

**Propuesta de política operacional de drenaje y despacho de tanques en una terminal  
petrolera usando simulación de Monte Carlo y machine learning**

**Antony Jesus Berdugo Morales**

**Universidad de La Sabana**

**Maestría de gerencia en ingeniería**

**Chía, 2024**

## Tabla de contenido

Resumen ejecutivo.....	3
Palabras clave .....	3
Introducción.....	3
Justificación.....	4
Objetivos .....	8
Marco teórico.....	8
Metodología.....	12
Desarrollo metodológico .....	13
a) Caracterización de política operacional.....	13
b) Simulación de política operacional .....	19
c) Evaluación y selección de nueva política operacional .....	32
Definición de nuevas políticas. ....	32
Evaluación de nuevas políticas. ....	33
Nueva política operacional seleccionada.....	37
Conclusiones .....	45
Investigación futura .....	46
Bibliografía .....	48
Apéndice A.....	51
Apéndice B. ....	60
Apéndice C.....	61
Apéndice D.....	62
Apéndice E. ....	74

## **Propuesta de política operacional de drenaje y despacho de tanques en una terminal petrolera usando simulación de Monte Carlo y machine learning**

### **Resumen ejecutivo**

Importante empresa de transporte de petróleo en Colombia en su mapa estratégico define como objetivo operacional estratégico el generar ahorros de 30 millones de dólares. Debido a que el agua presente en los sistemas de la compañía afecta el core del negocio (el transporte de petróleo), en el presente proyecto se aborda la operación de despacho y drenaje en una terminal petrolera buscando alinearla con el objetivo y mapa estratégico. Para ello se realiza un diagnóstico de la operación actual, se simula la operación usando Monte Carlo apoyado con una herramienta de predicción desarrollada usando machine learning, se proponen alternativas de políticas operacionales y se evalúan para finalmente seleccionar la que genera mayor ahorro (0,74 MUSD en VPN) (mejorando el uso de los recursos disponibles en la operación de drenaje).

### **Palabras clave**

Operación de drenaje, Monte Carlo, política operacional, petróleo, agua libre, Machine learning, CRISP-DM, sistema de apoyo a la decisión, Knime, Business process modeling.

### **Introducción**

En el presente trabajo de grado se busca la alineación de la operación con el mapa estratégico de una compañía de transporte de petróleo por medio de una propuesta de mejora de las políticas operacionales.

La empresa petrolera suministró información para los datos de la operación y la información estratégica se tomó del informe de sostenibilidad.

Importante compañía de transporte de petróleo realiza mapa estratégico en donde se define como objetivo estratégico operacional el ahorrar 30 millones de dólares en un horizonte de tiempo de 10 años, se observa que los objetivos estratégicos son grandes retos por lo que se hace necesario sugerir cambios que produzcan resultados importantes para alinear la operación con esta visión, proponiendo nuevas políticas operacionales.

Este trabajo se enfoca en el core del negocio, en el transporte y almacenamiento de petróleo y en uno de los puntos más importantes del negocio, la terminal petrolera de exportación, la cual tiene el almacenamiento más grande en la compañía. La operación en la terminal petrolera presenta problemas de eficiencia operativa, pues se encuentran altos niveles de agua. Por lo que el enfoque de las iniciativas es la reducción del agua en los sistemas de la terminal petrolera, específicamente en la operación de drenaje y despacho en la terminal petrolera.

El objetivo estratégico necesita de una política operacional coherente para que exista el correcto alineamiento estratégico, por ello es necesario escoger la política adecuada para lograrlo. En este trabajo se realiza un diagnóstico de la operación actual simulando su estado inicial y final previo a la exportación y la operación de exportación (incluyendo los drenajes) por medio de una simulación de Monte Carlo, para ello se crean distribuciones a partir de los datos recolectados para el estado inicial de la operación mientras que para el estado final se usa una herramienta de machine learning entrenada a partir de los datos recolectados siguiendo la metodología CRISP-DM y usando Knime; a partir del diagnóstico se definen las políticas operacionales a evaluar, entre estas se contempla una herramienta de inteligencia artificial con un sistema de apoyo a la decisión (DDS) para apoyar la operación de despacho y drenaje en la terminal petrolera, de estas nuevas políticas se selecciona la que genera mayor ahorro.

### **Justificación**

Enfocando la definición de Reich y Benbasat (Reich & Benbasat, 1996) en el ámbito operacional, el alineamiento estratégico es el nivel en el que la misión, los objetivos y los planes contenidos en la estrategia de negocio son compartidos y soportados por la estrategia operacional. El alineamiento estratégico también se ve como una relación entre la estrategia de

negocios y la infraestructura y procesos (Henderson & Venkatraman, 1993) o también puede verse como que “Todo rema en la misma dirección” según Abraham, G.A (Y. E. Chan & Reich, 2007). Este alineamiento se logra traduciendo los objetivos de unidad de negocio en objetivos personales (Campbell, 2005).

Además, el alineamiento estratégico es un proceso de cambio a través del tiempo y de continua adaptación (Henderson & Venkatraman, 1993), muchos autores lo definen como un proceso, no un estado (Henderson & Venkatraman, 1993; Rondinelli et al., 2001); donde se sugiere reajustar cuatro componentes estratégicos constantemente (Rondinelli et al., 2001): la estrategia de negocio, las decisiones de penetración de mercado, la gerencia de procesos y la estructura organizacional. Este proyecto se enfoca en la gerencia de procesos, evaluando las políticas operacionales actuales y proponiendo nuevas que encaminen la operación a apoyar los indicadores de los objetivos estratégicos.

La industria petrolera encuentra distintas amenazas como precios bajos, otras opciones de energía cada vez más competitivas y el contexto sociopolítico y ambiental, por lo que es necesario adaptarse. La empresa busca adaptarse por medio de un mapa estratégico que se transforma en objetivos estratégicos que a su vez pasan a ser realidad a través de iniciativas y proyectos para alinear las políticas actuales conforme a la visión.

Esta empresa busca adaptarse a los cambios que experimenta el mundo, observando el contexto operacional, nacional y mundial, su entorno económico, regulatorio, sociopolítico, ambiental y tecnológico, y las amenazas y oportunidades que puedan tener implicaciones en el negocio; para ello desarrolla un marco estratégico con metas a largo plazo que responden a los retos actuales en materia operativa, ambiental, social y de gobernanza. Este marco estratégico para el 2030 define tres perspectivas estratégicas relacionadas entre sí: sostenibilidad,

generación de valor con grupos de interés y adaptación organizacional. Estas perspectivas las desarrolla en 5 objetivos: utilidad operacional sostenida, impacto positivo al medio ambiente y a la sociedad, tecnología e innovación propulsoras de transformación, modelo organizacional óptimo y talento competente-apropiación de la cultura.

Como primer objetivo se tiene la utilidad operacional sostenida en donde la eficiencia y flexibilidad operativa es primordial (en las metas del marco estratégico 2030 se esperan reducciones de costos por \$30 millones de dólares), por medio de procesos mejorados en donde se espera intervenga la tecnología e innovación. Por lo anterior, se puede definir como objetivo operacional estratégico una reducción de costos en \$30 millones de dólares al 2030. La operación reportó en 2021 costos operacionales por 163,8 millones de dólares de los cuales 54,3 millones de dólares son costos variables.

El core del negocio es el transporte de petróleo, (productos M y P) el cual presenta un porcentaje de agua en suspensión. Para mejorar la eficiencia y la capacidad del oleoducto durante su operación se requiere reducir la cantidad de agua presente en el petróleo dado que no genera ningún beneficio, al contrario, genera problemas en la operación de despacho como: reclamaciones, aumentos en los costos de operación, baja calidad del producto y hasta riesgos a la seguridad.

Uno de los puntos más importantes del oleoducto es la terminal petrolera donde es entregado el petróleo a buques; punto crucial para la gestión del agua pues allí es drenada antes de ser entregada al buque. Esta operación es realizada por un grupo llamado “cuadrilla” el cual también presenta problemas operativos (falta de programación, desorden, entre otros). Por lo que esta tesis busca generar una propuesta de política de drenaje y despacho en los tanques de la terminal petrolera que este alineada con el objetivo estratégico “utilidad operacional sostenida”.

En la industria petrolera se encuentran análisis de inteligencia de negocio teniendo en cuenta la operación; como un estudio que propone la adopción de inteligencia de negocio y machine learning en toda la organización dentro de los sistemas de gestión de seguridad (Nakhla et al., 2021) y un estudio que por medio de un modelo de optimización reduce costos y riesgos de la cadena de suministro de petróleo y reduce los riesgos operativos con un ligero aumento en el costo total de la cadena de suministro (Kumar Tarej et al., 2022). Por lo que para reducir costos operativos es una buena opción el uso de machine learning y de modelos de optimización.

Para el desarrollo de la política operacional se valoran tecnologías que simulan la operación y la mejoran por medio de recomendaciones como un sistema de apoyo a la decisión (DDS). Respecto a sistemas de apoyo a la decisión (DDS) para la industria petrolera se encontró un sistema experto de diagnóstico de falla para el proceso de refinación de aceite lubricante basado en conocimiento, ayuda al operador a monitorear y diagnosticar una situación anormal (Qian et al., 2008), un sistema de apoyo a la decisión para el monitoreo y diagnóstico de procesos de separación en la producción de petróleo (C. W. Chan, 2005), una aplicación de mantenimiento predictivo, basada en datos con análisis multivariable para predecir el sistema, la falla temprana y su causa (Clavijo et al., 2019); sin embargo, ninguno de estos se enfoca en apoyar la operación de drenaje del tanque. Este trabajo de grado genera valor académico a la literatura de sistemas de apoyo a la decisión para las operaciones de la industria petrolera al aplicarlo a la operación de drenaje, además de incluir un desarrollo de machine learning.

Teniendo en cuenta lo anterior es necesario caracterizar el proceso de drenaje y despacho, entender el efecto del agua suspendida en la operación actual de la terminal petrolera y determinar que política operacional de despacho y drenaje puede reducir este efecto, por lo cual surgen las siguientes preguntas:

¿Qué tanto afecta el exceso de agua suspendida en el petróleo la operación de despacho y drenaje y su alineación a la estrategia operacional de la compañía?

¿Qué política puede alinear la operación de despacho y drenaje de tanques en la terminal petrolera con el objetivo estratégico operacional?

### **Objetivos**

Proponer una política operacional alineada al objetivo estratégico “utilidad operacional sostenida”, que mejore la operación de despacho y drenaje de los tanques en la terminal petrolera reduciendo los costos asociados al exceso de agua libre.

#### **Objetivos específicos**

- Caracterizar la operación de drenaje de tanques y despacho de petróleo actual por medio de un diagnóstico para entender cuáles son los recursos involucrados.
- Construir un modelo de simulación de la operación de drenaje y despacho de petróleo en la terminal petrolera para evaluar el desempeño de las políticas operacionales.
- Proponer una política operacional de drenaje y despacho de petróleo que esté alineada con el objetivo estratégico “utilidad operacional sostenida” en la terminal petrolera para mejorar la eficiencia y contribuir al indicador de reducción de costos (30MUSD).

### **Marco teórico**

Para entender la operación de despacho y drenaje se usa el modelado de procesos de negocio (BPM), el cual da a las organizaciones una forma sencilla de entender y optimizar los flujos de trabajo por medio de la creación de representaciones visuales basadas en datos de los procesos comerciales clave (IBM, 2021). Para realizarlo se comienza definiendo los roles implicados y conectándolos por medio de eventos o actividades, se realiza un diagrama de flujo

“feliz”, luego se agregan subprocesos si es necesario, después se agregan caminos alternativos, excepciones, actividades o artefactos requeridos relacionados a estos caminos, completando así las decisiones y los caminos de “NO” (Khabbazi et al., 2013). Esta metodología mejora la toma de decisiones en la asignación de recursos, la mejora de procesos y la estrategia empresarial en general, determina problemas en el proceso y además permite mantener bajos los costos operativos. (IBM, 2021).

También para entender la operación y encontrar las posibles causas de los problemas se usa el diagrama de Ishikawa o “causa y efecto”, que es una herramienta de evaluación de riesgos reconocida y validada, la cual representa posibles causas que conducen a un efecto determinado (Colli et al., 2019). Para realizar el diagrama se siguen los siguientes pasos: Definición del Problema, Identificación de las causas, lluvia de ideas y revisión de ideas (Bustamante Zulueta & Sosa Vergara, 2022). Como ventajas, es fácil su interpretación y análisis, y se pueden observar las causas de acuerdo con el problema y sus efectos; además permite trabajar con gran cantidad de información y determinar las posibles causas de un problema, lo que conlleva a aumentar la probabilidad de identificar las causas principales (Delgado et al., 2021).

Respecto a la operación de drenaje; cuando los tanques de almacenamiento petrolero reciben producto normalmente se debe esperar un tiempo a que decante el agua en suspensión, esta agua separada en el fondo es llamada agua de fondo; en este contexto el agua libre definida como agua en una fase separada es sinónimo de agua de fondo (American Petroleum Institute, 1994). El agua libre o de fondo es la que necesita drenarse y esta se drena hasta el nivel necesario para exportar el producto.

El proyecto desarrolla una herramienta de machine learning para la predicción de agua libre final, usada para la simulación de Monte Carlo y como parte de una propuesta de política

operacional. El machine learning se define como la capacidad de los sistemas para aprender a partir de datos de entrenamiento para un problema en específico para automatizar el proceso de construcción de modelos analíticos y resolver tareas asociadas (Janiesch et al., 2021).

Los modelos son desarrollados en el software de Knime (Konstanz Information Miner), el cual es una plataforma de ciencia de datos de acceso libre; permite preprocesamiento de datos, integración y aplicación de distintas técnicas de machine learning (Löfström et al., 2023). Tiene un entorno modular que permite un montaje visual, también permite el análisis interactivo de datos; visualmente está dividido en nodos con sus propios algoritmos que interactúan entre sí, lo que permite la creación de flujos de trabajo (Anggraeni et al., 2022).

Para desarrollar el software se sigue la metodología de CRISP-DM, esta metodología es la usada de facto como estándar para aplicar en proyectos de minería de datos (Schröer et al., 2021). Se describe como un proceso con seis fases iterativas que sigue un modelo jerárquico que contiene conjuntos de tareas descritas en cuatro niveles de abstracción de lo general a lo específico: fase, tarea genérica, tarea especializada e instancia de proceso mostradas en la figura (Chapman et al., 2000).

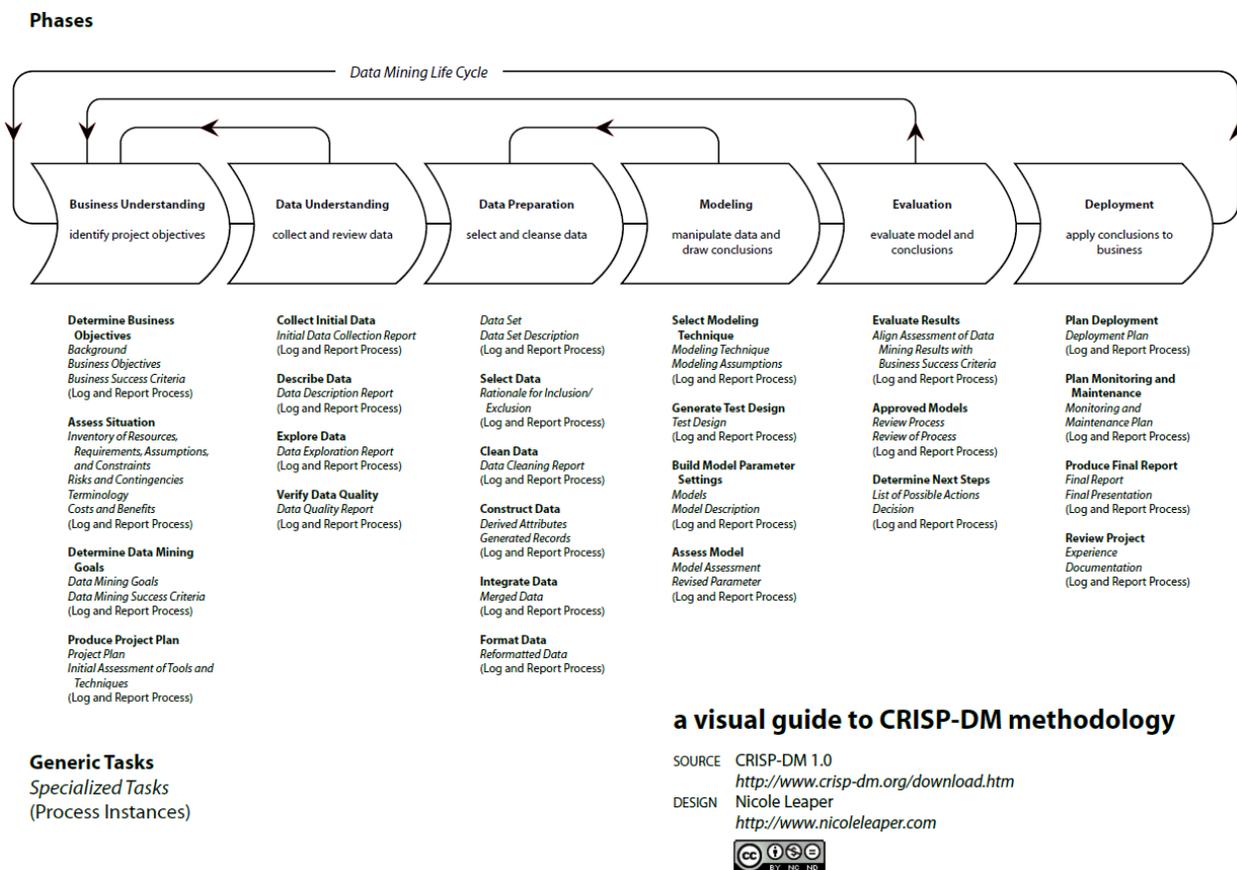


Figura 1. Proceso CRISP-DM

Los métodos de “forward feature selection” y “backward feature elimination” son usados para la selección de los atributos a usar en los modelos; estos evalúan múltiples subconjuntos agregando (forward feature selection) o eliminando (backward feature elimination) características para encontrar la combinación óptima que maximice el rendimiento del modelo (Weisberg, 2005).

