.



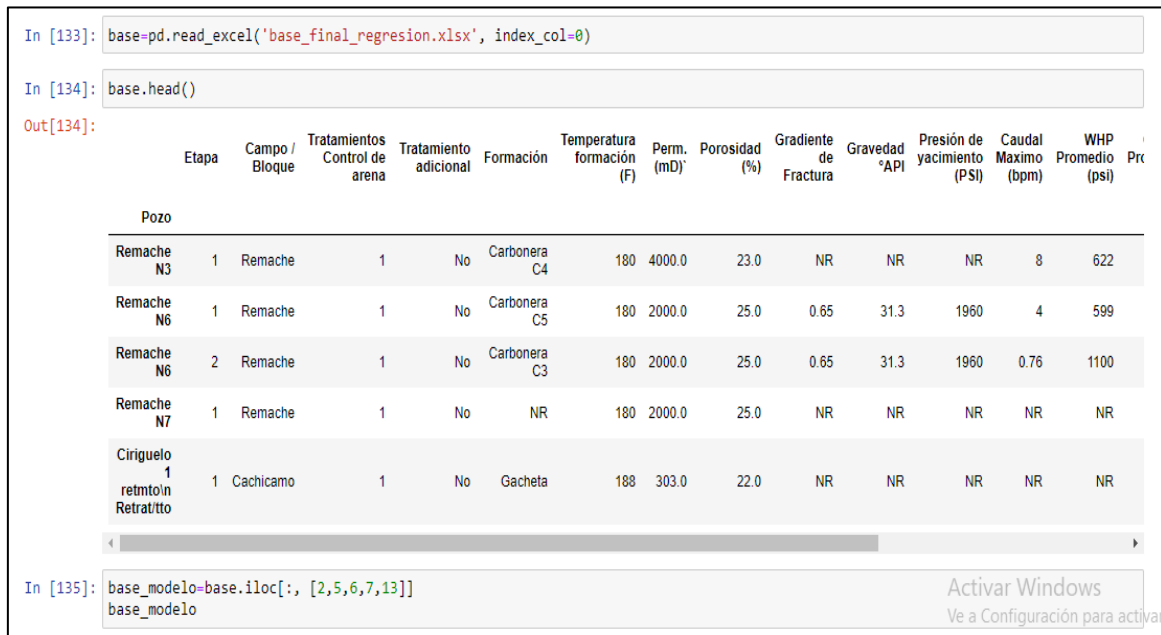
Figura 8 información de operaciones para HRWP y ZF en el campo X



*Nota: Comparativo de operaciones realizadas con HRWP Y ZF en el Campo X.*

Para iniciar con el desarrollo del modelo se cargará la base de datos al programador definiendo las librerías a utilizar (Random Forest para el desarrollo base y Numpi, Pylab, Stats models entre otras para la organización de datos y diseño de graficas). Luego se seleccionara la información necesaria de la base de datos teniendo en cuenta la definición de las variables críticas y se dividirán los datos en *train* (para entrenamiento) y *test* (para prueba y calibración).

Figura 9 Base de datos tomada de Python.



Nota: Base de datos desde Python

Para la fase de entrenamiento (train) se utilizara el 70% de la información mediante los métodos de regresión lineal de múltiple intercepto para determinar la influencia de las variables y árbol de decisión por regresión para definir el modelo estadístico, error típico y error cuadrático medio (rmse).

Después de definir los valores de confiabilidad iniciales se continuará con la fase de prueba y calibración del modelo (test) con el 30% de la información restante, asignándole la a variable de resultado = *TRATAMIENTO* valores Uno “1” para ZF y Cero “0” para HRWP, estableciendo finalmente el score de entrenamiento definitivo para este modelo.

#### 2.4. Implementación del modelo de Machine Learning con una selección de pozos del campo X

Se considerara una selección de 4 pozos de estudio del campo X que ya tienen establecidos un método de control de arena, teniendo en cuenta las variables

mencionadas anteriormente como parámetro de entrada para realizar el análisis en el modelo de predicción, los resultados arrojados corresponden al éxito de las operaciones realizadas sin el modelo machine Learning estimadas entre un 70 y un 80%, si el porcentaje de éxito es menor al 70% se estudiarán y analizarán las causas y así mismo se realizará una sugerencia de operación para cada pozo involucrado.

## **2.5. Evaluación económica**

Para el cumplimiento del último objetivo se deben tener en cuenta las predicciones realizadas por el machine Learning, se realizará una estimación económica general teniendo en cuenta costos base para la implementación de los métodos de control de arena sin la implementación de herramientas o elementos adicionales para HRWP y ZF en los campos de aplicación correspondientes a inversión del servicio, producción alcanzada después de la implementación del sistema de control de arena y eficiencia de la inversión

### 3. RESULTADOS Y ANÁLISIS DE RESULTADOS

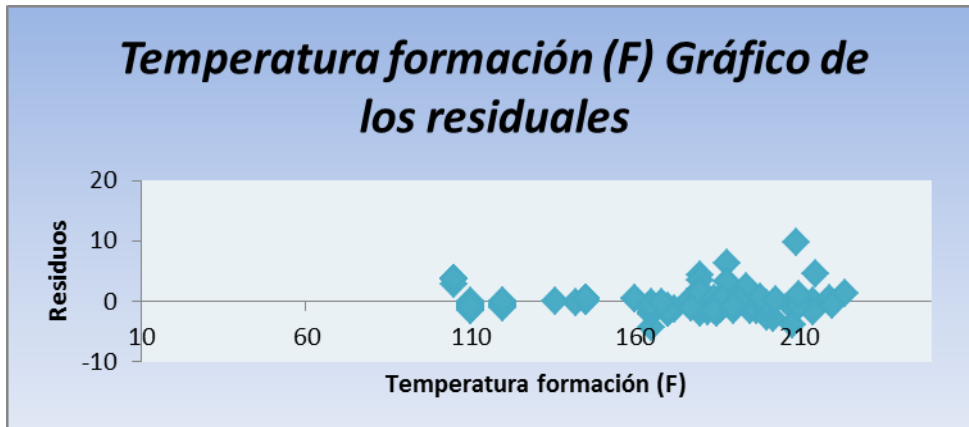
Este proyecto está orientado por la metodología descrita en la sección anterior, fue desarrollada en su respectivo orden obteniendo los siguientes resultados.

#### 3.1. Clúster Analysis

Teóricamente las variables temperatura de formación, permeabilidad, porosidad, gradiente de fractura, Gravedad API, presión de yacimiento,  $Q_0$  antes del tratamiento con High Rate Water Pack o Zeta Flow y porcentaje de agua y sedimentos (BSW) antes del tratamiento con High Rate Water Pack o Zeta Flow, son las más influyentes en el proceso de producción de arenas y finos en un pozo. Al desarrollar la base de datos se encontró que la ausencia de información en variables influye considerablemente en el porcentaje de confiabilidad final de este modelo, por lo tanto inicialmente se consideró la asunción de datos para las variables %BSW y Gravedad API, como una solución eficiente similar al overfitting realizado anteriormente, finalmente se descartaron por falta de información. Como consecuencia de esto se realizó el Cluster Analysis a las variables temperatura de formación, permeabilidad y porosidad que fueron clasificadas como variables independientes y caudal promedio de producción clasificada como variable dependiente.

Inicialmente se evaluaron las variables independientes de forma individual mostrando un comportamiento de residuales que representan la diferencia entre el valor real y el valor ajustado por la regresión representado en la figura 10, 11, y 12.