La estrategia de “hold out validation” es usada en la generación de los modelos, donde los datos se dividen en datos de entrenamiento y datos de validación, este método es considerado como uno de los más confiables para determinar la precisión de un modelo de predicción (Lee, Liong, & Jemain, 2018).

Para simular la operación de despacho y drenaje se usa la simulación de Monte Carlo, la cual es un procedimiento estadístico antiguo y muy utilizado para hacer inferencias basadas en una pequeña muestra (Sortino et al., 2010). La simulación de Monte Carlo explota la ley de los grandes números en donde juega con la probabilidad y el volumen (Gilli et al., 2019), convergiendo en el valor verdadero conforme aumenta el volumen de eventos en la simulación. Para realizar la simulación de Monte Carlo primero es necesario crear un modelo que transforme los datos de entrada en datos relevantes para el análisis, este modelo o algoritmo debe simplificar el sistema a simular. Luego se determinan las entradas al modelo, estas son generadas por una fuente de números aleatorios y luego son transformadas de tal manera que puedan alimentar el modelo. Luego es usado el modelo y se estiman las acciones a tomar mediante el análisis estadístico de las salidas del modelo. Como beneficios, el uso de esta técnica permite realizar simulaciones con una muestra pequeña de datos y evitar la realización de experimentos, además del uso de información determinística y estocástica (López Agüí, 2008).

## **Metodología**

### **Caracterización de política operacional**

Para desarrollar el proyecto primero es necesario realizar un diagnóstico de la operación de drenaje y de despacho, para esto se realiza un **BPM** en donde se representa el proceso de la operación. Luego por medio de un **diagrama de “causa y efecto”** se proponen los posibles problemas que aquejan a la operación, los cuales ayudan a la simulación de esta y a proponer nuevas políticas operacionales.

### **Simulación de política operacional**

Después se realiza una **simulación de Monte Carlo** en donde se representa la operación de drenaje y despacho según las políticas actuales, lo cual ayuda a completar el diagnóstico de la

política actual. Para realizar la simulación es necesario tener muchas muestras de datos, pero solo se tienen alrededor de 300, por lo que es necesario crear una distribución del estado inicial de los tanques y para el estado final de los tanques crear una herramienta de predicción a partir de esos datos iniciales; para esto siguiendo la metodología **CRISP-DM** se desarrolla una herramienta de predicción de agua libre final que usa inteligencia artificial por medio del software Knime.

Cabe aclarar que la herramienta de inteligencia artificial desarrollada entrega datos de entrada para la simulación de Monte Carlo, se observa su desarrollo en el Apéndice A.

### **Evaluación y selección de nueva política operacional**

Luego de conocer el estado del proceso se procede a plantear y a evaluar nuevas políticas que alinean la operación al objetivo estratégico operacional. Las políticas son evaluadas según los resultados que se obtienen al simularlas en **Monte Carlo**.

Por último, a partir de los resultados de las simulaciones se **selecciona** la política operacional que más reduce los **costos** a la operación. Escogiendo así una nueva política operacional que permite a la operación de drenaje y despacho de la terminal petrolera alinearse con los objetivos estratégicos operacionales.

## **Desarrollo metodológico**

### **a) Caracterización de política operacional**

Conforme al mapa estratégico realizado por la compañía para alinearse con su visión y adaptarse a los retos internos y externos, se definieron objetivos estratégicos, entre ellos el objetivo estratégico operacional, el cual busca aumentar la utilidad operacional, mejorar la eficiencia y flexibilidad operativa y por medio de procesos optimizados por medio de tecnología e innovación aportar 30 millones de dólares en beneficios para el 2030.

El corazón del negocio es el transporte de petróleo y dado que se presentan ineficiencias en este asociadas al alto nivel de agua en los sistemas, se busca reducirla proponiendo una política operacional adecuada. El punto seleccionado para desarrollar la nueva política operacional es la terminal petrolera pues es la estación más importante y grande para la operación entre las 7 estaciones con almacenamiento de la empresa; representa el 66% del almacenamiento sumando todas las estaciones, tiene la importancia estratégica de llevar a cabo el proceso de exportación y también es el punto más importante de drenaje entre las instalaciones de la empresa. Para mejorar la política actual primero se debe conocerla a detalle, por ello primero es necesario conocer la operación de drenaje y despacho de la terminal petrolera de exportación. Para conocer a detalle las actividades y recursos involucrados en el proceso se realiza un BPM del proceso de drenaje y despacho, el cual se muestra a continuación. Este proceso comienza con la llegada del petróleo a la estación, su medición y drenaje inicial, y luego con el día de exportación, su drenaje y despacho.

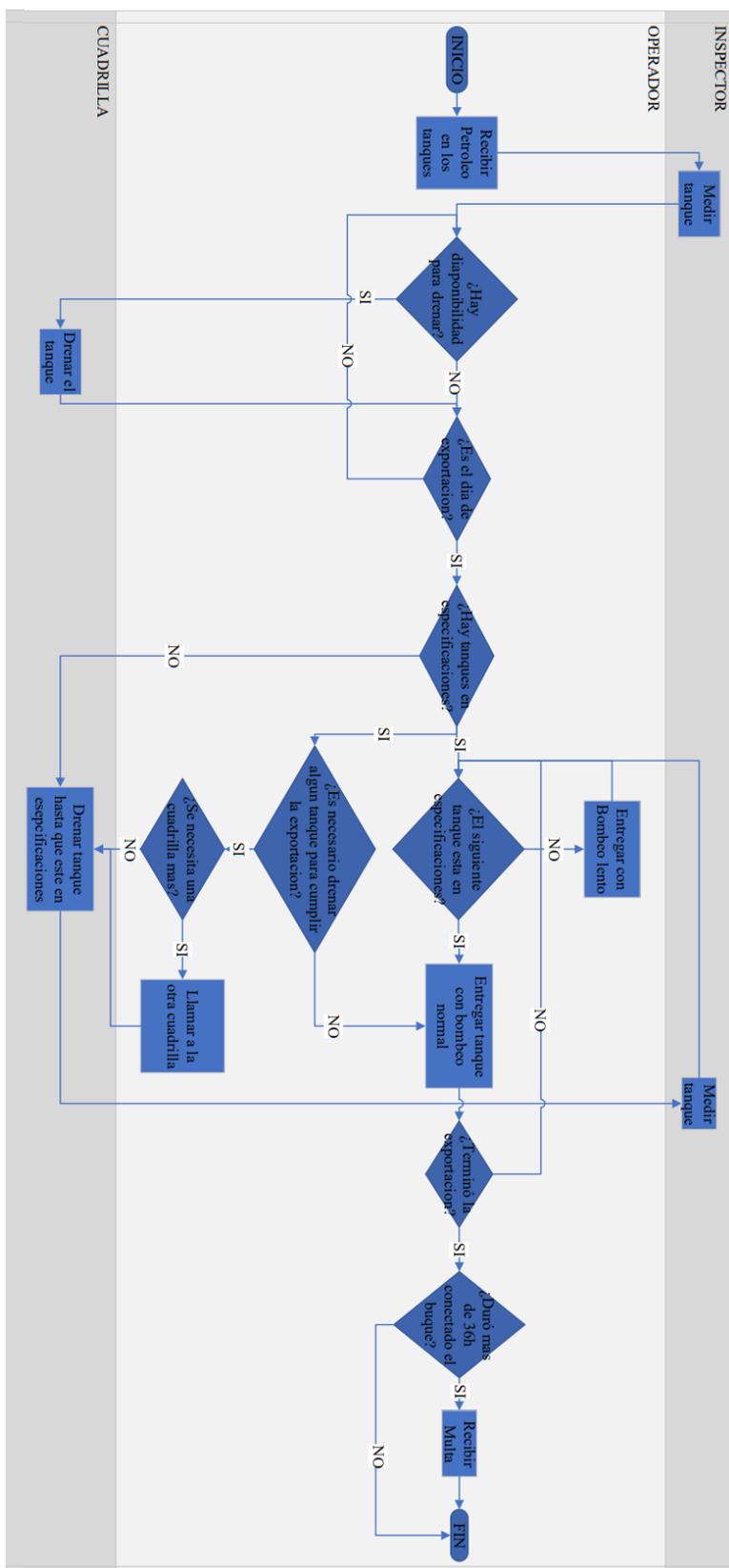


Figura 2. BPM de política operacional de drenaje y despacho de tanques

En el proceso encontramos tres roles distintos: el operador, el inspector y la cuadrilla, para 9 actividades y 8 decisiones. El proceso en la terminal petrolera comienza con el recibo de petróleo desde el oleoducto donde es medido volumétricamente junto con algunas de sus propiedades, luego se realiza el recibo de petróleo en los tanques (operador); luego de llenarse el tanque pasan horas hasta que es medido por el inspector confirmando las mediciones iniciales y determinando el agua libre inicial; luego si hay disponibilidad de tiempo de la operación y del personal este se drena, donde aproximadamente 30 minutos luego de drenarse se obtiene la medición del agua libre por el inspector. Mientras el tanque espera a ser exportado es drenado todo el tiempo posible para reducir los niveles de agua libre.

Las exportaciones suceden por ventanas de tiempo, las cuales son de 48 horas y por producto. En la ventana el buque tiene 36 horas para estar conectado, si este tiempo es excedido la empresa recibe una multa elevada por cada hora de más, por lo que normalmente la conexión se realiza a las 12 horas de iniciada la ventana. Cuando llega la ventana de exportación en caso de no haber tanques en especificaciones (con niveles aceptables de agua libre) es necesario realizar el drenaje de los tanques y luego medirlos (inspector) para verificar su estado; en caso de tener tanques para empezar la exportación esta se inicia (operador), pero si el siguiente tanque a exportar no está en especificaciones la exportación se realiza a una rata baja hasta que esté en especificaciones el siguiente tanque. Se asume que el drenaje de los tanques se comienza una hora antes de la ventana. Una vez exportado el volumen deseado se finaliza el proceso de despacho (operador). Cabe aclarar que para algunos tanques por restricciones no se pueden bombear a ratas altas por lo que se hace a ratas bajas, aunque el siguiente tanque esté listo para exportar.

La cuadrilla está activamente drenando los tanques y solo puede drenar un tanque a la vez, por lo que a veces es necesario tener un segundo equipo debido a la cantidad de agua libre presente; esta segunda cuadrilla no está en sitio y cuando se necesita esta aproximadamente 6 horas después de empezada la ventana pues se dan cuenta de que se necesita a las 3 horas de empezada y se demoran 3 horas más en empezar a drenar. En la terminal petrolera se tienen 3 productos (C, M, R), donde C y R son mezclados y cuentan con 5 tanques de almacenamiento y M cuenta con 2 tanques de almacenamiento.

Antes de la exportación el tanque puede quedar quieto hasta días, decantando el agua en emulsión y si es excesiva puede generar problemas en la exportación, pues se tendrían tanques fuera de especificaciones (no listos para exportar) ralentizando el bombeo, el gasto por bombeo lento es de \$1'355.299 pesos/h para los productos C y R (mezcla llamada producto P) y \$1'339.420 pesos/h para el bombeo del producto M. El flujo de drenado es de 146,2 BIs/h.

Tabla 1

*Velocidad de bombeo según producto y nivel de agua.*

<b>Producto</b>	<b>Estado del siguiente tanque.</b>	<b>Flujo de bombeo (BIs/h)</b>
M	Fuera de especificaciones	20.000
	En especificaciones	40.000
P	Fuera de especificaciones	25.000
	En especificaciones	36.000

La cuadrilla debe drenar los tanques priorizándolos según momento de exportación, nivel de agua libre y disponibilidad; esta drena según se lo permita la operación. En caso de

necesitarse una cuadrilla de más para drenar se pueden aumentar los tiempos de drenaje y a su vez los costos de operación y el riesgo a reclamaciones por agua presente en el petróleo en la exportación debido al retraso en su llegada. El costo de una cuadrilla de 3 obreros por hora es de \$43.249 pesos.

En la operación de despacho y drenaje, se tienen los siguientes problemas: no se sabe el volumen final de los tanques desde el comienzo de la operación y si estarán en especificaciones, lo que conlleva a no saber de la necesidad de una segunda cuadrilla hasta que son medidos los tanques y a no conocer los tiempos necesarios de drenado para definir cuanto tiempo llevará la operación de despacho; también se tiene una capacidad de drenaje limitada por cantidad de personal y altos tiempos de espera por apoyo. Por ello es necesaria una estrategia que responda a estas necesidades y que agilice y asegure la operación.

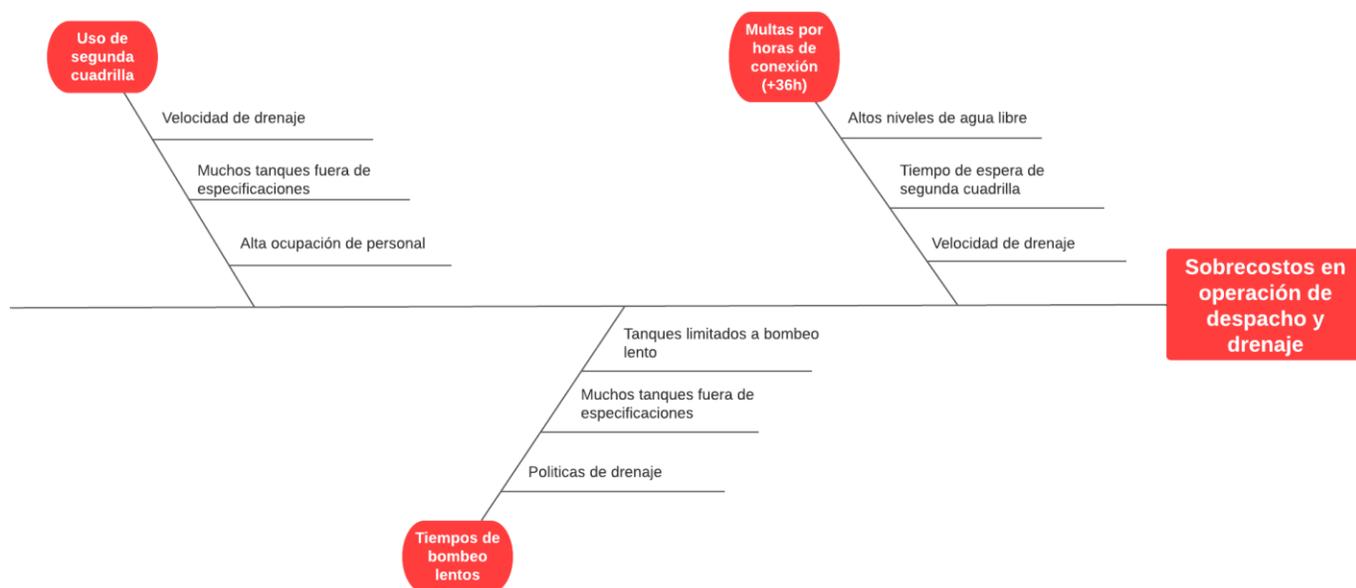


Figura 3. Diagrama de Pescado

Por medio de un diagrama de pescado se conocen las principales causas de los sobre costos en la operación de despacho y drenaje, donde tenemos las multas por horas de conexión, los tiempos de bombeo lentos y el uso de una segunda cuadrilla de drenaje. Para la causa de tiempos de bombeo lentos, como política actual de drenaje se bombea lento si el siguiente tanque a exportar esta fuera de especificaciones dándole tiempo a que lo drenen, sin embargo, este bombeo lento es un sobre costo; esta política no se puede cambiar debido a que si se para el despacho por acelerar el bombeo se pueden tener problemas de calidad de producto en la exportación y en la medición. También están las causas referentes a los activos, como los tanques y las facilidades de drenaje, pero estas no pueden ser abordadas debido a que llevaría a disminuir los inventarios en la operación, además de ser costosas. Otras causas son las referentes a la cuadrilla de drenaje, donde la cuadrilla normalmente está ocupada y la segunda cuadrilla se tarda 6 horas en empezar a drenar desde el inicio de la ventana de exportación. Por lo que la política a aplicar debe buscar disminuir el sobre costo por medio del correcto manejo del recurso humano en la operación.

#### **b) Simulación de política operacional**

Luego de conocer a detalle la política operacional actual, se procede a hacer una simulación de Monte Carlo para conocer los costos; donde se busca simular todo el comportamiento de la operación desde el recibo de los tanques, el estado final de estos y la operación de despacho y drenaje tomando como base la información suministrada por la empresa, correspondiente al histórico de un año de operación de la terminal petrolera.

Para la simulación de Monte Carlo primero se realiza un tratamiento estadístico para identificar el comportamiento del estado inicial de los tanques, donde se determina la distribución de probabilidad del agua libre inicial y el volumen de tanque inicial por cada tanque y producto (M y P), Apéndice D. Para realizarlo se toma el historial de operación de un año, se

agrupan los datos y se usa la herramienta “Input Analyzer” para conocer la distribución que mejor simula los datos teniendo en cuenta la prueba de chi cuadrado, también se tiene en cuenta la facilidad de calcular la inversa del tipo de distribución para la simulación. Además, algunas distribuciones se hacen empíricas por su poca relación con otros tipos de distribuciones.