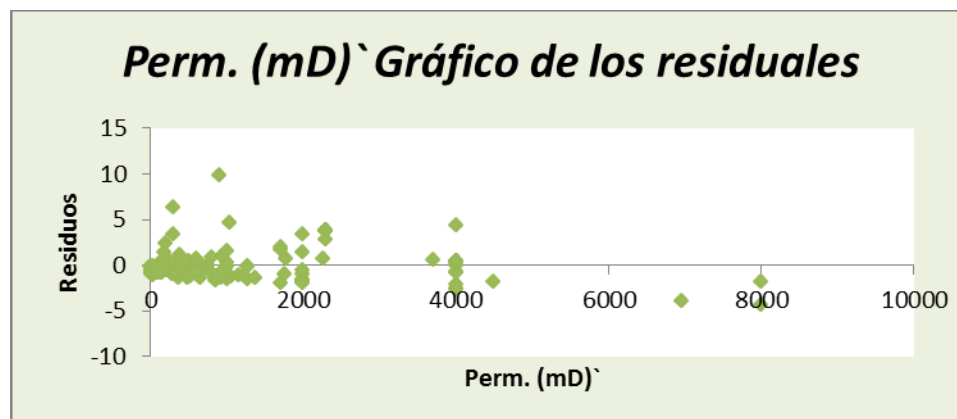
Figura 10 Grafico de residuales temperatura de formación.



Nota: resultados de los residuales de temperatura de formación

La figura 10 muestra el comportamiento residual de la variable temperatura de formación, observando valores asimétricos entre 100 y 150 °F que se encuentran fuera de los valores mínimos de temperatura de formación establecidos en la tabla 4 Variables críticas, esto se debe a la ubicación de los pozos en la subunidad A2 de la formación carbonera.

Figura 11 Grafico de residuales permeabilidad.

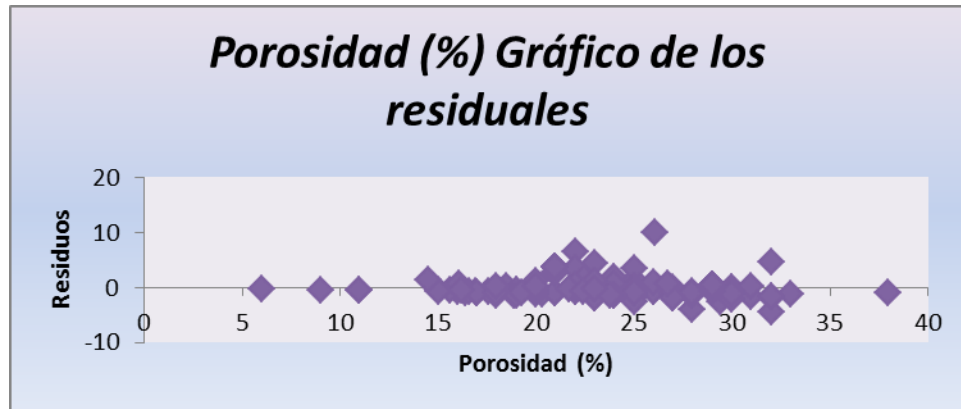


Nota: Resultados de los residuales de la variable permeabilidad

La figura 11 muestra el comportamiento residual de la variable permeabilidad, se observan valores atípicos cercanos a 4000MD, este valor se encuentra fuera del rango

mínimo promedio de la permeabilidad establecida en la tabla 4 Variables críticas, estos valores se justifican por la ubicación de los pozos en la subunidad arenosa carbonera C5.

*Figura 12 Grafico de residuales porosidad.*

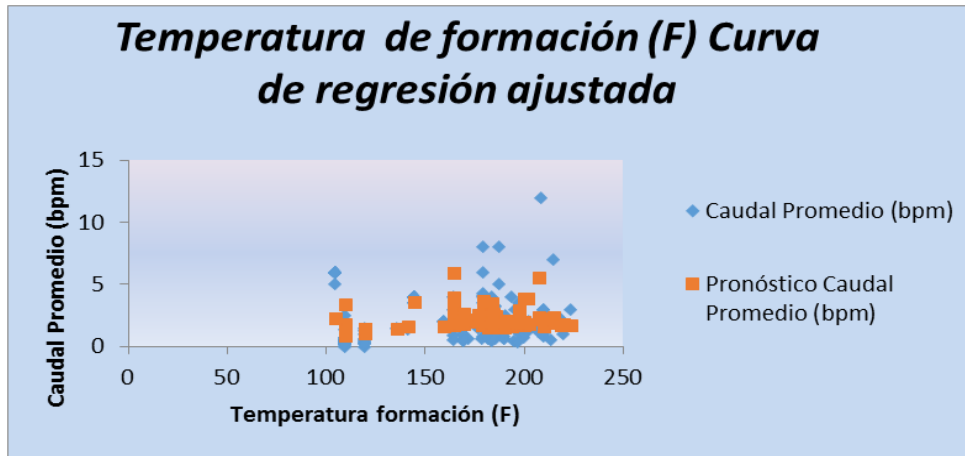


*Nota: Resultados de los residuales de la variable porosidad*

La figura 12 muestra el comportamiento residual de la variable porosidad, es una de las variables con más residuales por lo cual se debe tener en cuenta que sus valores son asimétricos y se encuentran entre 6% y 12% y corresponden a los mismos pozos con valores atípicos en la figura 10 ubicados en la subunidad A2 de la formación carbonera, esto puede representar un futuro descarte de esta variable dado que inicialmente no cumple con las premisas del modelo.

En general las variables temperatura de formación, permeabilidad y porosidad muestran un comportamiento de residuales poco asimétrico y con pocos valores atípicos exceptuando la variable porosidad, a partir de esto se procede a realizar el Cluster Analysis evaluando cada una de las variables independientes respecto a la variable dependiente Caudal Promedio de producción mediante curvas de regresión ajustada.

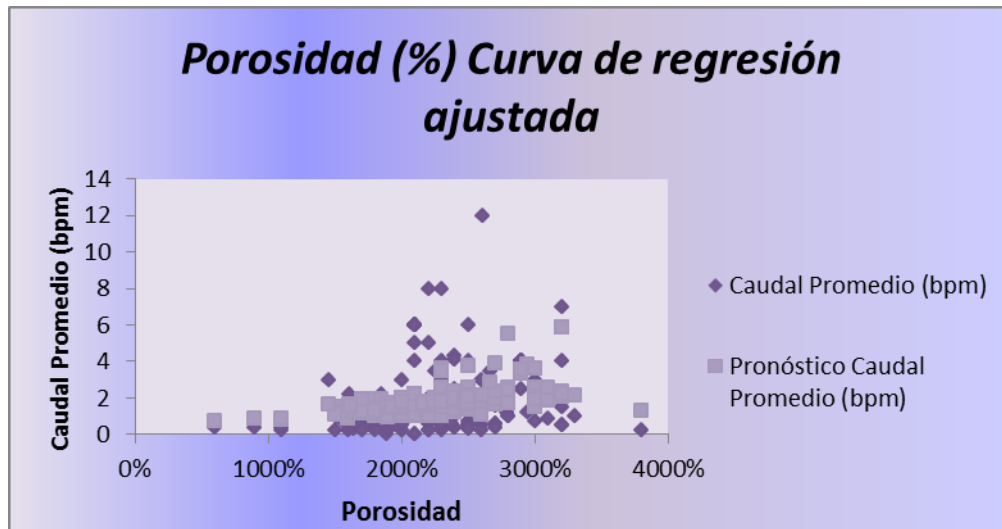
Figura 13 Curva de regresión ajustada de la variable temperatura.