Tabla 2

*Tipos de distribución Producto M.*

<b>Tipo de Tanque</b>	<b>Agua libre inicial</b>	<b>Volumen inicial del tanque</b>
Tanque pequeño	Beta	Beta
Tanque grande	Empírica	Beta

Tabla 3

*Tipos de distribución Producto P.*

<b>Tipo de tanque</b>	<b>Agua libre inicial</b>	<b>Volumen inicial del tanque</b>
Tanque 1	Exponencial	Beta
Tanque 2	Normal	Empírica
Tanque 3	Beta	Beta
Tanque 4	Empírica	Beta

Luego a partir de las distribuciones que explican el comportamiento inicial de los tanques y por medio de 10.000 datos aleatorios se crean 10.000 eventos de estados iniciales de los tanques (volumen del tanque y agua libre inicial) al aplicar la inversa de las distribuciones. La operación de despacho del producto M tiene en cuenta dos tanques, uno grande y uno pequeño; mientras que para el producto P se tienen 5 tanques de cuatro tipos.

Como siguiente paso se simula el estado final de los tanques, es decir el volumen final y el agua libre final previo a la operación de exportación. El volumen final tiene una relación lineal

con el estado inicial por lo que se aplican las fórmulas enunciadas más adelante para cada tanque y producto, donde “y” es el volumen final y “x” el volumen inicial.

$$\text{Producto M. Tanque Pequeño } y = 0,9978x + 265,08$$

$$\text{Producto M. Tanque grande } y = 0,9973x + 177,27$$

$$\text{Producto P. T1 } y = 1,0009 * x - 258,99$$

$$\text{Producto P. T2 } y = 0,9984 * x + 346,66$$

$$\text{Producto P. T3 } y = 0,9883 * x + 4860,9$$

$$\text{Producto P. T4 } y = 0,9987 * x + 248,77$$

Para seguir con la simulación de los estados finales de los tanques es necesario conocer el volumen de agua libre final, se realizan análisis tomando en cuenta los datos suministrados por la empresa para poder predecir con más exactitud el volumen de agua libre final a partir de todas las variables encontradas en este conjunto de datos. Para ello se sigue la metodología de minería de datos CRISP-DM, mostrada más a detalle en el Apéndice A; se comienza realizando análisis estadísticos usando la herramienta de Knime, también a partir de los datos recibidos se prueban distintos conjuntos de variables como predictoras para predecir el agua libre final y a su vez se prueban distintos modelos presentes en Knime; se encuentra que la mejor manera para describir el comportamiento del agua libre final es en base al agua libre inicial y por medio de un modelo de machine learning, específicamente un modelo de “Gradient Boosted Trees (Regression)”. El modelo desarrollado tiene una exactitud del 77% teniendo en cuenta la mayoría de los datos. Esta herramienta de predicción es usada para simular el agua libre final a partir de los 10.000 eventos de agua libre inicial creados anteriormente.

El uso de técnicas de analítica de datos y machine learning son de gran valor para simular procesos; en este caso se encontró al analizar los datos suministrados por la empresa que estos no

hacían referencia a los últimos movimientos de los tanques y por lo tanto existían movimientos (recibo, drenaje) entre los dos estados, por lo tanto se necesitaba un modelo que encontraría la relación entre las variables iniciales y el agua libre final teniendo en cuenta estos movimientos, es decir, simular la operación de la terminal; por medio de Knime se lograron validar desde modelos de regresión lineal, pasando por arboles de decisión y redes neuronales, permitiendo también validar distintas variables predictoras, además esta herramienta de analítica de datos permitió guardar el modelo y poder usarlo de manera muy practica logrando así predecir 10.000 eventos para la simulación de Monte Carlo.

Al tenerse simulado tanto el volumen de tanque final como el volumen de agua libre final se procede a simular la operación de despacho y drenaje. Para la simulación de Monte Carlo se desarrolla un modelo para la operación el cual tiene en cuenta el BPM descrito anteriormente y demás variables de la operación, como la capacidad de cada tanque, velocidades de bombeo y tiempos operativos. Para este modelo como entrada se tiene el volumen final de cada tanque y su volumen de agua libre final.

En el modelo primero se definen los tanques fuera de especificaciones; un tanque está en especificaciones si el agua libre está por debajo del volumen máximo aceptado para exportar, este volumen cambia según cada tipo de tanque. También se determina el tiempo de drenaje según la diferencia del volumen de agua libre final y el volumen mínimo aceptado, y el flujo de drenado de 146,2 Bls/h; a este tiempo se agregan 30 minutos de reposo.

Luego se define el tiempo de la operación de exportación. Como directrices para calcular el tiempo de operación de cada escenario se tienen las velocidades de bombeo relacionadas en la tabla 4 y los tiempos de drenaje, cabe aclarar que el tiempo de operación se cuenta a partir de 12h luego del comienzo de la ventana de exportación cuando el buque es conectado a la terminal

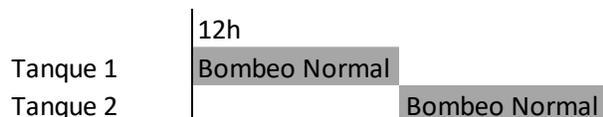
petrolera. En el cálculo se tiene en cuenta el tipo de producto y por consiguiente los tipos de tanques y sus estados previos a la exportación. Se aclara que no se puede despachar más de un tanque a la vez y solo se drena un tanque por cuadrilla.

Tabla 4

*Velocidad de bombeo según producto y nivel de agua.*

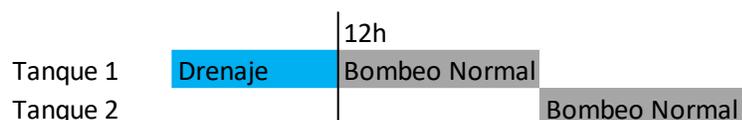
Producto	Estado del siguiente tanque.	Flujo de bombeo (Bls/h)
M	Fuera de especificaciones	20.000
	En especificaciones	40.000
P	Fuera de especificaciones	25.000
	En especificaciones	36.000

**Modelo de tiempo de operación para el producto M.** Para el producto M al haber 2 tanques, hay 3 tipos posibles operaciones de despacho: que todos los tanques estén en especificaciones, un solo tanque fuera de especificaciones y los dos tanques fuera de especificaciones. Se simula la operación a partir de estas 3 opciones, donde se prioriza el tanque con más tiempo de drenaje y el flujo de despacho es lento si el siguiente tanque esta fuera de especificaciones. A continuación, se presentan distintos ejemplos que se encuentran simulados en el modelo.



*Figura 4.* Modelo de tiempo de operación. Producto M. Caso 1: Tanques en especificaciones

Para el caso de que todos los tanques estén en especificaciones, luego de las 12 horas previas a la conexión; se suma el tiempo de exportación a velocidad de bombeo normal. Por lo anterior, para esta opción no hay tiempos de bombeo lentos ni horas de cuadrilla extra ni tiempos de conexión extensos.



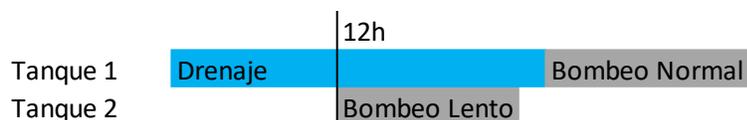
*Figura 5.* Modelo de tiempo de operación. Producto M. Caso 2A: 1 tanque fuera de especificaciones

Para el caso de que un tanque este en especificaciones y el tiempo de drenaje sea menor o igual a 12 horas; se suma el tiempo de exportación a velocidad de bombeo normal. Por lo anterior, para esta opción no hay tiempos de bombeo lentos ni horas de cuadrilla extra ni tiempos de conexión extensos.



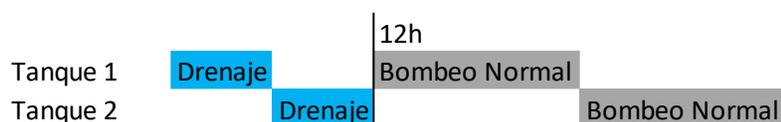
*Figura 6.* Modelo de tiempo de operación. Producto M. Caso 2B: 1 tanque fuera de especificaciones

Para el caso de que un tanque este en especificaciones y el tiempo de drenaje sea mayor a 12 horas; se suma el tiempo de drenaje mayor a 12 horas como bombeo lento del tanque 2, más el tiempo de bombeo normal para completar el despacho de este, más el tanque 1 a velocidad de bombeo normal.



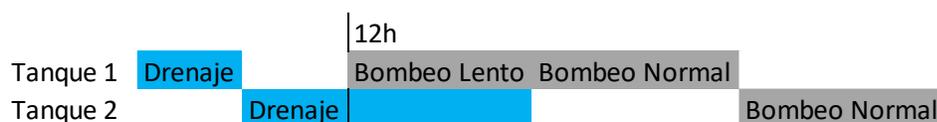
*Figura 7.* Modelo de tiempo de operación. Producto M. Caso 2C: 1 tanque fuera de especificaciones

Para el caso de que un tanque este en especificaciones y el tiempo de drenaje sea mayor a 12 horas más el tiempo de bombeo lento del tanque 2; se suma el tiempo de drenaje mayor a 12 horas, más el tanque 1 a velocidad de bombeo normal.



*Figura 8.* Modelo de tiempo de operación. Producto M. Caso 3A: 2 tanques fuera de especificaciones

Para el caso de que no haya tanques en especificaciones y el tiempo de drenaje sea menor o igual a 12 horas; se suma el tiempo de bombeo normal para los dos tanques. Por lo anterior, para esta opción no hay tiempos de bombeo lentos ni horas de cuadrilla extra ni tiempos de conexión extensos.



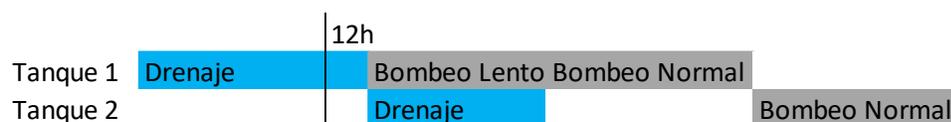
*Figura 9.* Modelo de tiempo de operación. Producto M. Caso 3B: 2 tanques fuera de especificaciones

Para el caso de que no haya tanques en especificaciones y el tiempo del segundo drenaje finalice después de las 12 horas; se suma el tiempo de bombeo lento para el primer tanque, más el bombeo normal para completar el despacho del tanque 1, más el bombeo normal del tanque 2.



*Figura 10.* Modelo de tiempo de operación. Producto M. Caso 3C: 2 tanques fuera de especificaciones

Para el caso de que no haya tanques en especificaciones, el tiempo de drenaje sea mayor a 12 horas y la finalización del drenaje del tanque 2 sea luego del bombeo lento del tanque 1; se suma el tiempo de drenaje mayor a 12 horas para el drenaje del tanque 2, más el bombeo normal del tanque 2.



*Figura 11.* Modelo de tiempo de operación. Producto M. Caso 3D: 2 tanques fuera de especificaciones

Para el caso de que no haya tanques en especificaciones y el tiempo de drenaje 1 sea mayor a 12 horas; se suma el tiempo de drenaje mayor a 12 horas para el tanque 1, más el tiempo de bombeo lento del tanque 1 igual al tiempo de drenaje del tanque 2, más el tiempo de bombeo normal faltante del tanque 1, más el tiempo de bombeo normal para el tanque 2.



*Figura 12.* Modelo de tiempo de operación. Producto M. Caso 3E: 2 tanques fuera de especificaciones

Para el caso de que no haya tanques en especificaciones y el tiempo del primer drenaje sea mayor a 12 horas y el segundo mayor al tiempo de bombeo lento del primer tanque; se suma

el tiempo de drenaje mayor a 12 horas para el tanque 1, más el tiempo de drenaje del tanque 2, más el tiempo de bombeo normal para el tanque 2.



Figura 13. Modelo de tiempo de operación. Producto M. Caso 3F 2 cuadrillas

Si se encuentra que el tiempo de operación será mayor a 36 horas, se decide usar una segunda cuadrilla. En este caso se tienen los dos tanques fuera de especificaciones y el tiempo de drenaje excede las 12 horas iniciales; se suma el tiempo de drenaje mayor a 12 horas necesario para comenzar a bombear, se suma el tiempo de bombeo lento mientras el otro tanque está en drenaje y el tiempo de bombeo normal necesario para lograr el despacho.

El modelo de tiempo de operación para una segunda cuadrilla puede tener distintas variaciones a la ilustrada anteriormente, como: que el drenaje 1 termine luego del drenaje del tanque 2 o que el drenaje del tanque 2 termine luego del bombeo lento del tanque 1.

**Modelo de tiempo de operación para el producto P.** Para el producto P hay 5 tanques, donde se pueden usar entre 3 y 5 en la operación según el volumen de cada uno al momento de exportar. Respecto al orden de los tanques, se priorizan por menor tiempo de drenaje.

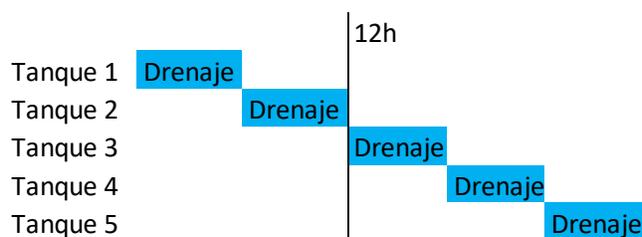
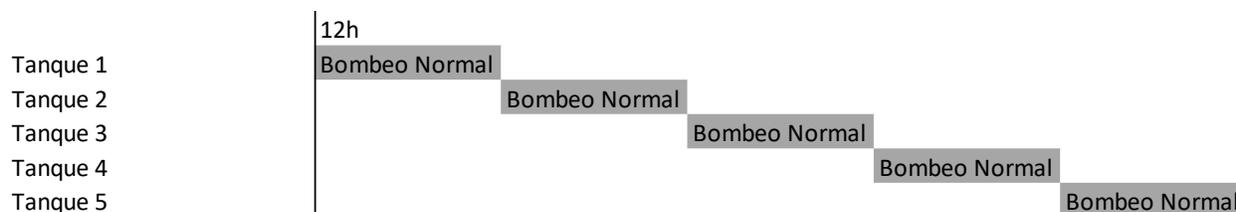


Figura 14. Operación de drenaje para 5 tanques fuera de especificaciones y una cuadrilla



Figura 15. Operación de drenaje para 5 tanques fuera de especificaciones y dos cuadrillas



*Figura 16.* Operación de despacho para 5 tanques, todos en especificaciones

El modelo de tiempo de operación para el producto P presenta variaciones según los tiempos de drenaje y los tiempos de bombeo de cada tanque, estos se tienen en cuenta en la simulación de Monte Carlo. En general los tiempos de operación se calculan sumando los tiempos de despacho para completar el volumen de exportación más los tiempos de drenaje mayores a 12 horas, más los tiempos de drenaje entre los tiempos de despacho. Cabe aclarar que los flujos de despacho son lentos si el siguiente tanque en la lista de priorización está fuera de especificaciones. La segunda cuadrilla se simula para tiempos de operación mayores a 36 horas.

Además del tiempo de operación determinado anteriormente, se calcula el tiempo de bombeo lento, el cual es la suma de los tiempos de bombeo lento de los tanques usados en la operación de exportación, también se calculan las horas de segunda cuadrilla que son la suma de las horas en que la segunda cuadrilla estuvo trabajando.

Luego se calcula el **sobrecosto de la operación**, sumando cada costo:

- Costo de segunda cuadrilla: El costo de una cuadrilla de 3 obreros por hora es de \$43.249 pesos y se multiplica por las horas de trabajo para cada evento.
- Costo por multas por conexiones extensas: Para cada evento se calcula el tiempo de operación y si este supera las 36 horas con las dos cuadrillas, cada hora de más es multiplicada por el valor de la multa.

- Costo por tiempo de bombeo lento: Se multiplica las horas de bombeo lento por su respectivo costo. El costo por bombeo lento es de \$1'355.299 pesos/h para el producto P y \$1'339.420 pesos/h para el bombeo del producto M.

Como resultado de la simulación de Monte Carlo se tiene el sobre costo de operación para cada evento, este se divide en tiempos extensos de conexión, bombeos lentos y uso de cuadrillas extras. Los costos por tipo de producto se totalizan ponderando la ocurrencia de las ventanas exportación para cada producto (Tabla 5).

Tabla 5

*Tabla de costos adicionales operativos promedio. (USD)*

<b>Costo</b>	<b>M</b>	<b>P</b>	<b>Ponderado</b>
Conexiones extensas	\$ 2.300	\$ 4.720	\$ 2.784
Tiempos de bombeo lentos	\$ 2.216	\$ 931	\$ 1.959
Uso de segunda cuadrilla	\$ 67	\$ 7	\$ 55
<b>Total</b>	<b>\$ 4.583</b>	<b>\$ 5.658</b>	<b>\$ 4.798</b>

Análisis de sensibilidad de la simulación de Monte Carlo. la mayoría de las variables usadas en el Monte Carlo presentan una distribución incluida en la simulación, sin embargo, algunas fueron dejadas fijas como el dólar y la ocurrencia de exportación por producto; este análisis presenta el impacto de las dos variables mencionadas anteriormente, el dólar y las ocurrencias de exportación por producto. El dólar es usado para calcular el costo de la cuadrilla y el costo de bombeo lento los cuales son dados en pesos, para la simulación base se tomó una conversión de 3714,28 pesos (valor promedio entre el periodo de tiempo de los datos tomados), como máximo se toma el máximo histórico a la fecha 4968,94 pesos por dólar y como mínimo se toma el valor mínimo presentado en el 2020 (año en que comienza la toma de datos usados en el

proyecto) 3253,89 pesos por dólar. Respecto a las ocurrencias se maximizan por cada producto hasta llegar a exportar solo producto M o P.

Tabla 6

*Tabla de resultados del análisis de sensibilidad. (USD)*

<b>Variación de sobrecosto de operación</b>	<b>Productos sin cambios</b>	<b>Solo producto M</b>	<b>Solo producto P</b>
Dólar sin cambios	\$ 4.798	\$ 4.583	\$ 5.658
Dólar Alto	\$ 4.289	\$ 4.006	\$ 5.421
Dólar Bajo	\$ 5.083	\$ 4.906	\$ 5.791

Según los resultados presentados en la tabla anterior se puede observar cómo varía el sobrecosto de operación según el cambio del dólar y de los productos exportados; el valor más alto corresponde a un dólar bajo y a la exportación del producto P con \$5791 USD por operación, esto representa un incremento del 21% del sobrecosto; el valor más bajo corresponde a un dólar alto y a la exportación del producto M con \$4006 USD por operación, esto es un 17% más bajo. El modelo es más sensible a los cambios de productos de exportación que al dólar, se encuentra que puede variar un rango del 22% con los productos y uno del 17% con el dólar.

En conclusión, como resultado de la simulación de Monte Carlo se tiene a la política actual con un sobrecosto de \$4798 USD por operación (Tabla 5), es decir \$8,64 MUSD en 10 años. Para la política actual el concepto más alto de costo se da por las conexiones que sobrepasan el tiempo establecido para el producto P. Además, se encuentra que el costo del uso de una segunda cuadrilla no es representativo en comparación con el costo total. Las multas por exceder el tiempo de conexión con el buque equivalen al 3% de los bombeos de producto P y un 4% para el producto M.

Según las proyecciones, los costos asociados a los tres puntos tratados (\$8,64 MUSD) son representativos para la estrategia de la compañía al acaparar aproximadamente un tercio de los objetivos, por lo que se confirma la necesidad de atacar estos problemas operativos.

Además, se calcula el costo operacional total sumando el sobre costo, asumiendo una cuadrilla fija todo el tiempo y teniendo en cuenta los bombeos normales. Como resultado se encuentra que el costo operacional es de \$8262 USD para el producto M, \$15546 USD para el producto P y \$9719 USD como ponderado para la operación.

Al aplicar la política adecuada se esperan beneficios económicos al disminuir el costo y tiempo de operación de la exportación. Cuando hay gran cantidad de agua en los tanques la tasa de bombeo debe ser más lenta, lo que además de ralentizar la operación, la encarece por el funcionamiento de las bombas; además, si se alcanzan más de 36 horas de conexión con el buque se generan multas.

También, como beneficio consecuente esta aumentar la calidad del petróleo; cuando hay gran cantidad de agua en los tanques se envían trazas de agua, lo que disminuye el valor comercial del petróleo y en algunos casos puede dar lugar a reclamaciones costosas, por ello la solución en consecuencia también cuidaría la reputación de la empresa. La presencia de agua en el sistema también causa problemas de corrosión e incrustaciones lo que aumenta los costos de transporte, afectando la vida útil del tanque e incrementando los costos de mantenimiento. Además, la presencia de agua en el tanque disminuye la capacidad de este y representa un costo de inventario al almacenarse un producto que no representa valor para el negocio. Por otro lado, un bajo nivel de agua libre también significa mayor seguridad; en caso de incendio en un tanque de almacenamiento la presencia de agua libre o emulsiones puede generar “Boilover”, donde el calor producido llega por convección al fondo del tanque y cuando el agua alcanza su punto de

ebullición, el vapor de agua generado expulsa el hidrocarburo del tanque expandiendo el incendio (Mandal et al., 2022).

### c) **Evaluación y selección de nueva política operacional**

#### **Definición de nuevas políticas.**

Es necesario proponer una política de operación alineada a los objetivos estratégicos operacionales, que mejore la operación de despacho y drenaje de tanques en la terminal petrolera reduciendo los costos asociados al exceso de agua libre en los sistemas optimizando la operación y los recursos humanos.

¿Qué política puede mejorar la operación de despacho y drenaje de tanques en la terminal petrolera? Teniendo en cuenta el diagnóstico las políticas deben abordar el recurso humano y la planeación de la operación de despacho y drenaje. Estas deberían reducir el costo de la operación por medio de la reducción de tiempos extensos de conexión, costos de recursos humanos y tiempos de bombeo lentos.

Conforme a las limitaciones expresadas en el diagnóstico como el tener solo una cuadrilla fija en campo y tener hasta 6 horas de espera para drenar con una segunda cuadrilla, se proponen las siguientes políticas para reducir los costos de operación:

1. **Agregar una segunda cuadrilla fija en campo.** Esta nueva cuadrilla de ser necesario apoya a la primera en la ventana de exportación actual, y si está libre apoya drenando los tanques de la siguiente ventana. Es posible adelantar 4 horas de drenado de la siguiente ventana de exportación cuando hay cambio de producto el día siguiente; sin embargo, esto no se puede hacer si el siguiente día se exporta el mismo producto, pues normalmente los tanques están indisponibles por la operación.

2. **Agregar una cuadrilla fija para 12h antes del cambio de producto de exportación y tener otra disponible para apoyo.** Tener una cuadrilla a la espera de apoyar la ventana de exportación actual de ser necesario y la otra disponible antes del cambio de producto.

3. **Usar una herramienta de planeación prescriptiva que apoye en la decisión de llamar una cuadrilla para la ventana actual de exportación.** Con el fin de eliminar los tiempos muertos en la llamada de cuadrillas adicionales se propone una herramienta que por medio de predicciones recomiende la planeación para la operación de drenaje y el llamado oportuno de cuadrillas. Esta herramienta (DDS) se basaría en la correcta predicción del agua libre final en los tanques según el estado inicial de estos, como un trabajo futuro de la “herramienta de predicción de agua libre final” con la cual el Monte Carlo fue construido.

4. **Usar una herramienta de planeación prescriptiva que apoye en la decisión de llamar una cuadrilla para la ventana actual de exportación y agregar una cuadrilla fija 12h antes del cambio de producto de exportación.** Además de la herramienta (DDS) usada en la opción 3, tener una cuadrilla fija para el drenado anticipado a las ventanas de exportación cuando hay cambio de producto.

### **Evaluación de nuevas políticas.**

La selección de la política se realiza en base al indicador de ahorro. Las iniciativas son simuladas modificando el Monte Carlo con el que se evaluó anteriormente el resultado de la política actual.

1. **Agregar una segunda cuadrilla fija en campo.** Como beneficios de esta política se tiene, la eliminación del tiempo de espera para la operación de la segunda cuadrilla; también se adelanta el drenado de tanques para la siguiente ventana de exportación cuando hay cambio de producto, siempre y cuando esta cuadrilla adicional no este ocupada apoyando la ventana actual, **esto equivale al 96% del tiempo para el producto P y el 57% del tiempo para el producto M.** Como puntos en contra se tiene el costo fijo de una cuadrilla de más.

Tabla 7

*Tabla de resultado de simulación política 1. (USD).*

<b>Costo</b>	<b>M</b>	<b>P</b>	<b>Ponderado</b>
Conexiones extensas	\$ 1.952	\$ 4.338	\$ 2.429
Tiempos de bombeo lentos	\$ 1.950	\$ 874	\$ 1.734
Uso de segunda cuadrilla	\$ 559	\$ 559	\$ 559
<b>Total</b>	<b>\$ 4.460</b>	<b>\$ 5.771</b>	<b>\$ 4.723</b>

La Política 1 tiene un sobrecosto de \$4723 USD por operación, proyectados a \$8,5MUSD en 10 años, 1,6% de ahorro. Como cambio importante se tiene el agregar el costo de una cuadrilla disponible todo el tiempo, aumentando considerablemente el costo del recurso humano. Sin embargo, esto disminuye el costo total al reducir los otros costos asociados a conexiones extensas y tiempos de bombeo lentos, aunque solo representa un 1,6% de ahorro.

2. **Agregar una cuadrilla fija para 12h antes del cambio de producto de exportación y tener otra disponible para apoyo.** Como beneficios de esta política está el tener fija una cuadrilla que adelante horas de drenaje antes de cambiar la ventana de

exportación a un nuevo producto y tener la posibilidad de llamar a otra cuadrilla cuando la operación de la ventana de exportación actual lo requiera. Como desventaja se tiene el aumento nuevamente a 6 horas de espera para la entrada en operación de la cuadrilla de apoyo.

Tabla 8

*Tabla de resultado de simulación política 2. (USD).*

<b>Costo</b>	<b>M</b>	<b>P</b>	<b>Ponderado</b>
Conexiones extensas	\$ 2.050	\$ 4.428	\$ 2.526
Tiempos de bombeo lentos	\$ 2.195	\$ 887	\$ 1.933
Uso de segunda cuadrilla	\$ 79	\$ 56	\$ 74
<b>Total</b>	<b>\$ 4.324</b>	<b>\$ 5.372</b>	<b>\$ 4.533</b>

La Política 2 tiene un sobrecosto de \$4533USD por operación, proyectado a \$8,16MUSD en 10 años, lo que representa un ahorro del 5,5%. Como cambio importante se tiene el aumento leve del costo del recurso humano por la cuadrilla fija; sin embargo, esto disminuye el costo total al reducir los otros costos asociados a conexiones extensas y tiempos de bombeo lentos.

3. **Usar una herramienta de planeación prescriptiva que apoye en la decisión de llamar una cuadrilla para la ventana actual de exportación.** Esta herramienta se desarrollaría en base a la “herramienta de predicción de agua final” nombrada en el presente trabajo. Como beneficio de esta política está el reducir en 3 horas la espera de la segunda cuadrilla, al predecir cuándo es necesitada.

Tabla 9

*Tabla de resultado de simulación política 3. (USD).*

<b>Costo</b>	<b>M</b>	<b>P</b>	<b>Ponderado</b>
Conexiones extensas	\$ 2.174	\$ 4.660	\$ 2.671
Tiempos de bombeo lentos	\$ 2.058	\$ 916	\$ 1.830
Uso de segunda cuadrilla	\$ 52	\$ 7	\$ 43
<b>Total</b>	<b>\$ 4.285</b>	<b>\$ 5.583</b>	<b>\$ 4.544</b>

La Política 3 tiene un sobrecosto de \$4544 USD por operación, proyectado a \$8,18 MUSD en 10 años y con un ahorro del 5,3%. Como resultado importante se presenta una reducción de costos en cada ítem.

4. **Usar una herramienta de planeación prescriptiva que apoye en la decisión de llamar una cuadrilla para la ventana actual de exportación y agregar una cuadrilla fija 12h antes del cambio de producto de exportación.** Tomando como base la política 3, se busca optimizarla añadiendo una cuadrilla fija para 12h antes del comienzo de ventanas de exportación con cambio de producto. Como beneficio de esta política está el reducir en 3 horas la espera de la segunda cuadrilla y el aprovechar las horas de drenado anticipado para las ventanas mencionadas.

Tabla 10

*Tabla de resultado de simulación política 4. (USD).*

<b>Costo</b>	<b>M</b>	<b>P</b>	<b>Ponderado</b>
Conexiones extensas	\$ 1.928	\$ 4.370	\$ 2.416
Tiempos de bombeo lentos	\$ 2.041	\$ 876	\$ 1.808
Costo de personal	\$ 64	\$ 56	\$ 63
<b>Total</b>	<b>\$ 4.034</b>	<b>\$ 5.302</b>	<b>\$ 4.287</b>

La Política 4 tiene un sobrecosto de \$4287 USD por operación, proyectado a \$7,72 MUSD en 10 años, es decir un 10,6% de ahorro (0,92 MUSD). Se aumenta levemente el costo del recurso humano en un 14%, pero se disminuye el costo total al reducir los otros costos asociados a conexiones extensas y tiempos de bombeo lentos. Al agregar recurso humano (cuadrillas) en horas estratégicas, se logra compensar su costo al reducir el costo de los demás conceptos.

#### **Nueva política operacional seleccionada.**

La política seleccionada es la 4 “Usar una herramienta de planeación prescriptiva que apoye en la decisión de llamar una cuadrilla para la ventana actual de exportación y agregar una cuadrilla fija 12h antes del cambio de producto de exportación”, debido a que tiene un ahorro del 10,6% de los sobrecostos, casi el doble que la segunda mejor política operacional propuesta; desglosando el ahorro por concepto, se encuentra que se reduce un 13% el costo asociado a conexiones extensas, un 8% el costo asociado a tiempos de bombeo lentos y se aumenta un 14% el costo asociado a uso al recurso humano adicional.

Tabla 11

*Resumen de ahorro según políticas.*

Resumen de ahorro según políticas	
Política 1	1,6%
Política 2	5,5%
Política 3	5,3%
Política 4	10,6%

En la figura 17 se presenta el BPM de la nueva política operacional de despacho y drenaje de tanques de la terminal petrolera, donde las actividades en verde son las agregadas, se pasa de 9 a 10 actividades y se reduce de 8 a 7 las decisiones en comparación con el BPM original. El uso de la herramienta de apoyo a la decisión es la columna vertebral de esta política al recomendar los planes de despacho y drenaje, además de la recomendación de llamar a nuevas cuadrillas, lo que conduce a la reducción de costos en la operación. El uso de esta herramienta consumiría poco tiempo (2 min por cada drenaje), según las dinámicas de exportación el día en que más se podría usar consumiría 15 min de trabajo, lo cual no afecta las cargas de trabajo del operador.

Al aplicar herramientas de simulación y de inteligencia artificial y apoyar la política operacional en un sistema de apoyo a la decisión (DDS) se logra reducir el costo operacional; optimizando el recurso humano en la operación; el cual, aunque aumenta su costo al aumentar el uso de una segunda cuadrilla, lo compensa reduciendo los tiempos de bombeo lentos y las conexiones extensas de los buques, de esta manera se logra impactar un objetivo estratégico adicional: “Modelo organizacional optimo”. La reducción de tiempos de bombeos lentos significa una reducción en el consumo energético de la terminal petrolera, lo cual también impacta positivamente el objetivo ambiental de la compañía.

La política seleccionada se puede dividir en dos puntos, el cambio en el recurso humano y el uso de una herramienta de apoyo a la decisión. Respecto a la herramienta de apoyo a la decisión, se basaría en un desarrollo en la plataforma de Knime la cual es gratuita; El precio de la herramienta desarrollada puede estar alrededor de los 40 millones de pesos teniendo en cuenta otros desarrollos para la industria, sin embargo, no se cargaría con este precio a la empresa con el fin de validar la herramienta y seguir adelante con más desarrollos.

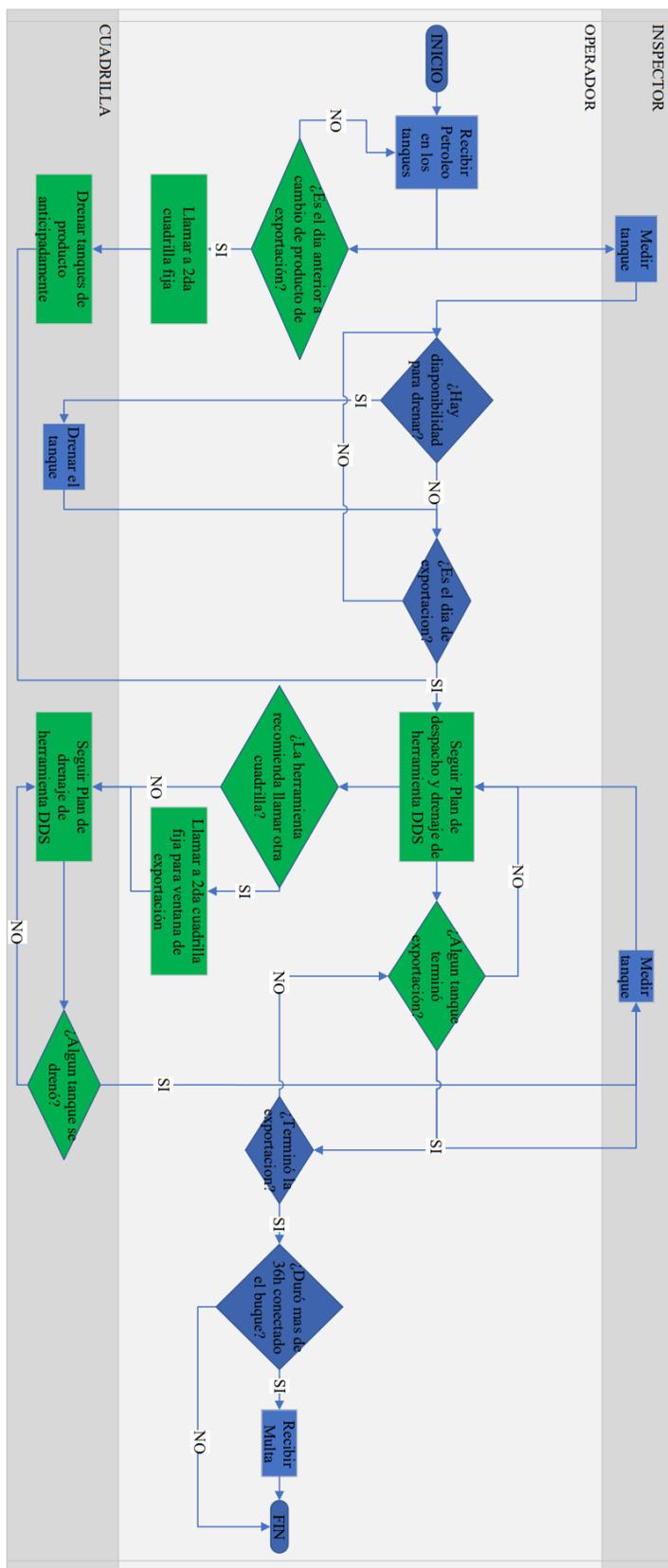


Figura 17. BPM de nueva política operacional de despacho y drenaje de tanques.

Como plan de implementación de la nueva política operacional se proponen las etapas, actividades y KPI's (fórmulas en Apéndice E) en la tabla 12. Se comenzaría con unas pruebas en campo, para probar el modelo de machine learning; luego se realizaría la instalación de la herramienta en los equipos de la operación y se haría la capacitación a los distintos turnos de la operación; una vez hecha la capacitación se comenzaría a usar la herramienta de manera progresiva en la operación hasta estabilizar el proceso, en este punto el seguimiento sería crucial para asegurar la nueva política operativa. Además de las actividades mencionadas anteriormente, se propone realizar un seguimiento trimestral de KPI's (Impacto a la operación y seguimiento de exactitud del modelo) y una actualización anual de la herramienta para adaptarse a posibles cambios en la operación y/o en las características de los productos.