Nota: Resultados de regresión ajustada de la temperatura

La figura 13 demuestra la relación entre los valores del eje X Temperatura de formación y el eje Y Caudal promedio, fuera de la línea promedio se encuentran los valores atípicos que no representan el comportamiento de las variables entre sí.

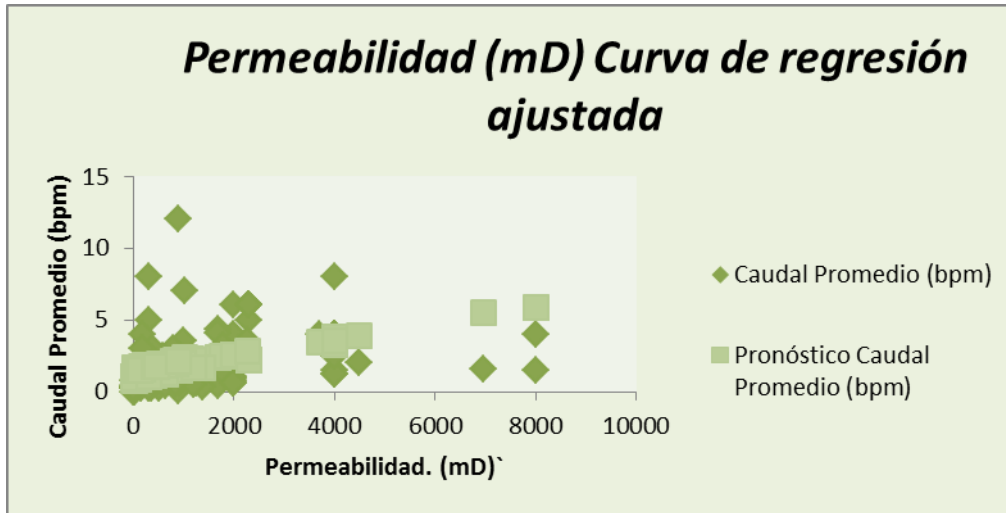
Figura 14 Curva de regresión ajustada de la variable POROSIDAD.



Nota: Resultados la curva de regresión de la porosidad

La figura 14 demuestra la relación de las variables porosidad y caudal promedio, de la cual se esperaba una mayor densidad de valores atípicos dados los valores mostrados en la figura 11.

*Figura 15 Curva de regresión ajustada de la variable permeabilidad.*



*Nota: Resultados de curva de regresión ajustada de permeabilidad*

La figura 15 demuestra la relación de las variables permeabilidad y caudal promedio, dada la figura 10 se esperaba un comportamiento lineal que permitiera ajustar sus datos pero se evidencia la dispersión por los valores entre 4000 y 8000 MD.

Las variables evaluadas entre si demuestran un comportamiento de “conglomerados”, la agrupación de los valores de cada una representa una similitud de comportamiento entre sí, lo que permite la homogeneización de la información a utilizar, en conclusión la información a trabajar es óptima para el desarrollo del modelo.

A partir de esto se realizaron los cálculos de probabilidad normal obteniendo los siguientes resultados:



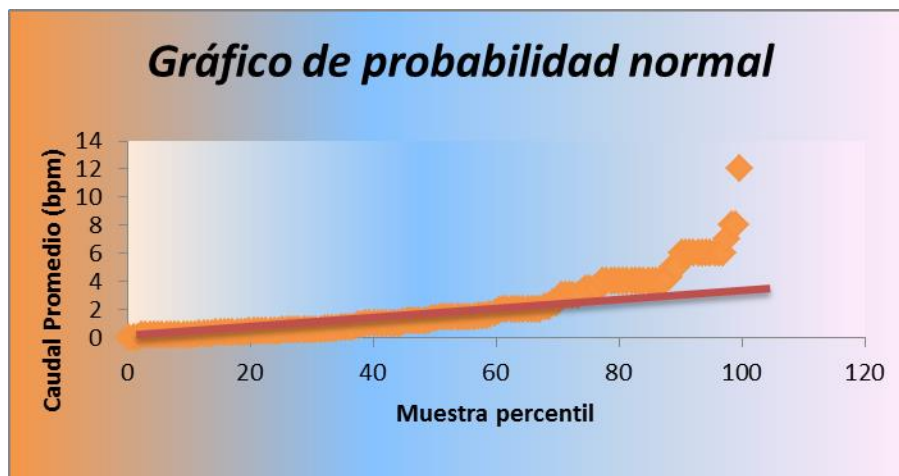
Tabla 5 Percentil de regresión.

<b>Percentil</b>	<b>Caudal Promedio (bpm)</b>
0,33	0,00
1,01	0,00
1,60	0,05
2,36	0,24
3,04	0,24

Nota: La presente tabla muestra el percentil del caudal promedio

La tabla 5 representa la estadística descriptiva clave para el proceso de análisis exploratorio de datos, la muestra percentil para cada valor de caudal representa su comportamiento en función de la muestra, es decir el caudal de cada pozo teniendo en cuenta los valores de temperatura de formación, permeabilidad y porosidad relacionado con los demás en un orden de menor a mayor.

Figura 16 grafico de probabilidad normal.



Nota: Resultados de caudal promedio respecto a percentil

Dada la densidad de la muestra se muestra un comportamiento lineal en la figura 16, la mayor cantidad de puntos se encuentran dentro de la distribución y no muestran sesgo

significativo, es decir la distribución normal se ajusta de forma conveniente a los datos, se representa de forma numérica en la tabla 6

Es ideal que los valores percentiles se comporten de manera lineal ascendente provocados por un sobreajuste, estos valores se tuvieron en cuenta para realizar la estadística de regresión mostrada a continuación:

Tabla 6 estadísticas de la regresión.

<b>Estadísticas de la regresión</b>	
Coeficiente de correlación múltiple	0,77
Coeficiente de determinación R <sup>2</sup>	0,59
R <sup>2</sup> ajustado	0,58
Error típico	1,89
Observaciones	148

*Nota: La presente tabla muestra los resultados de la estadística de la regresión*

Como se observa en la figura 16 un porcentaje significativo de datos se encuentra fuera de la línea de distribución normal (línea roja), por lo cual se obtiene un coeficiente de determinación del 59%, es decir más de la mitad del modelo explica la variabilidad de la muestra en torno a su media, se debe tener en cuenta que este porcentaje no indica si este modelo de regresión es correcto.

Se establece un coeficiente de correlación de variables o correlación múltiple de 0,77 y un error típico RMSE de 1,89 que se debe tener en cuenta para establecer los valores correctos de un resultado siento este el límite superior e inferior de la respuesta.