En caso de que los KPI's de seguimiento sean incumplidos, se propone adelantar la actualización del modelo de ML y de solo fallar el KPI de impacto a la operación se propone analizar la política implementada respecto a los cambios operativos y hacer las modificaciones pertinentes.

Tabla 12

*Plan de implementación de nueva política operacional.*

<b>Etapa</b>	<b>Actividades</b>	<b>Responsable</b>	<b>KPI</b>	<b>Hito</b>
1. Pruebas en campo	Probar modelo de machine learning en campo.	Desarrollador	Exactitud del modelo* Aprobado si $\geq 77\%$	Modelo de machine learning validado
2. Instalación	Instalar Knime y modelo ejecutable en cada computador. (Ingeniero TIC)  Validar instalación correcta (Desarrollador)	Ingeniero TIC  Desarrollador	Porcentaje de computadores instalados  Aprobado si es 100%	Instalación de Herramienta de apoyo a la decisión en la operación

3. Capacitación a la operación	Se capacita a la operación por medio de: - Video de introducción (impacto, funcionamiento, cultura empresarial) - Video de capacitación (uso) - Reunión de 2 horas por persona (4)	Desarrollador  Operación	Porcentaje de operadores capacitados  Aprobado si es 100%	Operación capacitada
4. Implementación y estabilización	Asegurar la implementación de la nueva política operacional y la estabilización de la operación.: - Implementar la nueva política operacional. - Recomendaciones de la herramienta de apoyo a la decisión y recomendaciones de cuadrilla antes de cambio de producto. - Por una duración de tres meses observar los KPI's para asegurar la implementación y estabilización	Operación  Ingeniero TIC	Seguimiento de exactitud del modelo*. Aprobado si $\geq 77\%$ . Para cada mes  Implementación de política (DDS) Aprobado si es 100% Para último mes  Implementación de política (Cambio de producto) Aprobado si es 100% Para último mes  Impacto a la operación (Ahorro%) Aprobado si $> 0\%$ Para último mes	Nueva política operacional implementada

Nota. Exceptuando 3% de la operación, el cual representa los datos donde el modelo de ML expresa errores altos. Operativamente se recomienda para un tanque en específico del producto P, si hay drenajes importantes para volúmenes iniciales menores a 700bbls asumir el tanque en especificaciones.

Tabla 13.

*Cronograma de implementación de nueva política operacional.*

<b>Plan de implementación</b>	<b>Mes</b>				
<b>Actividades</b>	1	2	3	4	5
1. Pruebas en campo					
2. Instalación					
3. Capacitación a la operación					
4. Estabilización de la operación					

Sobre la gestión del cambio se aseguraría teniendo en cuenta dos aspectos, el personal y los procesos. Respecto al personal, se realizarían capacitaciones con videos y reuniones conforme se enuncia en las actividades del plan de implementación, también desarrollando un procedimiento sobre el uso de la herramienta y actualizando el procedimiento del cargo de la operación; en los videos de introducción se buscaría crear conciencia sobre la importancia de la política operacional en cuanto a beneficios esperados, objetivos organizacionales y cultura organizacional; además, la nueva política operacional incluye una herramienta de apoyo a la decisión la cual ayudaría a los procesos del cargo implicado. Respecto a los procesos, en el presente trabajo hay un análisis rigurosos de cómo se vería afectado y de los beneficios esperados, los procesos también se asegurarían con los procedimientos nombrados anteriormente, con el seguimiento trimestral de los indicadores y la actualización anual del modelo de Machine Learning.

Detallando el plan de inversión necesario para la política operacional se encuentra el costo de implementación, el costo de seguimiento trimestral y el costo de actualización del modelo, se descontarían al ahorro generado calculado anteriormente. La actualización del software tendría un costo de \$5.400 USD teniendo en cuenta la mano de obra y los costos en la industria petrolera. El costo de implementación seria de \$412 USD, dados por la mano de obra

necesaria para hacer seguimiento durante la estabilización, la instalación de la herramienta y las horas hombre de la operación para su capacitación. El costo del seguimiento trimestral equivaldría a \$79 USD. Además, se aclara que la herramienta no requiere un costo de instalación o un nuevo equipo para la operación.

Por lo tanto, el plan de inversión para la nueva política operacional tendría una inversión inicial de \$412 USD y un flujo de caja anual de \$86.213 USD, este flujo de caja se obtiene de los ahorros generados por la nueva política operacional la cual tiene en cuenta disminución de multas, mayor eficiencia en el bombeo de petróleo y el aumento del recurso humano en la operación, además de descontar el plan de inversión; esto representaría un VPN a 10 años de \$736.885 USD; para la tasa de descuento se usa la tasa de descuento del 2021 de una empresa colombiana del mismo sector y de tamaño similar. El costo anual de la política operacional haría parte del OPEX de la compañía al cubrirse por sus mismos beneficios durante el mismo periodo de tiempo.

La nueva política operacional tendría un ahorro neto de \$0,74 MUSD (VPN) proyectados a 10 años logrando aportar al objetivo de “utilidad operacional sostenida”. El alineamiento estratégico es medido respecto al indicador de ahorro esperado, para calcular este porcentaje de ahorro esperado se usa como base los costos operativos anuales de \$163,8 MUSD y los costos operativos variables anuales de \$54,3 MUSD, donde si se asume que el objetivo de treinta millones hace referencia a cada concepto se esperarían reducir un 2% y un 6% respectivamente. Al compararlo con el ahorro esperado del 10,3% para el sobrecosto operacional por la nueva política operacional se puede concluir que se están alineando los sobrecostos operacionales respecto a lo esperado para los costos variables; respecto al costo operacional total, este pasaría de \$9719 USD a \$9208 USD, es decir una disminución del 5,3% siendo mayor que la reducción

del 2% esperada para la operación, logrando el alineamiento estratégico con la política operacional propuesta, con un 283% (5,3% ahorro operación/ 2% ahorro toda la operación) respecto a lo esperado. Mostrando así que al usar herramientas de simulación y de inteligencia artificial en la operación se logra reducir los sobrecostos y generar ahorros importantes con una sola política aplicada en una estación.

Al aplicar de manera exitosa herramientas tecnológicas de vanguardia se apoya también el objetivo estratégico de “tecnología e innovación propulsoras de transformación”.

Se propone proyectar este ahorro aplicando esta política en las demás estaciones, logrando así apuntalar el objetivo de ahorrar 30 millones de dólares en 10 años. Además, al aplicar esta política en estaciones que alimentan la terminal petrolera se lograría reducir el agua libre en cadena hasta llegar a disminuir el agua libre en la terminal petrolera misma, aumentando así los ahorros; por lo anterior, al asumir que los ahorros son proporcionales al almacenamiento gestionado por la nueva política operacional, al implementarse en las demás estaciones estos podrían alcanzar \$ 1,2 MUSD traídos al presente, sin tener en cuenta el precio de nuevos desarrollos o la actualización de estos.

Para continuar con el alineamiento estratégico se propone considerar optimizar otros procesos usando herramientas de simulación y de inteligencia artificial que permitan obtener diagnósticos de valor y planear la operación con herramientas prescriptivas que reduzcan drásticamente los costos operativos.

### **Conclusiones**

A partir del BPM y el diagrama Ishikawa aplicados para caracterizar la política operacional implementada se logran determinar los problemas asociados a la operación. Problemas operativos como extensos tiempos de conexión al buque, bombeos lentos y uso

inadecuado del recurso humano, estos son abordados al momento de desarrollar las nuevas políticas operacionales evaluadas.

Se logra desarrollar una herramienta de predicción que representa en gran medida ( $R^2=0,991$ ) el comportamiento del agua libre final en los tanques según su estado final siguiendo la metodología de CRISP-DM la cual es muy fácil de aplicar. Esta herramienta es de gran utilidad para la simulación de la operación en Monte Carlo y para el desarrollo de la propuesta operacional seleccionada. La operación presenta como resultado un sobre costo de 8,6 millones de dólares en 10 años.

El uso de herramientas de apoyo a la decisión con inteligencia artificial en la operación puede reducir los sobrecostos y generar ahorros importantes con una sola política aplicada; además permite que la operación de despacho y drenaje de tanques se alinee al objetivo estratégico. Por medio de la política operacional de despacho y drenaje propuesta aplicada se puede reducir en 0,74 millones de dólares (VPN) los sobrecostos operacionales en la terminal petrolera y el aplicar esta política en las otras 6 estaciones permitiría amplificar los ahorros de la política operacional de despacho y drenaje, beneficiando principalmente al objetivo de la “utilidad operacional sostenida” y a su vez el mapa estratégico empresarial. También se afectan otros objetivos organizacionales positivamente como el impacto positivo al medio ambiente, un modelo organizacional óptimo y tecnología e innovación propulsoras de transformación. Además, la reducción del tiempo de operación mitiga el riesgo de multas costosas por conexiones extensas con buques.

### **Investigación futura**

Desarrollar la herramienta prescriptiva propuesta en la política seleccionada, de manera que sea funcional para la operación, y que por medio de predicciones recomiende la planeación

de la operación de drenaje y el llamado oportuno de cuadrillas; como sistema de apoyo a la decisión se basaría en la predicción del agua libre final en los tanques. Además, se propone un nuevo desarrollo que permita que esta se actualice automáticamente según los datos de la operación; esta actualización permitiría a la herramienta mantener o mejorar su exactitud ante los cambios de los productos en años posteriores.

Según los resultados de la política seleccionada se encuentra que la reducción de horas en la llegada de la segunda cuadrilla es importante, por lo que si se logra predecir el agua libre inicial aun antes de que el bache llegue a la estación se podría eliminar este tiempo. En este caso de trabajo futuro se lograría una mejora a la política propuesta, al tener un sobre costo de operación de \$4110 USD (Tabla 13), es decir un ahorro del 14,3% (\$1,24MUSD sumados en 10 años).

Tabla 13

*Tabla de resultado de simulación política futura. (USD).*

<b>Costo</b>	<b>M</b>	<b>P</b>	<b>Ponderado</b>
Conexiones extensas	\$ 1.830	\$ 4.332	\$ 2.330
Tiempos de bombeo lentos	\$ 1.942	\$ 872	\$ 1.728
Costo de personal	\$ 50	\$ 56	\$ 51
<b>Total</b>	<b>\$ 3.822</b>	<b>\$ 5.261</b>	<b>\$ 4.110</b>

Se propone desarrollar un plan de implementación para aplicar a otros procesos herramientas de simulación y de inteligencia artificial que permitan obtener diagnósticos de valor y planear la operación con herramientas prescriptivas que reduzcan drásticamente los costos operativos.

Lo mencionado anteriormente nos lleva a las siguientes preguntas de investigación como trabajo futuro:

¿Cómo construir la herramienta propuesta de manera que sea funcional para la operación y cuál es la exactitud del nuevo modelo de actualización para la herramienta desarrollada?

¿Cómo predecir el agua libre final en los tanques antes de que el producto llegue a la estación?

¿Cuáles nuevas iniciativas o políticas pueden reducir los costos operativos usando la inteligencia artificial?

### Bibliografía

- American Petroleum Institute. (1994). Manual de Estándares de Medición de Petróleo Capítulo 1 — Vocabulario.
- Anggraeni, D., Sugiyanto, K., Zam, M. I. Z., & Patria, H. (2022). Stock price movement prediction using supervised machine learning algorithm: KNIME. *Jurnal Akun Nabelo: Jurnal Akuntansi Netral, Akuntabel, Objektif*, 4(2), 671-681.
- Barrantes Aguilar, L. (2019). Diferencias en la estimación del coeficiente de curtosis en diferentes softwares estadísticos. *Revista E-Agronegocios*, 5(2). <https://doi.org/10.18845/rea.v5i2.4456>
- Bustamante Zulueta, N. W., & Sosa Vergara, J. C. (2022). Modelo de optimización basado en la metodología Lean Manufacturing y PDCA para mejorar la productividad del proceso de ejecución de enchapado en un edificio multifamiliar [Universidad Tecnológica del Perú]. [https://repositorio.utp.edu.pe/bitstream/handle/20.500.12867/7142/N.Bustamante\\_J.Sosa\\_Tesis\\_Titulo\\_Profesional\\_2022.pdf?sequence=1&isAllowed=y](https://repositorio.utp.edu.pe/bitstream/handle/20.500.12867/7142/N.Bustamante_J.Sosa_Tesis_Titulo_Profesional_2022.pdf?sequence=1&isAllowed=y)
- Campbell, B. (2005). Alignment: Resolving Ambiguity within Bounded Choices. PACIS.
- Casson, R. J., & Farmer, L. D. (2014). Understanding and checking the assumptions of linear regression: A primer for medical researchers. *Clinical and Experimental Ophthalmology*, 590-596.
- Chan, C. W. (2005). An expert decision support system for monitoring and diagnosis of petroleum production and separation processes. *Expert Systems with Applications*, 29(1), 131–143. <https://doi.org/https://doi.org/10.1016/j.eswa.2005.01.009>
- Chan, Y. E., & Reich, B. H. (2007). IT alignment: What have we learned? *Journal of Information Technology*, 22(4), 297–315. <https://doi.org/10.1057/palgrave.jit.2000109>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Daimlerchrysler, T. R., Shearer, C., & Daimlerchrysler, R. W. (2000). Step-by-step Data Mining Guide. SPSS Inc, 78, 1–78.

- Clavijo, N., Melo, A., Câmara, M. M., Feital, T., Anzai, T. K., Diehl, F. C., Thompson, P. H., & Pinto, J. C. (2019). Development and Application of a Data-Driven System for Sensor Fault Diagnosis in an Oil Processing Plant. In *Processes* (Vol. 7, Issue 7). <https://doi.org/10.3390/pr7070436>
- Colli, L. F. M., da Silva, L. C. R. P., de Sousa, V. P., de Padula, M., & Cabral, L. M. (2019). Evaluation of the effectiveness of the notification process in the area of health products. *Health Policy and Technology*, 8(2), 105–110. <https://doi.org/https://doi.org/10.1016/j.hlpt.2019.03.005>
- Delgado, D., Cobo, D., Pérez, K., Pilacuan, R., & Rocha, M. (2021). El Diagrama De Ishikawa Como Herramienta De Calidad En La Educación: Una Revisión De Los Últimos 7 Años. *Tambara*, 84, 1212–1230.
- Gilli, M., Maringer, D., & Schumann, E. (2019). Chapter 6 - Generating random numbers (M. Gilli, D. Maringer, & E. B. T.-N. M. and O. in F. (Second E. Schumann (eds.); pp. 103–132). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-815065-8.00017-0>
- Henderson, J. C., & Venkatraman, H. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, 32(1), 472–484. <https://doi.org/10.1147/sj.382.0472>
- IBM. (2021). What Is Business Process Modeling? <https://www.ibm.com/cloud/blog/business-process-modeling>
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685–695. <https://doi.org/10.1007/s12525-021-00475-2>
- Khabbazi, M. R., Hasan, M. K., Sulaiman, R., & Shapi'i, A. (2013). Business process modelling in production logistics: Complementary use of BPMN and UML. *Middle East Journal of Scientific Research*, 15(4), 516–529. <https://doi.org/10.5829/idosi.mejsr.2013.15.4.2280>
- Knime. (2022, 4 20). Knime Hub/Nodes/Naive Bayes Learner. Retrieved from <https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.mine.bayes.naivebayes.learner3.NaiveBayesLearnerNodeFactory4>
- Knime. (2022, 4 20). Knime Hub/Nodes/PNN Learner (DDA). Retrieved from <https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.mine.bfn.radial.RadialBasisFunctionLearnerNodeFactory>
- Kumar Tarei, P., Kumar, G., & Ramkumar, M. (2022). A Mean-Variance robust model to minimize operational risk and supply chain cost under aleatory uncertainty: A real-life case application in petroleum supply chain. *Computers & Industrial Engineering*, 166, 107949. <https://doi.org/https://doi.org/10.1016/j.cie.2022.107949>
- Lee, L. C., Liong, C. Y., & Jemain, A. A. (2018). Validity of the best practice in splitting data for hold-out validation strategy as performed on the ink strokes in the context of forensic science. *Microchemical Journal*, 125-133.
- Löfström, T., Bondaletov alexanderbondaletov, A., Redfield, redfieldse A., Henrik Boström, S., Johansson, U., Papadopoulos, H., An Nguyen, K., Boström, H., & Carlsson, L. (2023).

- Tutorial on using Conformal Predictive Systems in KNIME. *Proceedings of Machine Learning Research*, 204, 602–620. <https://proceedings.mlr.press/v204/lofstrom23a.html>
- López Agüí, J. C. (2008). Guía básica para la simulación de Monte Carlo. In *Guía básica para la simulación de Monte Carlo*.
- Mandal, C. S., Agarwal, M., & Ignatius, P. (2022). External floating roof tank boil over: Causes, prevention and management-A comprehensive review. *International Journal of Engineering, Science and Mathematics*, 11(10), 1-18.
- MN, H., & Basheer, I. (2003). Comparison of logistic regression and neural network-based classifiers for bacterial growth. *Food Microbiology*, 43-55.
- Nakhal A, A. J., Patriarca, R., Di Gravio, G., Antonioni, G., & Paltrinieri, N. (2021). Investigating occupational and operational industrial safety data through Business Intelligence and Machine Learning. *Journal of Loss Prevention in the Process Industries*, 73, 104608. <https://doi.org/https://doi.org/10.1016/j.jlp.2021.104608>
- Qian, Y., Xu, L., Li, X., Lin, L., & Kraslawski, A. (2008). LUBRES: An expert system development and implementation for real-time fault diagnosis of a lubricating oil refining process. *Expert Systems with Applications*, 35(3), 1252–1266. <https://doi.org/https://doi.org/10.1016/j.eswa.2007.07.061>
- Reich, B., & Benbasat, I. (1996). Measuring the Linkage between Business and Information Technology Objectives. *MIS Quarterly*, 20, 55–81. <https://doi.org/10.2307/249542>
- Rish, I. (2001). IBM Research Report An empirical study of the naive Bayes classifier. *Computer Science*.
- Rondinelli, D., Rosen, B., & Drori, I. (2001). The struggle for strategic alignment in multinational corporations:: Managing readjustment during global expansion. *European Management Journal*, 19(4), 404–416. <https://doi.org/DOI:> ,
- Sortino, F., van der Meer, R., Plantinga, A., & Kuan, B. (2010). Chapter 3 - Beyond the Sortino Ratio (F. B. T.-T. S. F. for C. P. Sortino (ed.); pp. 23–52). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-12-374992-5.00003-X>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/https://doi.org/10.1016/j.procs.2021.01.199>
- Tawhid, A., Teotia, T., & Elmiligi, H. (2021). Chapter 13 - Machine learning for optimizing healthcare resources. In P. Kumar, Y. Kumar, & M. A. Tawhid, *Machine Learning, Big Data, and IoT for Medical Informatics* (pp. Pages 215-239). Barcelona: ACADEMIC PRESS.
- Vu, H. L., Ng, K. T., Richter, A., & An, C. (2022). Analysis of input set characteristics and variances on k-fold cross validation for a Recurrent Neural Network model on waste disposal rate estimation. *Journal of Environmental Management*, 114869.
- Weisberg, S. (2005). *Applied Linear Regression*. Hoboken: John Wiley & Sons, Inc.