Tabla 7 Análisis de varianza

	<b>Grados de libertad</b>	<b>Suma de cuadrados</b>	<b>Promedio de los cuadrados</b>	<b>F</b>	<b>Valor crítico de F</b>
Regresión	3	762,32	254,10	70,72	3,59E-28
Residuos	145	520,97	3,59	-	-
Total	148	1283,30	-	-	-

*Nota: La presente tabla muestra los resultados del análisis de varianza*

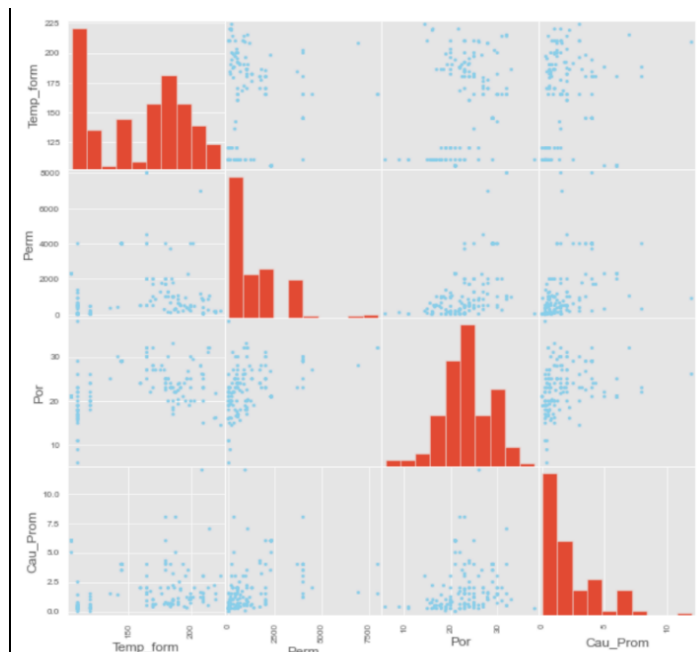
Como resultado de la regresión realizada el valor crítico de F obtenido de la tabla 7 demuestra que solo son necesarios tres grados de libertad para la estimación de cada muestra representada en la predicción. Un valor crítico de F mayor o igual a 1 representa la dispersión alta entre los datos.

### 3.2. Entrenamiento del modelo predictivo (70% de información)

Para entrar en la fase de entrenamiento se empleó el 70% de la información que corresponde a 123 pozos, esto con el fin de que el modelo reconozca las variables anteriormente mencionadas, teniendo en cuenta que el 82% de los datos utilizados corresponden a trabajos realizados con ZF y se procede a cargar la base de datos filtrada.

Se inicia con la regresión lineal simple de múltiple intercepto arrojando la matriz de correlación, así como los gráficos de dispersión que no muestran algún tipo de relación lineal entre las variables.

Figura 17 Grafico de dispersión de variables



Nota: Dispersión de variables, tomada de Python

En la figura 17 se evidencia el comportamiento de las variables evaluadas entre si por lo cual la correlación de una variable entre si misma siempre será uno (1) y no presentara dispersión, si la correlación de las variables entre si es muy cercana a cero menor relación existirá entre si.

*Figura 18 Dispersión de variables*

	Temp_form	Perm	Por	Cau_Prom
Temp_form	1.000000	-0.000705	0.290233	0.045692
Perm	-0.000705	1.000000	0.504541	0.427616
Por	0.290233	0.504541	1.000000	0.203874
Cau_Prom	0.045692	0.427616	0.203874	1.000000

*NOTA: Resultado de la dispersión de variables*

Se determina pues a partir de la figura 18 que la variable que mejor se correlaciona con las demás es la permeabilidad, teniendo en cuenta que su correlación con las variables porosidad y Caudal promedio son las más cercanas a uno 1.

La regresión realizada en el modelo demuestra que la variable más significativa es la permeabilidad, lo cual se ajusta a lo mostrado en la matriz de correlación Figura 19 esta se definió a partir del parámetro  $P > |t|$  cuyo valor debe ser cero

Figura 19 Matriz de correlación

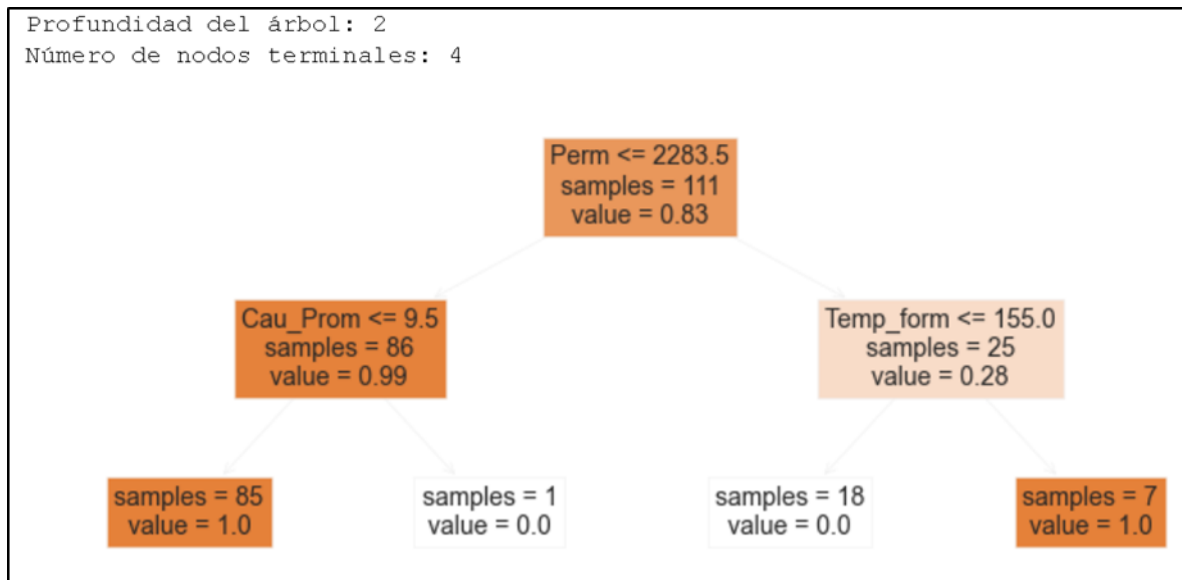
OLS Regression Results						
Dep. Variable:	Cau_Prom	R-squared (uncentered):	0.594			
Model:	OLS	Adj. R-squared (uncentered):	0.586			
Method:	Least Squares	F-statistic:	70.72			
Date:	Sun, 06 Dec 2020	Prob (F-statistic):	3.09e-28			
Time:	00:45:46	Log-Likelihood:	-303.13			
No. Observations:	148	AIC:	612.3			
Df Residuals:	145	BIC:	621.3			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Temp_form	0.0059	0.004	1.547	0.124	-0.002	0.013
Perm	0.0005	0.000	4.871	0.000	0.000	0.001
Por	0.0157	0.029	0.543	0.588	-0.041	0.073
Omnibus:	68.621	Durbin-Watson:	1.109			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	244.706			
Skew:	1.772	Prob(JB):	7.29e-54			
Kurtosis:	8.209	Cond. No.	397.			
Warnings:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

*Nota Matriz de correlación, tomada de Python*

Se plantearon tres modelos de arbol de decision con 2, 4 y 5 ramas, de lo cual se identifico una variacion significativa en la correlacion de las variables entre si y asi mismo la variacion en el porcentaje de confiabilidad.

Para el desarrollo de este modelo se identifico que al relacionarse las variables teorica y experimentalmente, a medida que el arbol aumenta su numero de ramas, su porcentaje de confiabilidad disminuye, antes de determinar estos tres modelos de arbol de descicion de determino que un arbol de una rama no posee porcentaje de confiabilidad dado que no correlaciona todas las variables, esta es la razon por la cual se inicia en el planteamiento de un arbol de decision de 2 ramas que arrojó los siguientes resultados.

Figura 20 Árbol de clasificación de 2 ramas.