## Apéndice A.

### Desarrollo de herramienta de predicción de agua libre final.

#### 1. *Comprensión del negocio*

Específicamente para el desarrollo de la herramienta el objetivo de negocio es reducir el tiempo de residencia del agua en los tanques por medio de la predicción del agua libre final en los tanques usando un modelo con machine learning; se define como criterio de éxito de negocio predecir por lo menos con un 70% de exactitud el agua libre final.

El objetivo de la minería de datos es:

- Predecir el agua decantada final en los tanques en operación.

El criterio de éxito de minería de datos se define como:

- Una exactitud mayor al 70% para la predicción del agua libre final en los tanques. (MAPE<30%)

#### 2. *Comprensión de los datos*

Se usan datos históricos de un año de operación de los tanques, en donde se tiene el ultimo recibo en tanques (estado inicial) y el estado antes de exportar (estado final). Luego de depurar los datos se concluye con 327 datos y 16 atributos. Apéndice B.

Es realizada una correlación de datos en Knime. Las correlaciones positivas mayores a 0,90 son obvias, pues la fecha inicial y final, el agua libre inicial y final, el volumen del tanque (TOV) Inicial y final y el %SW inicial y final están directamente relacionados. Las demás correlaciones son débiles o nulas.

Al realizar el análisis estadístico se encuentra que los volúmenes de tanque iniciales y finales en mínimos y máximos son cercanos; mientras que en el agua libre son idénticos los rangos, aunque la media sube para el agua libre final, esto se debe a la decantación; para el % SW inicial y final se observa que la media, dimensión y rango de datos bajan para el %SW final,

en términos de negocio es bueno pues significa que disminuyó la cantidad de agua en suspensión pasando a agua libre.

Al analizar los datos se refleja que hay tanques con movimientos, es decir, recibos, drenajes o ambos entre las mediciones; aunque los datos obtenidos muestran tanques con movimientos entre las mediciones, siguen siendo importantes para los objetivos de minería de datos pues los datos reflejan la operación en la terminal petrolera; es decir que el modelo al momento de predecir puede tener en cuenta posibles recibos y drenajes que no se ven reflejados entre las dos mediciones. Se encuentran inconsistencias en el significado de los datos, como disminución en el volumen de agua aun cuando el volumen del tanque no cambia, esto muestra que existen errores en la medición inicial, tales errores de medición se asumirán en la herramienta para que en la predicción se tenga en cuenta la operación incluyendo los errores.

### **3. Preparación de los datos**

Para seleccionar los atributos de los dataset a usar en el modelamiento, los atributos se identifican como variables de entrada, de salida u objetivo, Apéndice C. Las variables seleccionadas para el modelo en un inicio son las de entrada y objetivo, estos atributos se tienen al inicio de la operación en la primera medición del tanque, a excepción del tiempo transcurrido que también es agregado. Luego las variables para los modelos se escogen a partir de los métodos de “forward feature selection” y “backward feature elimination” aplicados a distintos modelos predictivos usando Knime. Además, se eliminan los datos atípicos del data set.

Para realizar la selección de atributos es necesario usar los modelos adecuados para los datos, por ello en la etapa de modelamiento estos son escogidos: Simple regression tree, gradient boosted trees (regression), random forest (regression) y tree ensemble (regression).

Como se observa en el diagrama de selección de atributos para el modelo de agua libre (Figura A1) existen dos secciones, una para probar con el método “forward feature selection” y

otra para probar el método de “backward feature elimination”. Cada metanodo tiene la estructura descrita en el diagrama mostrado en la figura A2.

El orden de los datos afecta el resultado de las técnicas de modelamiento al cambiar los datos tomados en cada muestreo; por lo que, para evitar este problema al evaluar un modelo en la selección de datos, cada modelo es validado 20 veces tomando muestras aleatorias por medio del nodo “Partitioner” en Knime, el cual usa la técnica de validación cruzada (Vu, Ng, Richter, & An, 2022).

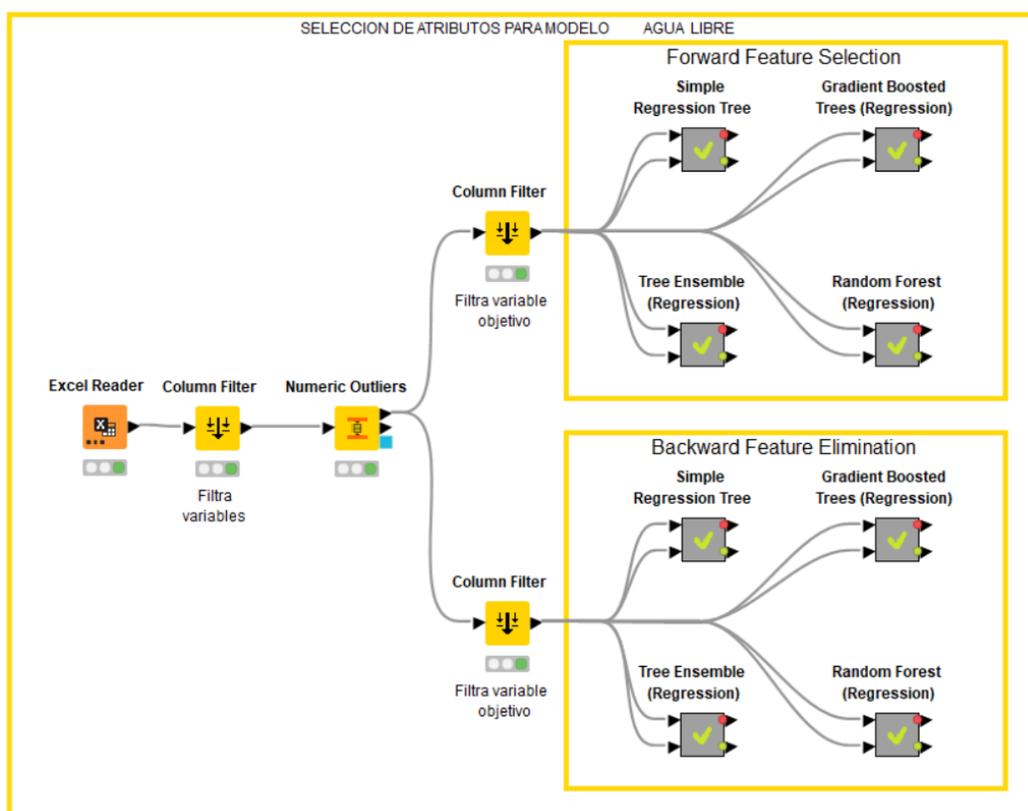


Figura A1. Diagrama de selección de atributos para el modelo de agua libre

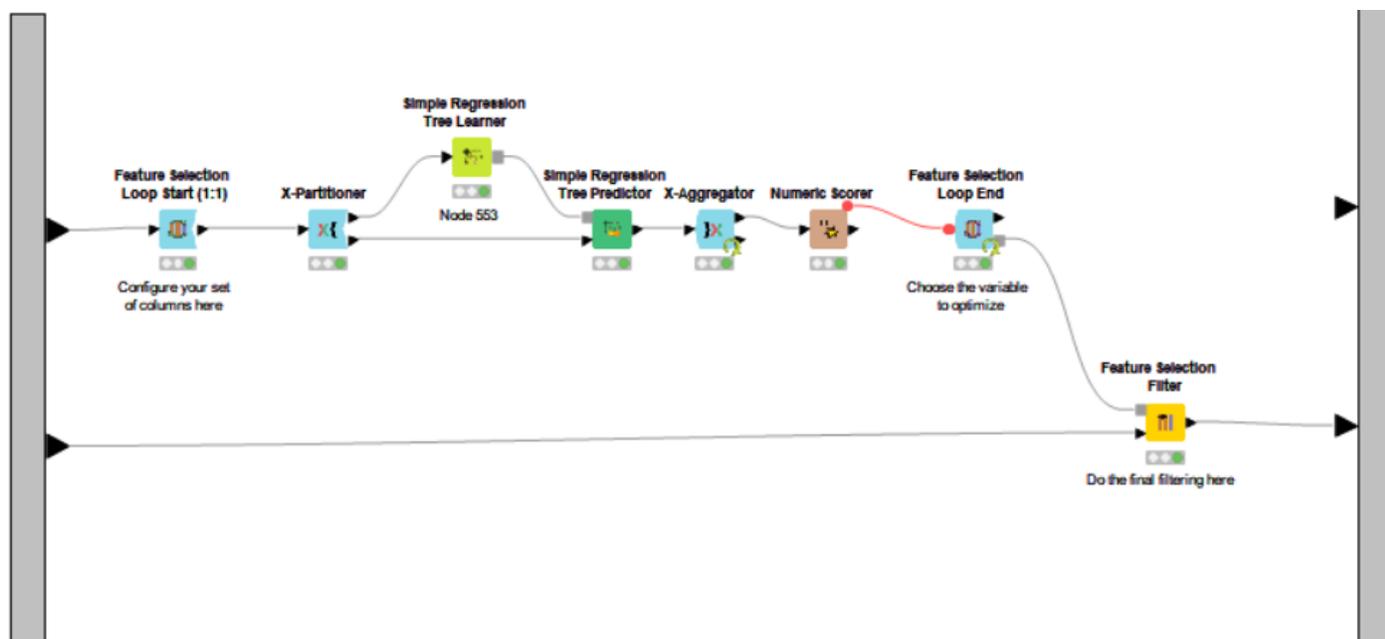


Figura A2. Diagrama de metanodo de selección de atributos

El modelo con los mejores resultados es “random forest regression” en el método de backward feature elimination con un RMSE de 997,4, donde se obtienen cuatro variables: Hora, Tipo de medición, agua libre inicial y %SW inicial. Por lo que los atributos para el modelamiento son 5 (contando la variable objetivo agua libre final), teniendo, así como dataset final 303 datos con 5 atributos.

#### 4. Modelamiento

inicialmente se presentan 6 técnicas de modelado preliminares entre las cuales se escogen las más adecuadas: Red neuronal probabilística (PNN), simple regression tree, gradient boosted trees (regression), random forest (regression), tree ensemble (regression) y linear regression; debido a que estos 6 modelos pueden tratar con atributos nominales y numéricos de entrada y con un atributo objetivo numérico, los demás modelos en Knime no pueden tratar con los atributos de esta manera o predicen valores faltantes con los datos dados. Entre las técnicas

seleccionadas preliminarmente se encuentra que los modelos de Naive Bayes, redes neuronales probabilísticas (PNN) y regresión lineal tienen suposiciones que no se cumplen para los datos.

El modelo de Naive Bayes asume que cada variable predictora es independiente (condición de Markov); aunque al no cumplirse la suposición el modelo sigue presentando buenos resultados (Rish, 2001). Además, en Knime al usarse este modelo para las variables numéricas se calcula la distribución gaussiana, la cual no se cumple para todos los datos (Knime, 2022), Por lo que no es recomendable usar Naive Bayes como técnica de modelado.

El modelo de redes neuronales probabilísticas (PNN) asume una distribución gaussiana debido a la clasificación de Bayes (MN & Basheer, 2003), en Knime el modelo se basa en el método DDA (Dynamic Decay Adjustment) donde las reglas generadas por los atributos numéricos son definidas por funciones gaussianas de alta dimensionalidad (Knime, 2022). No todas las variables presentan una distribución normal o gaussiana como el %SW inicial, Por lo que no es recomendable usar PNN como técnica de modelado.

El modelo de regresión lineal presenta diferentes suposiciones (Casson & Farmer, 2014). La suposición de linealidad entre las variables predictoras y la de resultado no se cumple para todas las variables, observando la correlación lineal entre el agua libre final y las demás variables. Aunque el agua libre inicial tiene una correlación lineal fuerte con el agua libre final, no se cumple la suposición de normalidad al observar que los residuales no tienen una distribución normal al usar como variable predictora el agua libre inicial; al calcular su curtosis es de 1,038 lo cual confirma que no sigue una distribución normal (Barrantes, 2019).

Por lo tanto, las técnicas de modelado seleccionadas luego de revisar las suposiciones son las siguientes: Simple regression tree, Gradient boosted trees (regression), Random forest (regression), Tree ensemble (regression)

Para validar cada modelo generado se usa la estrategia de “hold out validation” (Lee, Liong, & Jemain, 2018). El dataset es separado aleatoriamente en proporción 70/30, donde el 70% de los datos son para entrenar el modelo y el 30% para validarlo. Como diseño de comprobación, para comprobar la calidad del modelo se usa el coeficiente de determinación y el RMSE (Tawhid, Teotia, & Elmiligi, 2021).

Los modelos son generados en Knime siguiendo los parámetros por defecto dados por la herramienta y validados siguiendo el diagrama de la figura A3. En la figura A4 se observa el diagrama de flujo usado en Knime para probar las 4 técnicas de modelado seleccionadas.

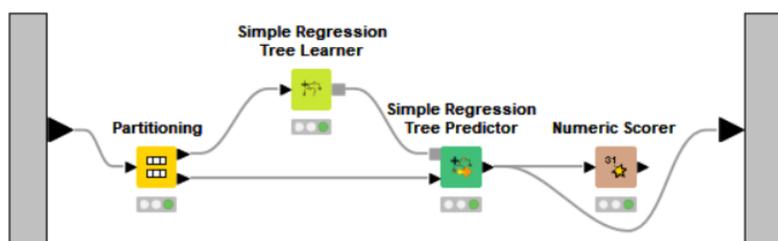


Figura A3. Diagrama metanodo de validación de modelo.

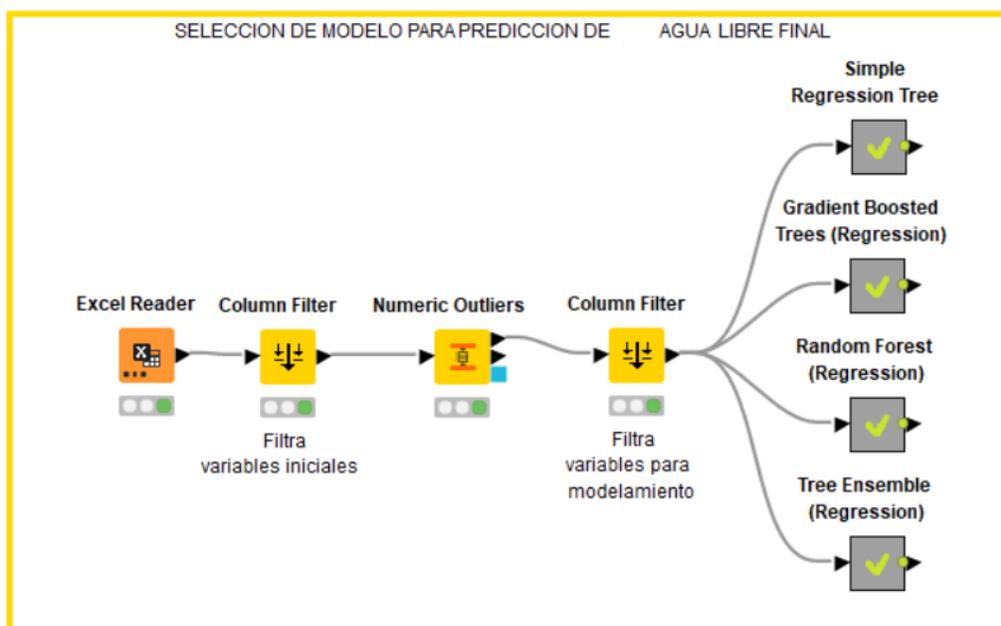


Figura A4. Diagrama de proceso de selección de modelo de agua libre final

Luego de generar los modelos estos se califican con las métricas mencionadas en el diseño de comprobación. los modelos se evalúan con el coeficiente de determinación  $R^2$  y el RMSE. Los resultados de los modelos de regresión para la predicción del agua libre final son buenos para todas las técnicas de modelado probadas; debido a que todas tienen un  $R^2$  cercano a 1 (Tabla A1) lo cual muestra que los modelos representan la mayoría de los datos. El RMSE más bajo y  $R^2$  mas alto corresponden al modelo de gradient boosted trees (regression) siendo así el mejor modelo para ambas métricas.