Nota: Resultado árbol de clasificación tomada de python

Porcentaje de confiabilidad: 100%

Se observa en la figura 20 la jerarquía de las variables planteadas en la regresión, este árbol está constituido por 2 ramas principales y 4 nodos terminales que permite relacionar todas las variables disponibles con el número de muestras arrojando un porcentaje de confiabilidad de 100%

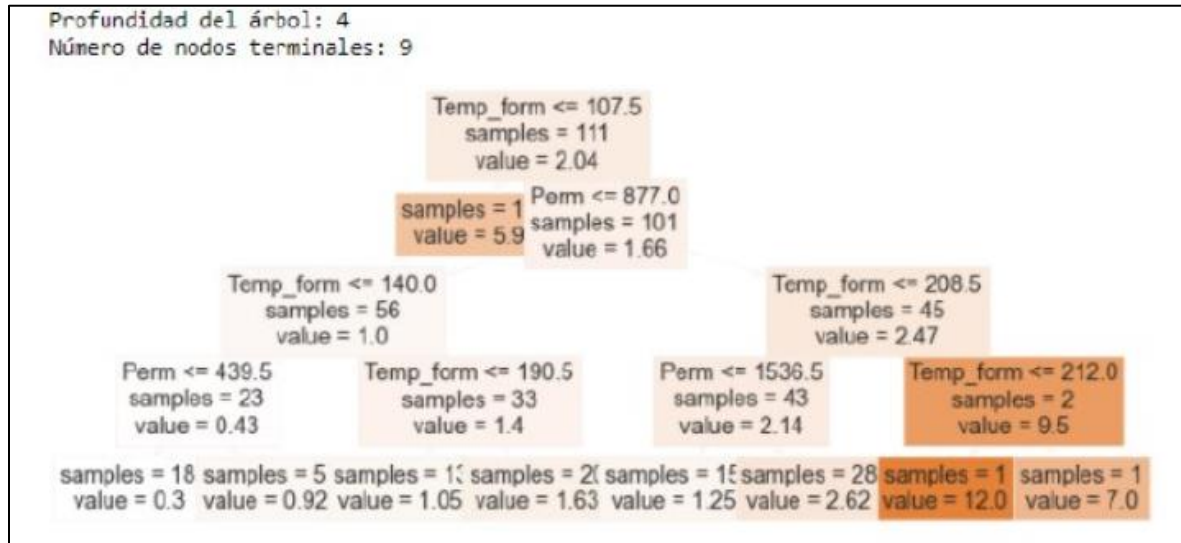
Figura 21 importancia de variables

	predictor	importancia
1	Perm	0.617191
0	Temp_form	0.320046
3	Cau_Prom	0.062763
2	Por	0.000000

Nota: Resultado de importancia de las variables tomada de python

Como resultado se obtiene que la variable de mayor importancia es la permeabilidad, esto como consecuencia de los valores arrojados por la matriz de correlación mostrada en la figura 18.

Figura 22 Árbol de clasificación de 4 ramas.



Nota: Resultado árbol de clasificación tomada de python

Porcentaje de confiabilidad: 32%

Se observa en la figura 22 el cambio en la jerarquía de las variables planteadas, en este caso el tallo del árbol se constituye por la variable temperatura de formación, este árbol está constituido por 4 ramas y 9 nodos terminales, se observa que dado el aumento de nodos las variables ya no se relacionan de la manera correcta.

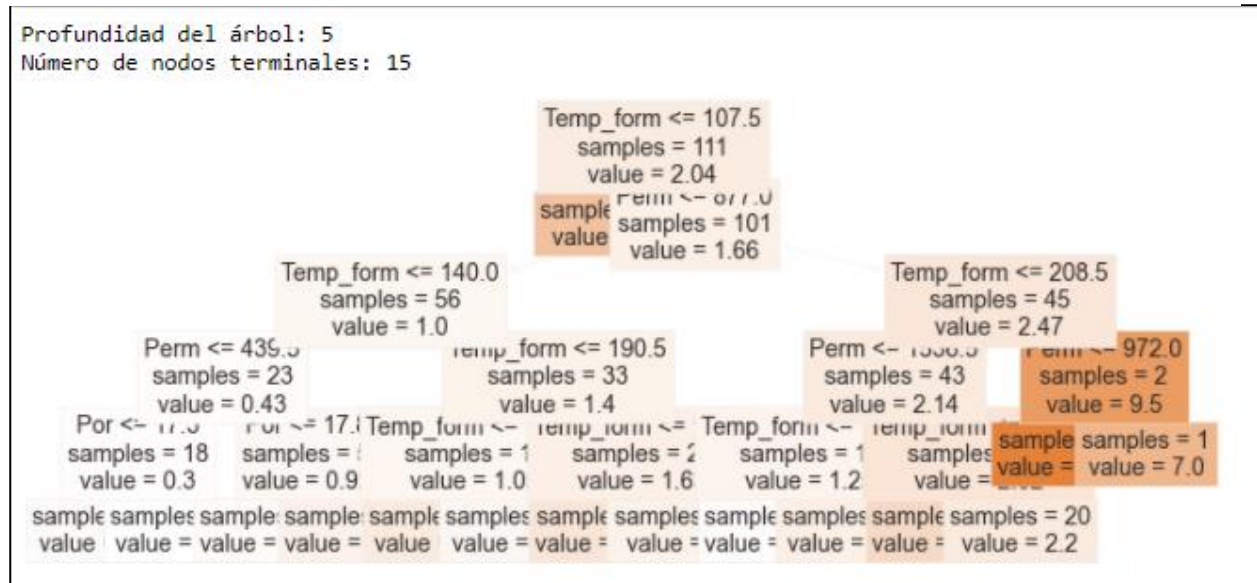
Figura 23 Importancia de variables

	predictor	importancia
0	Temp_form	0.800463
1	Perm	0.199537
2	Por	0.000000

Nota: Resultado de importancia de las variables tomada de python

En el anterior árbol del problema se evidencia un cambio de variable principal, en este caso es temperatura de formación, esta variable aporta un valor de importancia del 80% similar al árbol de decisión de la figura 20, como consecuencia de esto el porcentaje de confiabilidad disminuye al 32%.

Figura 24 Árbol de clasificación de 5 ramas.



Nota: Resultado árbol de clasificación tomada de python

Porcentaje de confiabilidad: 20%

Se observa en la figura 24 una estabilidad en la jerarquía de las variables planteadas, en este caso el tallo del árbol se mantiene en la variable temperatura de formación, este árbol está constituido por 5 ramas y 15 nodos terminales, se observa nuevamente que dado el aumento de nodos las variables ya no se relacionan de la manera correcta y afecta significativamente el porcentaje de confiabilidad estableciéndolo en 20%.



Figura 25 importancia de variables

	predictor	importancia
0	Temp_form	0.778250
1	Perm	0.218815
2	Por	0.002935

Nota: Resultado de importancia de las variables tomada de python

Dado el resultado del nivel de confiabilidad del árbol de decisión de 2 ramas se decide descartar los otros dos, de esto se obtienen cuatro variables de importancia, permeabilidad con 62%, temperatura de formación con 32%, caudal promedio con 6% respectivamente y la variable porosidad no aporta ningún porcentaje de influencia, por lo tanto no se tendrá en cuenta.