Tabla A1

*Selección de modelo de predicción de agua libre final.*

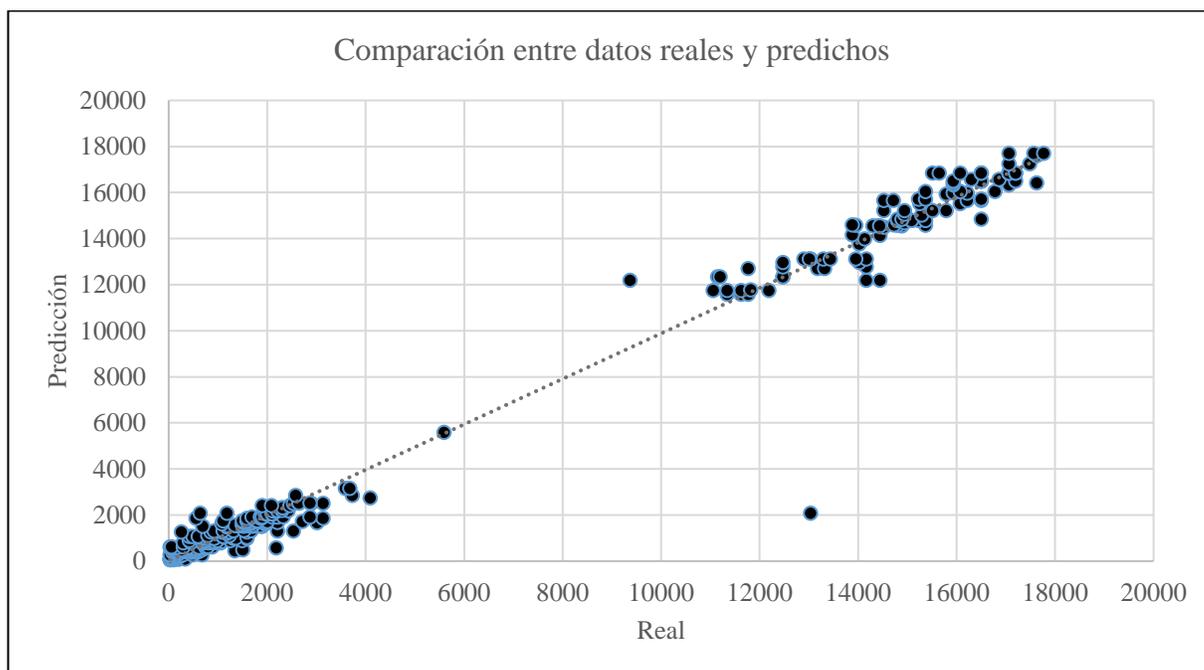
<b>Selección de modelo de predicción de agua libre final</b>		
<b>Modelo</b>	<b><math>R^2</math></b>	<b>RMSE</b>
Simple Regression Tree	0,987	772,129
<b>Gradient Boosted Trees (Regression)</b>	<b>0,991</b>	<b>621,701</b>
Random Forest (Regression)	0,983	876,138
Tree Ensemble (Regression)	0,978	1002,851

Al validar el modelo con datos distintos a los de entrenamiento se encuentra que es más exacto al usar solamente la variable predictora de agua libre inicial al tener un  $R^2$  de 0,991 y un error porcentual absoluto medio (MAPE) del 60,4%, más bajo que el modelo anterior de 4 variables (64,7%), aunque presenta un RMSE un poco más alto de 650,852. Además, al usar esta variable con los otros modelos validados anteriormente, GBT sigue siendo el más exacto.

Tabla A2

*Evaluación de modelo seleccionado (GBT y como variable predictora el agua libre inicial)*

Conjunto de Datos	R2	RMSE	MAPE
Entrenamiento (70%)	0,983	858,308	0,495
Validación (30%)	0,991	650,892	<b>0,604</b>
Todos (100%)	0,985	801,673	0,528
Todos incluyendo Outlayers	0,986	787,607	0,522



*Figura A5. Comparación entre datos reales y predichos.*

Este modelo mejora su exactitud con los datos nuevos, como es demostrado en la validación y aun al tomar los outlayers. Al revisar a detalle los datos evaluados, se observa que el modelo no predice con exactitud los ensayos de un tanque en rangos muy bajos de agua libre final en específico que representan solo el 5% de los datos; en donde si no se tienen en cuenta estos datos el MAPE del modelo llega al 23% lo cual es aceptable, al dar una exactitud del 77%.

### **5. Evaluación**

Como se observa en la etapa anterior se obtuvieron buenos resultados al tener 77% de exactitud en las predicciones. Por lo tanto, el criterio de éxito de minería de datos y el objetivo de negocio se cumplen. En conclusión, la herramienta permite predecir el comportamiento del agua libre final en los tanques y apoyar en la simulación de la operación de despacho y drenaje en la terminal petrolera.

## Apéndice B.

### Atributos iniciales.

Atributos luego del primer tratamiento de datos, se tienen 327 datos con 16 atributos.

Tabla B1

*Descripción de atributos iniciales.*

<b>Atributo</b>	<b>Descripción</b>	<b>Tipo de atributo</b>
Fecha Inicial	Fecha de último movimiento reportado en el tanque	Cuantitativo intervalo
Fecha Final	Fecha de medición previa a exportación	Cuantitativo intervalo
Hora Inicial	Hora de último movimiento reportado en el tanque	Cuantitativo intervalo
Hora Final	Hora de medición previa a exportación	Cuantitativo intervalo
Tipo de medición Inicial	Tipo de medición inicial del tanque, medición directa o ponderada	Cualitativo Binario
Tipo de medición Final	Tipo de medición final del tanque, medición directa o ponderada	Cualitativo Binario
Tanque inicial	Tanque donde esta almacenado el producto al inicio	Cualitativo Nominal
Tanque final	Tanque donde esta almacenado el producto al final	Cualitativo Nominal
Producto inicial	Tipo de producto al inicio, puede ser tres productos distintos	Cualitativo Nominal
Producto final	Tipo de producto al final, puede ser tres productos distintos	Cualitativo Nominal
TOV inicial	Volumen inicial observado del tanque	Cuantitativo continuo
TOV Final	Volumen final observado del tanque	Cuantitativo continuo
Agua libre inicial	Volumen de agua libre inicial en el tanque	Cuantitativo continuo
Agua libre final	Volumen de agua libre final en el tanque	Cuantitativo continuo
%SW inicial	Porcentaje inicial de agua y sedimento	Cuantitativo continuo
%SW Final	Porcentaje final de agua y sedimento	Cuantitativo continuo

## Apéndice C.

### Conjunto de variables según tipo.

Lista de atributos luego de la etapa de comprensión de los datos, del cual se seleccionarán las variables para el modelo.

Tabla C1

*Tipos de atributos para Dataset.*

<b>Variable</b>	<b>Tipo</b>
Fecha Inicial	Variable de entrada
Fecha Final	Variable de salida
Hora Inicial	Variable de entrada
Hora Final	Variable de salida
Tipo de medición	Variable de entrada
Tanque	Variable de entrada
Producto	Variable de entrada
TOV inicial	Variable de entrada
TOV Final	Variable de salida
Agua libre inicial	Variable de entrada
Agua libre final	Variable salida (Variable objetivo)
%SW inicial	Variable de entrada
%SW Final	Variable de salida
Tiempo transcurrido	Variable de salida (calculada)

## Apéndice D.

### Distribuciones para Monte Carlo.

#### Producto M. Tanque pequeño. Agua libre inicial.

Distribución Beta =  $124 + 4,68e+03 * \text{BETA}(0,848, 2,49)$

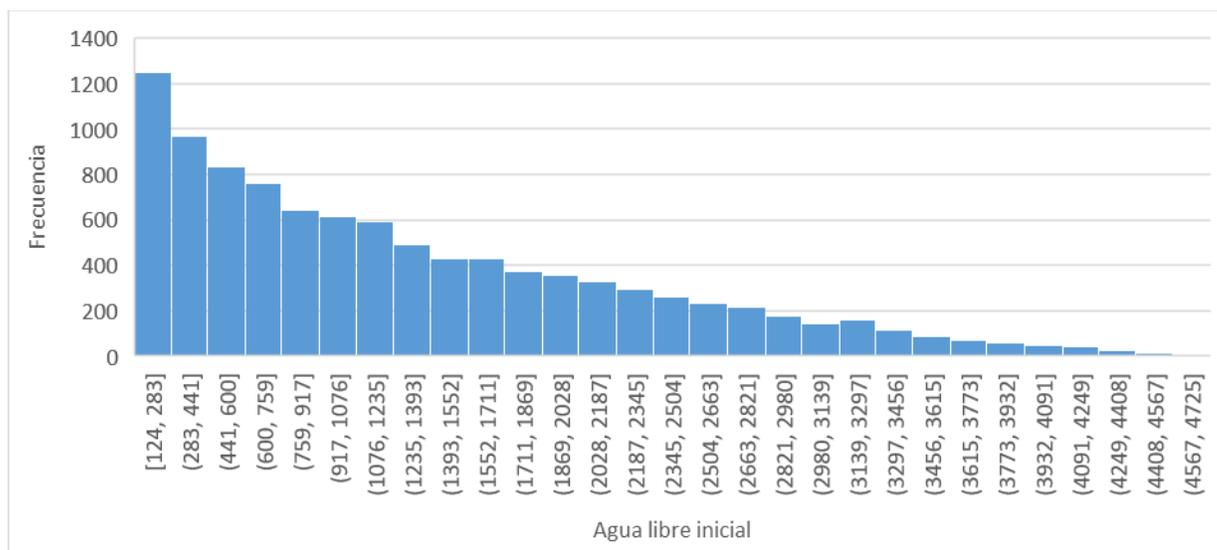
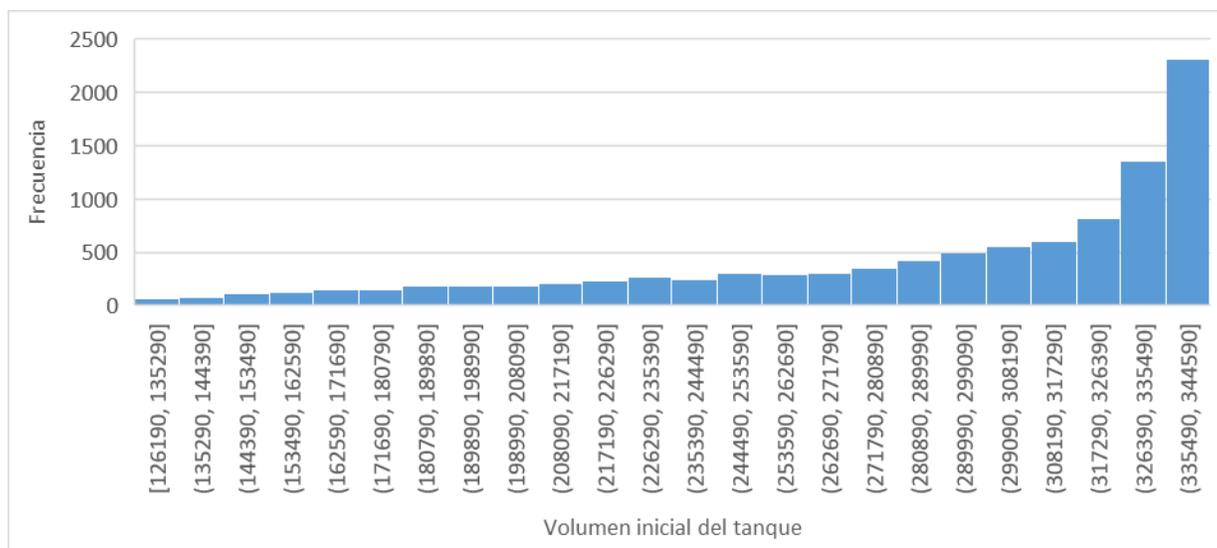


Figura D1. Histograma del agua libre inicial para el Producto M. Tanque pequeño.

### Producto M. Tanque pequeño. Volumen inicial del tanque.

Distribución Beta =  $1,26e+05 + 2,14e+05 * \text{BETA}(1,3, 0,41)$



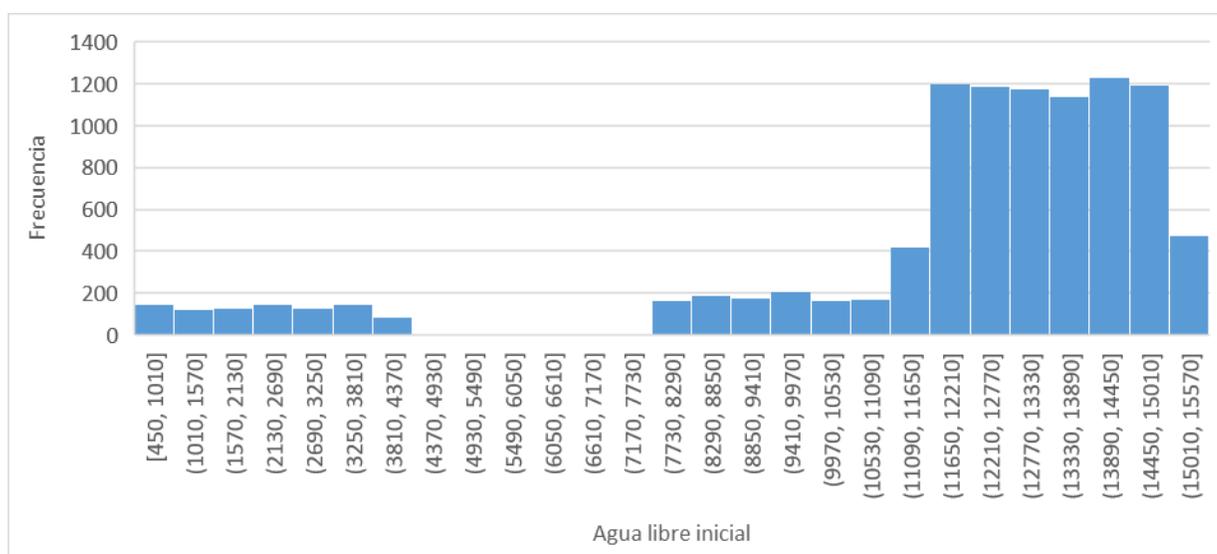
*Figura D2.* Histograma del volumen inicial del tanque para el Producto M. Tanque pequeño.

**Producto M. Tanque grande. Agua libre inicial.**

Tabla D1

*Distribución empírica del agua libre inicial para el Producto M. Tanque grande.*

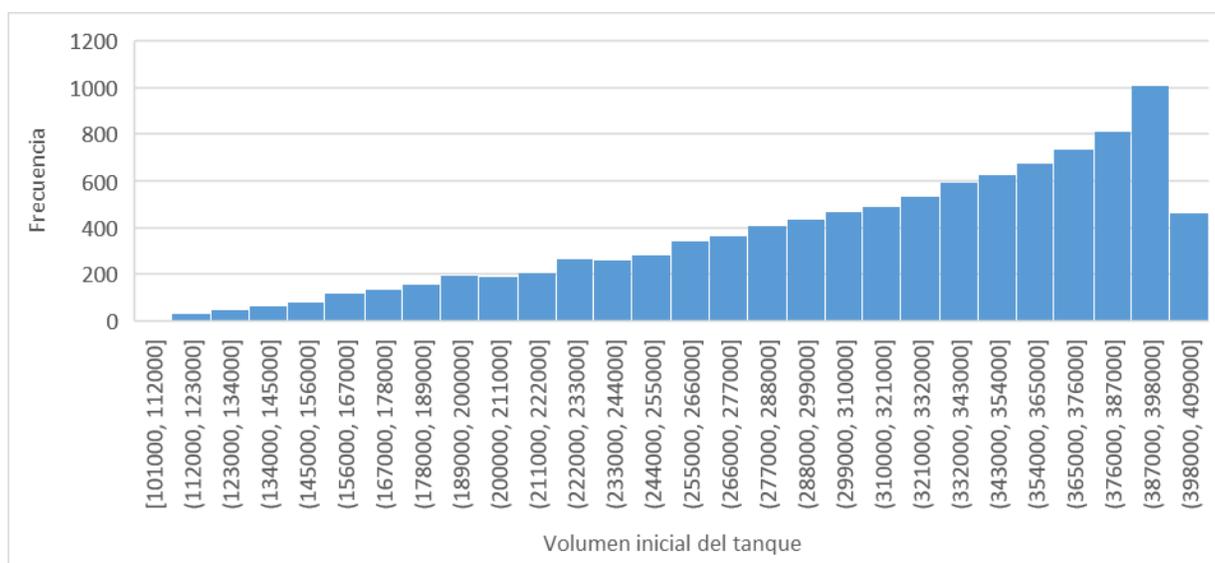
Intervalo		P
Min	Max	
439,34	4139,34	0,10
4139,34	7839,34	0,00
7839,34	11539,34	0,12
11539,34	15239,34	0,78



*Figura D3. Histograma del agua libre inicial para el Producto M. Tanque grande.*

**Producto M. Tanque grande. Volumen inicial del tanque.**

Distribución Beta =  $9,97e+04 + 3,02e+05 * \text{BETA}(2,07, 0,85)$



*Figura D4.* Histograma del volumen inicial del tanque para el Producto M. Tanque grande.