### 3.3 Calibración del modelo predictivo (30% de información)

A partir de esto se inició la fase de prueba con el 30% de la información correspondiente a 36 pozos, esto con el fin de calibrar el modelo. Se obtuvieron las siguientes predicciones:

Tabla 8 Predicción del entrenamiento

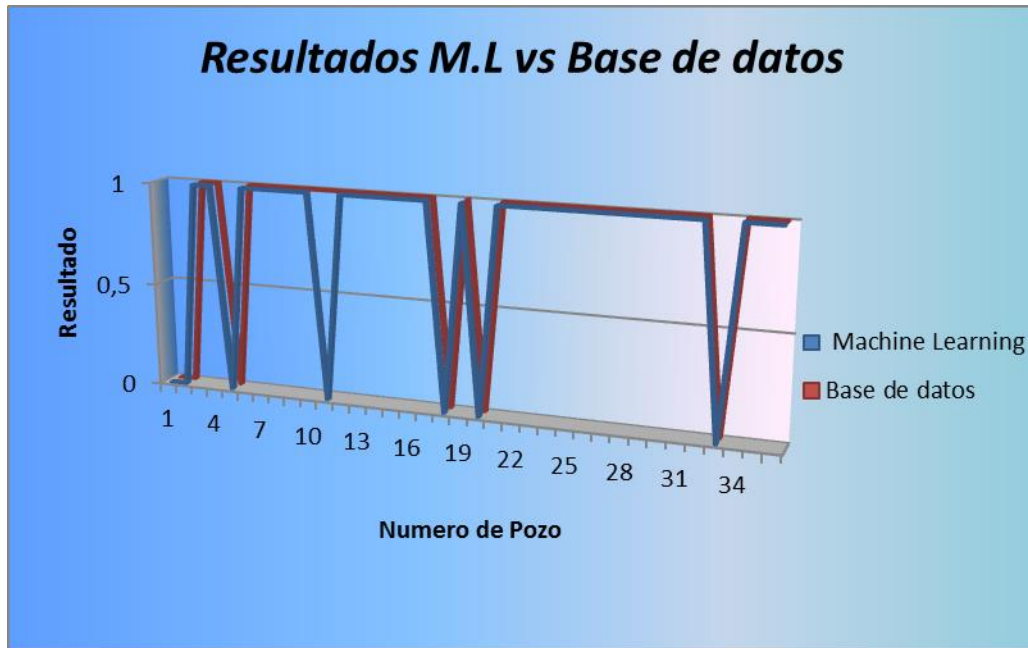
Pozo	Resultado	Variable de resultado
1	0	HRWP
2	0	HRWP
3	1	ZF
4	1	ZF
5	0	HRWP
6	1	ZF
7	1	ZF
8	1	ZF
9	1	ZF

10	1	ZF
11	0	HRWP
12	1	ZF
13	1	ZF
14	1	ZF
15	1	ZF
16	1	ZF
17	1	ZF
18	0	HRWP
19	1	ZF
20	0	HRWP
21	1	ZF
22	1	ZF
23	1	ZF
24	1	ZF
25	1	ZF
26	1	ZF
27	1	ZF
28	1	ZF
29	1	ZF
30	1	ZF
31	1	ZF
32	1	ZF
33	0	HRWP
34	1	ZF
35	1	ZF
36	1	ZF

*Nota: En la presente tabla se evidencian los resultados del ML*

Los resultados de la predicción realizada por el Machine Learning con el 30% de la información muestran una diferencia respecto a la base de datos con el pozo 11 representado en la figura 26, representando un valor atípico que puede ser consecuencia de la ausencia de datos en el gradiente de fractura, dado que es indispensable para la implementación de la herramienta HRWP, a partir de esto se determina que los datos fueron clasificados de manera correcta.

Figura 26 Resultados M.L vs Base de datos



*Nota: Resultado del programa ML y base de datos suministrada*

El score de entrenamiento del modelo o  $R^2$  es de 1 o confiabilidad del 100% y error cuadrático medio (rmse) de 1.85, mostrando un ajuste perfecto en el modelo, esto se debe a la baja cantidad de muestras de HRWP, provocando overfitting o sobreajuste, esto se soluciona tomando más muestras para HRWP o balanceando las muestras.

### 3.4. Evaluación económica

Se procede a realizar la predicción para los cuatro pozos establecidos al inicio de esta investigación, cuyos nombres fueron modificados por acuerdos de confidencialidad y corresponden a : X1, X2, X3 y X4 y de los cuales se conoce la siguiente información:

Tabla 9 información de pozos de prueba.

<b>Pozo</b>	<b>T de formación °F</b>	<b>Permeabilidad MD</b>	<b>Porosidad %</b>	<b>Q prom BPM</b>
X1	224	184	14,5	3
X2	220	90	16,2	1
X3	200	4,2	23	1,5
X4	184	-	24	0,9

*Nota: En la presente tabla se muestra la información de los pozos de prueba*

Se debe tener en cuenta que el pozo X4 no posee información para la variable permeabilidad, por lo cual los resultados pueden variar.

Tabla 10 Predicción para selección de pozos

<b>Nombre del pozo</b>	<b>Método de control de arena en base de dato</b>	<b>Resultado en ML</b>	<b>Variable resultado</b>
X1	Zeta Flow	1	Zeta Flow
X2	Zeta Flow	1	Zeta Flow
X3	Zeta Flow	1	Zeta Flow
X4	HRWP	1	Zeta Flow

*Nota: La presente tabla muestra la predicción para los 4 pozos de selección*

En la tabla 9 se evidencia los resultados arrojados por la predicción del Machine Learning y las operaciones realizadas por la empresa X en la selección de pozos.

Figura 26 ejemplo de predicción FUENTE: Tomado de Python

```
In [148]: modelo.predict(X = X_test.loc[["Chipiron 09"]])
Out[148]: array([1.])
```

Nota: modo de predicción

La figura 27 representa la forma en que el programador arroja los resultados para un pozo seleccionado.

De acuerdo a la predicción realizada con el modelo diseñado, los resultados indican que la empresa X ha trabajado el 75% de los pozos con el método de control de arena más adecuado, esto obedece a los valores estimados en la metodología con un rango entre el 70% y 80% de éxito operacional.

El resultado arrojado por el Machine Learning respeta los rangos de aplicación de ZF establecidos en la metodología, operacionalmente la variable más importante para la aplicación de este método de control de arena en un pozo es la permeabilidad, dado que se deben tener en cuenta estos valores para los porcentajes de adición del químico SandAid, se concluye que por omisión de variables el modelo evalúa las demás las variables disponibles que se encuentran dentro de los rangos de aplicación para Zeta Flow y se evidencia el correcto entrenamiento del modelo.

### 3.4.1. Costos por inversión de servicio.

El costo de inversión inicial por métodos de control de arena (HRWP) correspondiente al pozo X4 para 8ft de tratamiento e implementando aditivos como salmuera, geles lineales ligeros y fluido apuntalante fue de:

Tabla 11 Costo de tratamiento HRWP

COSTO HRWP (USD)			
POZO	PERSONAL + EQUIPOS	ADITIVOS	TOTAL (USD)
X4	\$ 100.000,00	\$ 22.000,00	\$ 122.000,00

Nota: La presente tabla muestra los costos de tratamiento para HRWP

Una vez implementado el método de control de arena Zeta flow para el pozo X4 el costo de inversión sin adición de tratamientos y/o químicos adicionales generado para 8ft de tratamiento y aditivo SandAid fue de:

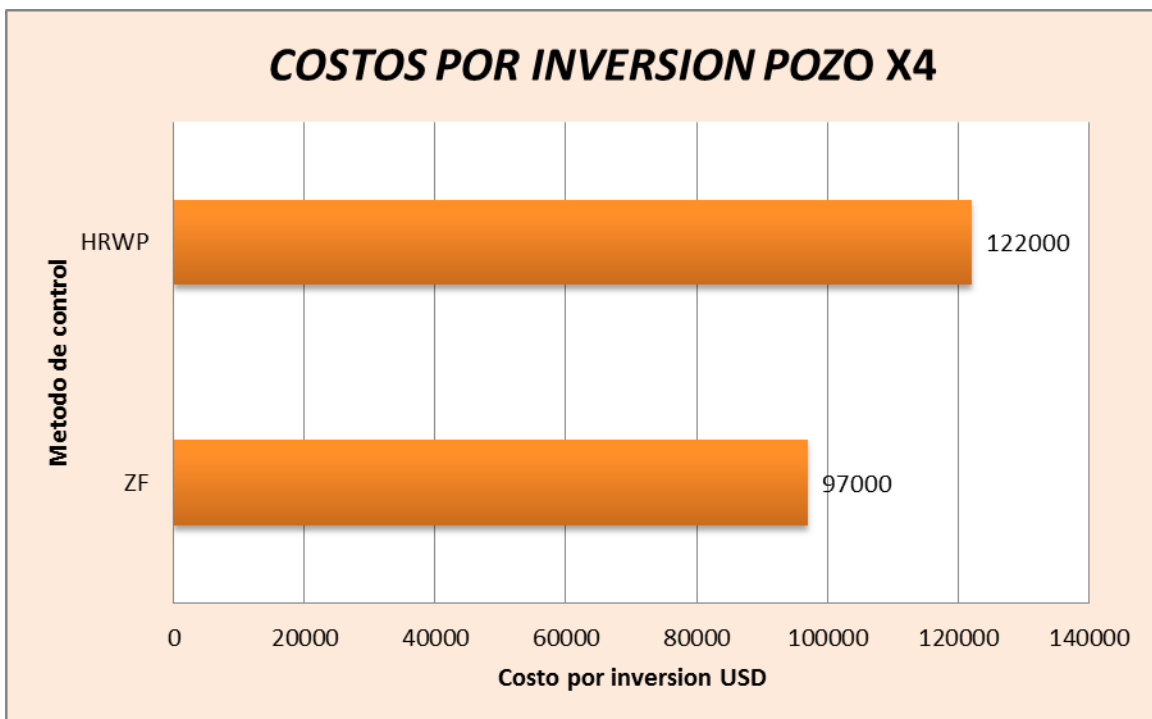
*Tabla 12 Costo de tratamiento ZF*

<b>COSTO ZETA FLOW (USD)</b>			
<b>POZO</b>	<b>PERSONAL + EQUIPOS</b>	<b>ADITIVOS</b>	<b>TOTAL (USD)</b>
X4	\$ 65.000,00	\$ 32.000,00	\$ 97.000,00

*Nota: La presente tabla muestra los costos para el método de control de arena ZF*

Al cambiar el método de control de arena se evidencia una reducción por costos de inversión del 20% que corresponden a 25.000 USD. Figura 28

*Figura 27 Costos por inversión de operación.*

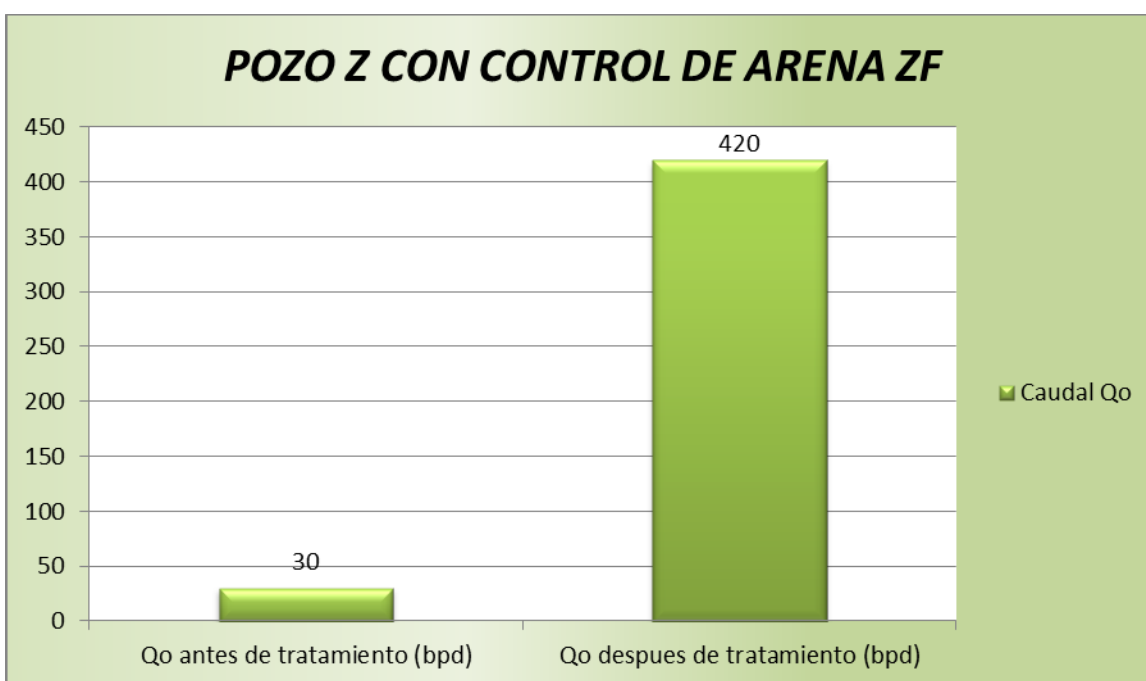


Nota: Comparativo del costo de operación entre ZF y HRWP

### 3.4.2. Producción alcanzada

Una vez aplicado el método de producción de arena sugerido por el machine Learning se evidenció un cambio en el %BSW cuyo valor inicial correspondía al 85% y cuyo valor final se desconoce por acuerdos de confidencialidad.

Figura 28 Producción alcanzada en el pozo seleccionado



Nota: Resultado de la mejora de caudal

El Cambio de tratamiento cumplió con su objetivo principal que es aumentar el Caudal de producción que cambio de 30 BPD a 420 BPD mostrado en la figura 19, al no poseer la información de %BSW final, se concluye que dada la teoría las variables  $Q_o$  y %BSW son inversamente proporcionales, es decir hasta la fecha el %BSW disminuyó.

### 3.4.3. Eficiencia de la inversión

Como se mencionó anteriormente el costo base para la operación fue de 97.000USD, se debe tener en cuenta la variación del barril de petróleo referencial Brent desde el inicio de la operación (septiembre de 2020) y para evaluar la eficiencia de la inversión se tendrá en cuenta solo el primer mes.

- Barril de petróleo referencia Brent para apertura del mes de septiembre 2020: **40,85 USD**
- Barril de petróleo referencia Brent para cierre del mes de septiembre 2020: **45,78 USD**
- Valor promedio de petróleo referencia Brent para el mes de septiembre 2020: **43,31 USD**

*Tabla 13 Ingreso estimado.*

<b>Precio Petróleo Brent (USD)</b>	<b>Qo (BPD)</b>	<b>Días de operación</b>	<b>Ingreso total (USD)</b>
\$ 43,31	420	30	\$ 545.706,00

*Nota: La presente tabla representa los ingresos estimados*

*Tabla 14 Ganancia estimada.*

<b>Ingreso total (USD)</b>	<b>Costos de operación (USD)</b>	<b>Ganancias (USD)</b>
\$545.706,00	\$97.000	\$448.706,00

*Nota: La presente tabla representa las ganancias estimadas*

#### Calculo de eficiencia de inversión:

*Ecuación 1 Cálculo de eficiencia de inversión*

$$ROI (\%) = (\text{Beneficio neto} / \text{Inversión}) \times 100$$



Donde:

**ROI(%):** *Eficiencia de inversión*

**Beneficio neto:** *ganancias*

Entonces:

$$\text{ROI (\%)} = (448.706 / 97.000) \times 100$$

$$\text{ROI (\%)} = 462,5835052\%$$

La eficiencia de la inversión de la operación corresponde a un 462% sin tener en cuenta otros costos operacionales como equipos de producción, personal, entre otros. Se determina que la inversión de un método de control de arena adecuado en un pozo representa un beneficio económico significativo relacionándolo directamente con el caudal de producción diario, lo cual es una propuesta interesante a la hora de adquirir un servicio de operaciones para el control de arena ofrecido por la empresa X.