### Producto P. Tanque 1. Agua libre inicial.

Distribución Exponencial =  $27 + \text{EXPO}(317)$

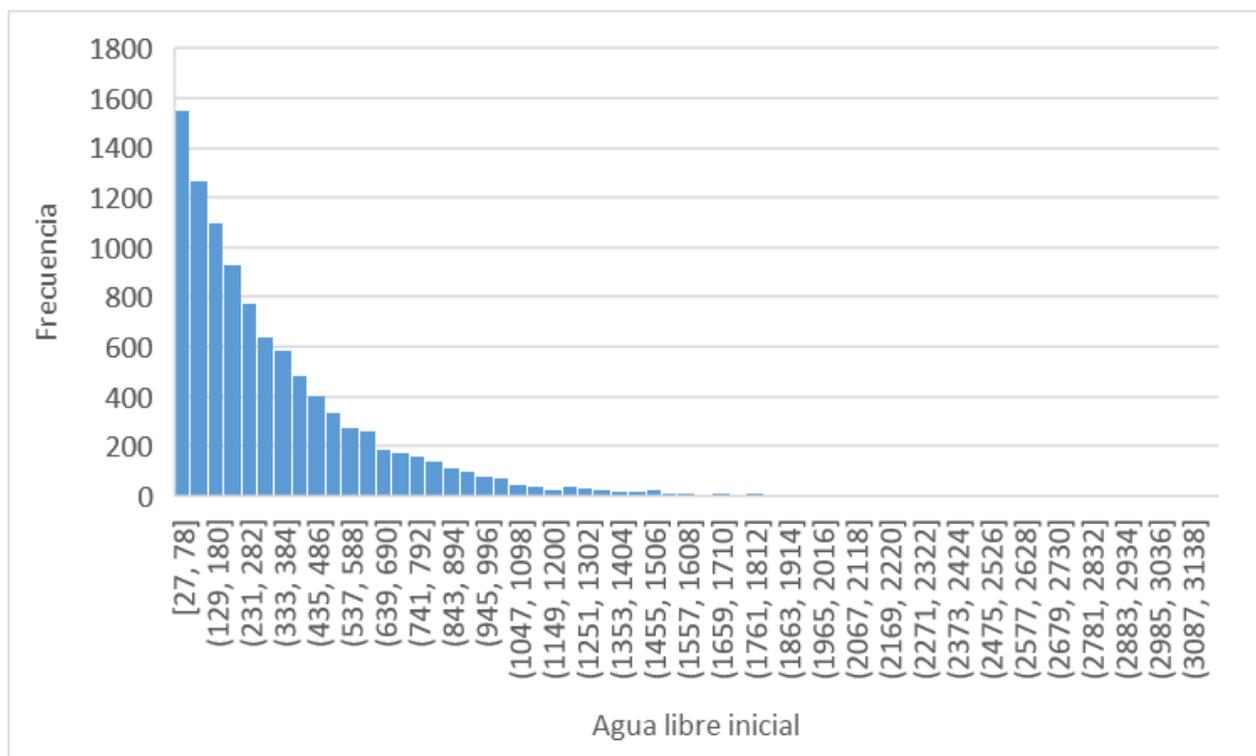


Figura D5. Histograma del agua libre inicial para el Producto P. Tanque 1.

### Producto P. Tanque 1. Volumen inicial del tanque.

Distribución Beta =  $1,7e+05 + 3,52e+05 * \text{BETA}(0,483 , 0,26)$

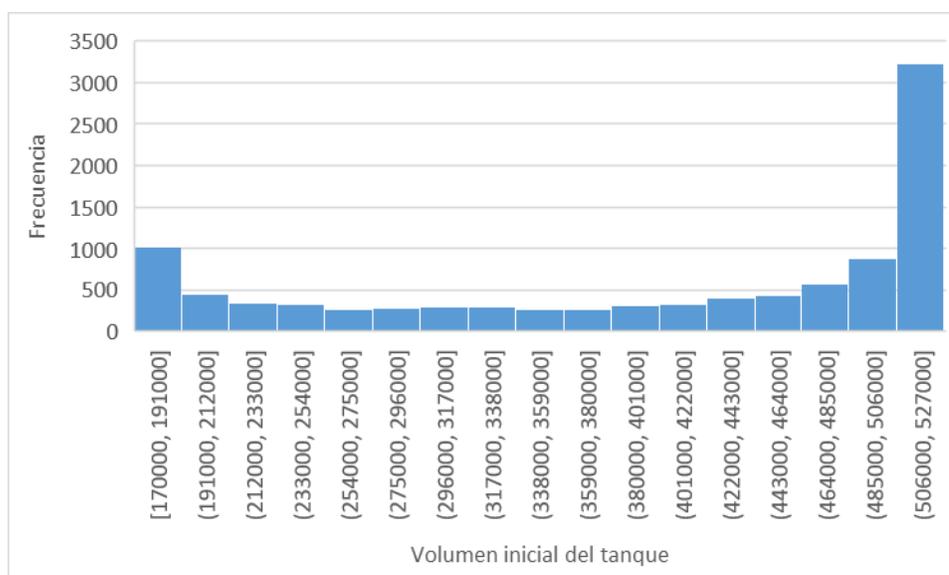


Figura D6. Histograma del volumen inicial del tanque para el Producto P. Tanque 1.

### Producto P. Tanque 2. Agua libre inicial.

Distribución Normal =  $NORM(1,62e+03, 249)$

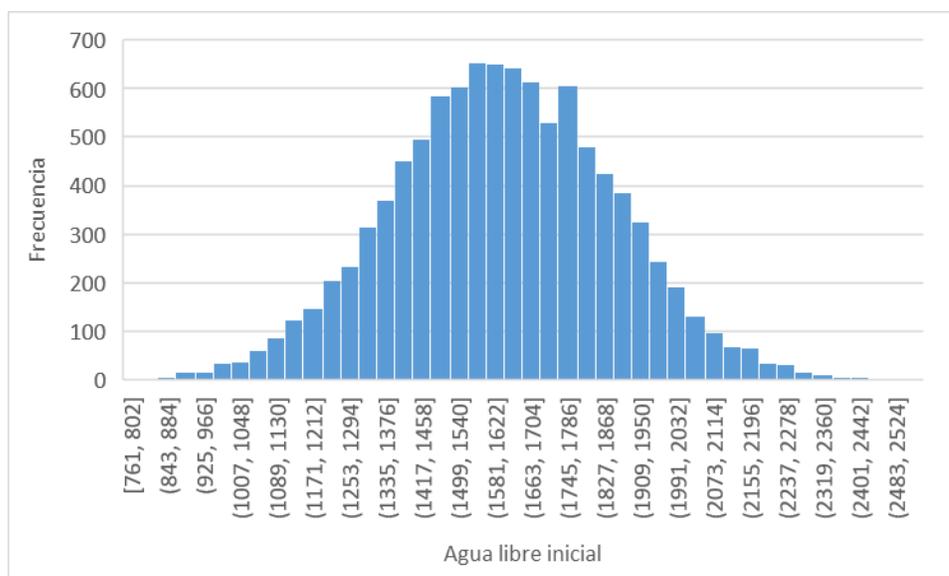


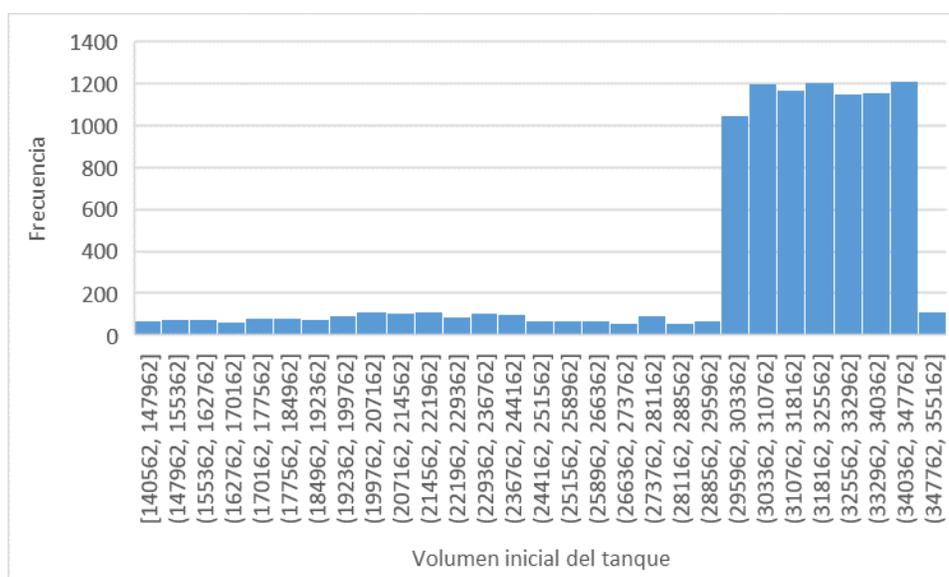
Figura D7. Histograma del agua libre inicial para el Producto P. Tanque 2.

## Producto P. Tanque 2. Volumen inicial del tanque.

Tabla D2

*Distribución empírica del volumen inicial del tanque para el Producto P. Tanque 2.*

Intervalo		P
Min	Max	
140553,37	192553,37	0,05
192553,37	244553,37	0,08
244553,37	296553,37	0,05
296553,37	348553,37	0,83



*Figura D8.* Histograma del volumen inicial del tanque para el Producto P. Tanque 2.

### Producto P. Tanque 3. Agua libre inicial.

Distribución Beta =  $1,21e+04 + 5,65e+03 * \text{BETA}(2,92, 1,38)$

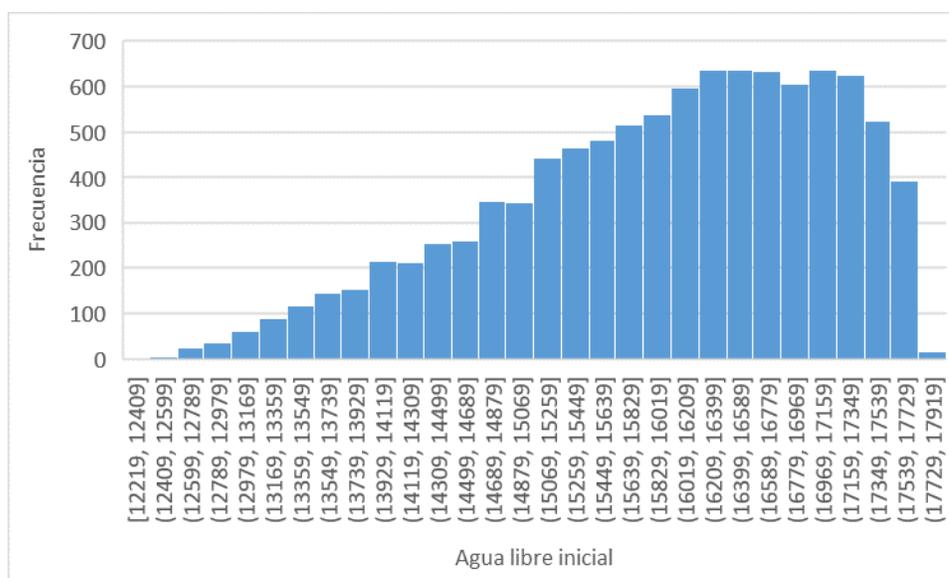


Figura D9. Histograma del agua libre inicial para el Producto P. Tanque 3.

### Producto P. Tanque 3. Volumen inicial del tanque.

Distribución Beta =  $1,49e+05 + 2,54e+05 * \text{BETA}(0,927, 0,52)$

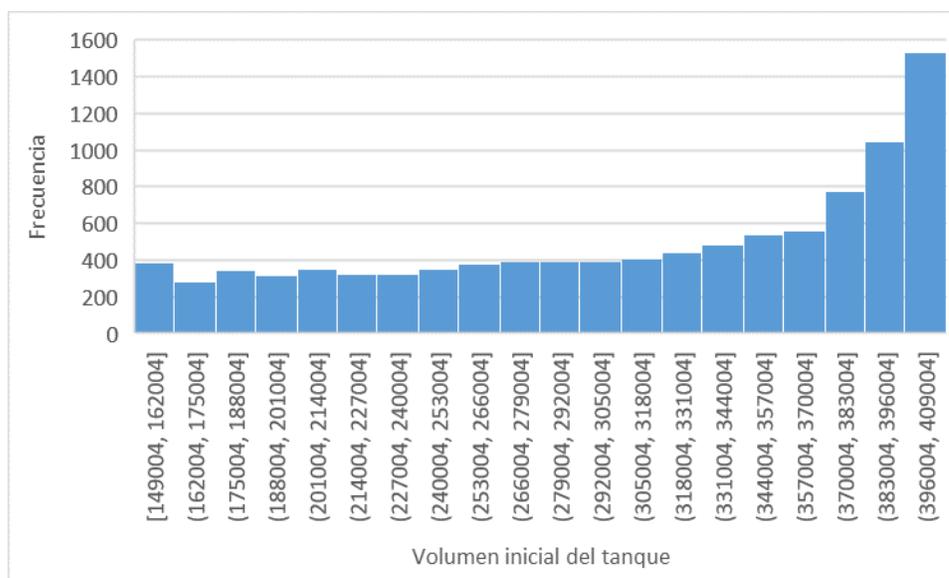


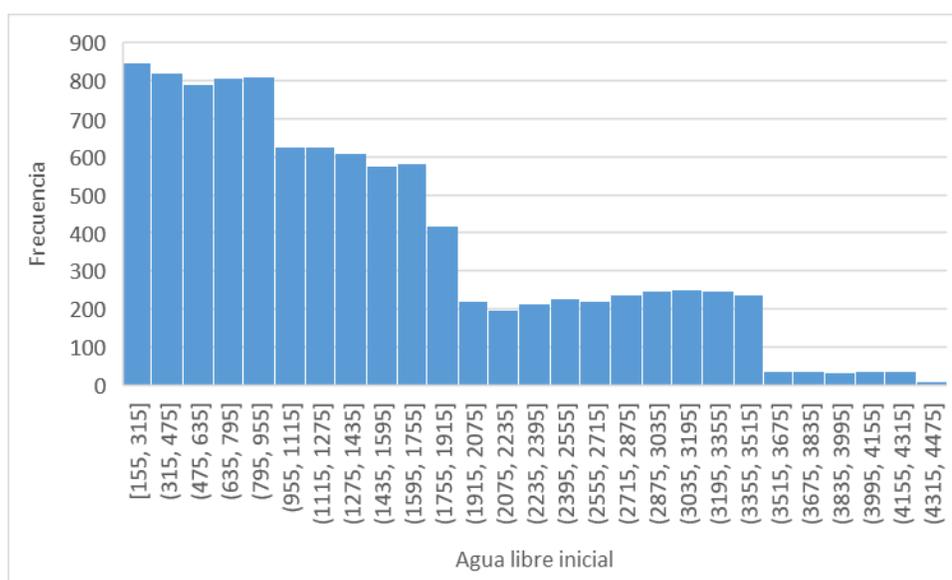
Figura D10. Histograma del volumen inicial del tanque para el Producto P. Tanque 3.

### Producto P. Tanque 4. Agua libre inicial.

Tabla D3

*Distribución empírica del agua libre inicial para el Producto P. Tanque 4.*

Intervalo		P
Min	Max	
154,34	994,34	0,43
994,34	1834,34	0,74
1834,34	2674,34	0,85
2674,34	3514,34	0,98
3514,34	4354,34	1,00



*Figura D11.* Histograma del agua libre inicial para el Producto P. Tanque 4.

### Producto P. Tanque 4. Volumen inicial del tanque.

Distribución Beta =  $9,73e+04 + 3,04e+05 * \text{BETA}(0,889, 0,3)$

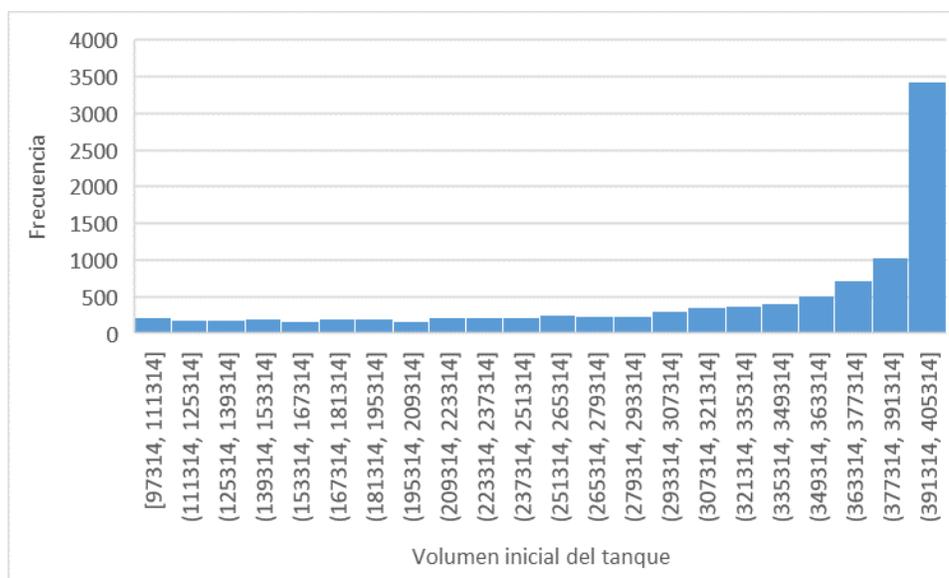


Figura D12. Histograma del volumen inicial del tanque para el Producto P. Tanque 4.

## Apéndice E.

### Fórmulas de KPI's.

Tabla E1

*Fórmulas de KPI's*

KPI	Formula
Exactitud del modelo* Aprobado si $\geq 77\%$	Promedio $(1 -  \text{agua libre final predicha} - \text{agua libre final real de tanque}  / \text{agua libre final real de tanque}) * 100\%$
Porcentaje de computadores instalados  Aprobado si es 100%	$(\text{Computadores con herramienta instalada} / \text{Computadores en la operación}) * 100\%$
Porcentaje de operadores capacitados  Aprobado si es 100%	$(\text{Operadores capacitados} / \text{Operadores}) * 100\%$
Seguimiento de exactitud del modelo. Aprobado si $\geq 77\%$ . Para cada mes	Promedio $(1 -  \text{agua libre final predicha} - \text{agua libre final real de tanque}  / \text{agua libre final real de tanque}) * 100\%$  -Medición mensual
Implementación de política (DDS) Aprobado si es 100% Para último mes	$(\text{llamada de segunda cuadrilla por herramienta} / \text{llamada de segunda cuadrilla}) * 100\%$  -Medición mensual
Implementación de política (Cambio de producto) Aprobado si es 100% Para último mes	$(\text{llamadas de cuadrilla de apoyo antes de cambio} / \text{cambio de producto de exportación}) * 100\%$  -Medición mensual

Impacto a la operación (Ahorro%) Aprobado si > 0% Para último mes	$\frac{((\text{Sobrecosto anterior} - \text{Sobrecosto real}) / \text{Sobrecosto anterior}) * 100\%}{\text{Medición mensual}}$
---	--