## CONCLUSIONES

A partir de la implementación del modelo Machine Learning en los 4 pozos de evaluación, se determinó que para los pozos X1, X2, X3 y X4 el mejor mecanismo de control de arena es Zeta flow.

Se evidencio que el cambio de mecanismo de control de arena en el pozo X4 cumplió con el objetivo principal presentando un aumento en el caudal de producción que cambio de 30 BPD a 420 BPD, así mismo una reducción de costos del 20% correspondiente a 25.000 USD, el %BSW inicial corresponde al 85%, se determinó a partir de la teoría que son dos variables inversamente proporcionales, dados los resultados de caudal de producción se determinó que dado el aumento significativo de producción el %BSW disminuyó considerablemente pero no fue posible cuantificarlo.

A partir del análisis exploratorio de datos (Cluster Analysis) se evidenció que las variables %BSW y Gravedad API presentan una incertidumbre de datos del 80% y 87% respectivamente, lo que impide la correcta implementación para el diseño del modelo predictivo.

Mediante el diseño de tres (3) árboles de decisión se determinó un modelo de confiabilidad del 100% teórico y 78% real constituido por dos ramas y 4 nodos, del cual la variable permeabilidad representa el 62% de influencia, temperatura de formación 32%, caudal promedio 6% y porosidad 0%. Este árbol de decisión permite relacionar de manera correcta las variables involucradas sin disminuir considerablemente el score de entrenamiento.

A partir de la implementación del modelo machine Learning se evidenció que el el 75% las operaciones realizadas por la empresa X en el campo X de los llanos orientales se han realizado con el método de control de arena más adecuado y eficiente, es importante resaltar que el comparativo de resultados entre el machine Learning y la base de datos realizado a partir del entrenamiento con el 30% de la información arrojo 1 de 36 de resultados diferentes, esto se debe a que es un pozo nuevo y a la ausencia

de información en la variable gradiente de fractura que se tuvo en cuenta para el desarrollo de la base de datos.

Los resultados encontrados dejan una ventana por estudiar que es incluir la variable de producción de crudo y agua total antes y después de las operaciones, así como como el respectivo seguimiento (información) a operaciones de mantenimiento post trabajo de sand control.

## BIBLIOGRAFÍA

- [1] R. Letterman, Coagulation & Flocculation, Zeta-Meter, Zeta-Meter, 1993.
- [2] W. Col, «Evaluacion del tratamiento del SandAid y analisis granulometrico,» Weatherford , 2013.
- [3] G. R. P., Matematicas distreta y combinatoria, Pearson, 1998.
- [4] D. M. M. Remotti, Intr oducción a lasNormas enEspañol, Mexico : Foro lineamientos de medicion, 2012.
- [5] Leslie Bellorin, WTC Specialist, «Curso introductorio al control de arena,» Ecopetro S.A, Colombia, 2018.
- [6] Weatherford, «Zeta Flow Overview,» Colombia , 2018.
- [7] E. Alpaydin, Introduction to machine Learning, Massachusetts: The MIT press, 2004.
- [8] C. McPhee, «Symposium and Exhibition on Formation Damace Control,» de *OMV Pakistan* , Louisiana, 2004.
- [9] D. D. C. K. A. K. C.A.M Veeken, «Sand Production Prediction Review: Developing an Integrated Approach,» *SPE*, pp. 3-6, 9 Oct 1991.
- [10] P. S. Rovina, «Sand Control in Open Hole Horizontal Wells-Case Histoies in Developments,» *Petrobras-Halliburton*, pp. 2-6, 12 April 1999.
- [11] J. D. W. William K Ott, World oil modern sandface completion practices handbook, Houston, TX: Gulf Pub. Co, 2005.
- [12] R. S. L. V. K. S. B. G. J. S. DS Treybig, «Consolidating Sand with New Generation Zeta Potential Altering Systems,» *Society of Petroleum Engineers*, vol. 1, nº SPE-179022-MS, pp. 11-14, 2016.

- [13] D. S. J. L. G. F. Z. S. K. E. F. J. David Johnson, «Enhancing Gas and Oil Production With Zeta Potential Altering System,» *Society of Petroleum Engineers*, nº 2-3, 2010.
- [14] W. Col, «Sand control. Gravel Pack Systems,» Weatherford , 2012.
- [15] W. Col, «Evaluacion del bombeo de la grava,» Weatherford , 2013.
- [16] W. Col, «Documents Zeta Flow@Sand-Conglomeration services».
- [17] M. A. Garcia Isabel, Introduccion a la programacion con Pyton, Universitat Jaume: sapientia, 2003.
- [18] J. Mojica, «Cuencas Catatumbo, Cesar – Ranchería, Cordillera,» *Subdireccion tecnica ANH*, p. 31, 2009.
- [19] Weatherford, «Conglomeración de arena SandAid: mayores índices de flujo sin arena, menor producción de agua,» *Egypth Oil&Gas*, p. 1, 2014.
- [20] J. P. Lorduy, «Caracterizacion petrofisica de un area en el bloque CPO 16 en la cuenca de los llanos orientales,» Universidad EAFIT, facultad de Geologia , Medellin , 2016.
- [21] U. t. d. prospeccion, *Evaluacion de las cuencas y estructuracion de escenarios de oferta de hidrocarburos convencionales y no convencionales*, Bogota : UPME, 2018.
- [22] ANH, «Cuenca llanos orientales-Integracion geologica de la digitalizacion y analisis de nucleos,» Agencia nacional de hidrocarburos, 2012.
- [23] J. R. J. D. P. E. R. D. O. M. C. Pablo Leyva, Diagnostico ambiental y lineamientos para el uso sostenible del area Caño Limon - Estero de Lipa, Bogotá: IDEAM, 2000.

## ANEXOS

### RECOMENDACIONES

Si se desea un modelo de predicción machine Learning con mayor porcentaje de confiabilidad se recomienda enfatizar en la recolección de información especialmente de los pozos ubicados en el campo X lo que también influye en la dificultad de la predicción del comportamiento del campo en producción, cabe resaltar que el diseño del modelo permite incluir más variables como por ejemplo otros métodos de control de arena o los métodos de control de arena trabajados con combinaciones mecánicas como Gravel Pack, Frack pack o adiciones químicas con ácidos, orgánicos o WecFlow.

Tener en cuenta la necesidad y características de cada pozo para la implementación y selección del método de control de arenas y finos más adecuado debido a que se usan métodos que no cumplen las expectativas de la empresa y el cliente final afectando la variable costo – beneficio.

En lo posible solicitar los datos de caudales de producción antes y después de la operación a la empresa a cargo, esto permitirá dar un mayor sustento financiero al éxito de las predicciones arrojadas por el Machine Learning,

Si en algún momento el modelo de predicción establecido logra alcanzar un porcentaje de confiabilidad mayor o igual al 95% real se recomienda tener en cuenta el criterio humano para ratificar las decisiones tomadas debido a que se deben evaluar los criterios físicos del yacimiento, estado mecánico, parámetros de inyección, bombeo y el daño de formación, dado que se deben tomar decisiones relacionadas directamente con la ingeniería que no se asocian al modelo de predicción machine Learning,