

**UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ**

**Colegio de Ciencias e Ingenierías**

**Mejorar el proceso de la asignación de SKUs a los canales de los  
A-Frames**

**en el Centro de Distribución de Corporación GPF**

**Camilo Ignacio Acosta Acosta**

**Sabrina Nicoll Cárdenas Padilla**

**Daniela Del Rocío Castro Benalcázar**

**Ingeniería Industrial**

Trabajo de fin de carrera presentado como requisito  
para la obtención del título de  
**INGENIERO INDUSTRIAL**

Quito, 11 de diciembre de 2024

**Universidad San Francisco de Quito USFQ**

**Colegio de Ciencias e Ingenierías**

**HOJA DE CALIFICACIÓN  
DE TRABAJO DE FIN DE CARRERA**

**Mejorar el proceso de la asignación de SKUs a los canales de los A-Frames  
en el Centro de Distribución de Corporación GPF**

**Camilo Ignacio Acosta Acosta  
Sabrina Nicoll Cárdenas Padilla  
Daniela Del Rocío Castro Benalcázar**

**Nombre del profesor, Título académico**

Carlos Suárez, PhD

Quito, 11 de diciembre de 2024

## © DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Camilo Ignacio Acosta Acosta

Código: 323454

Cédula de identidad: 1725959355

Nombres y apellidos: Sabrina Nicoll Cárdenas Padilla

Código: 216306

Cédula de identidad: 1755012018

Nombres y apellidos: Daniela Del Rocío Castro Benalcázar

Código: 213658

Cédula de identidad: 1726436296

Lugar y fecha: Quito, 11 de diciembre del 2024

## **ACLARACIÓN PARA PUBLICACIÓN**

**Nota:** El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETheses>.

## **UNPUBLISHED DOCUMENT**

**Note:** The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETheses>.

## RESUMEN

Este trabajo de fin de carrera presenta una propuesta para mejorar el proceso de asignación de SKUs a los canales de los A-Frames en el Centro de Distribución de la Corporación GPF. Se identifica una disminución en el porcentaje de líneas despachadas por los A-Frames, pasando del 40% al 34.2%, lo que sugiere un desaprovechamiento de las capacidades de estos equipos automatizados. La metodología empleada integra la complementariedad de DMAIC (Definir, Medir, Analizar, Mejorar, Controlar) y CRISP-DM (Cross-Industry Standard Process for Data Mining) para abordar el problema de manera estructurada.

El estudio utiliza herramientas como Gemba Walk, análisis FODA y clasificación ABC para diagnosticar la situación actual. Se propone una aplicación en Python que automatiza el proceso de selección de cambios de ubicación de SKUs, mejorando la eficiencia operativa. Además, se implementan modelos de aprendizaje automático, incluyendo Random Forest y Gradient Boosting, KNN y Naive Bayes para optimizar la asignación de productos.

Los resultados indican que la automatización del proceso actual y el uso de algoritmos de aprendizaje supervisado pueden aumentar significativamente las líneas despachadas. Este trabajo no solo busca resolver el problema inmediato, sino también establecer un marco para futuras aplicaciones en la gestión de inventarios y la mejora continua en la Corporación GPF.

*Palabras Clave:* Centro Distribución, A-Frame, Picking Automático, Metodología DMAIC, Minería de Datos, Aprendizaje Automatizado, Mejora Continua.

## **ABSTRACT**

This final project presents a proposal to improve the process of assigning SKUs to the channels of the A-Frames at the Distribution Center of the GPF Corporation. A decrease in the percentage of lines dispatched by the A-Frames is identified, dropping from 40% to 34.2%, which suggests underutilization of the capabilities of these automated systems. The methodology employed integrates DMAIC (Define, Measure, Analyze, Improve, Control) and CRISP-DM (Cross-Industry Standard Process for Data Mining) to address the problem in a structured manner.

The study uses tools such as Gemba Walk, SWOT analysis, and ABC classification to diagnose the current situation. A Python application is proposed to automate the process of selecting SKU location changes, improving operational efficiency. Additionally, machine learning models, including Random Forest and Gradient Boosting, are implemented to optimize product allocation.

The results indicate that automating the current process and using supervised learning algorithms can significantly increase the dispatched lines. This work not only aims to solve the immediate problem but also to establish a framework for future applications in inventory management and continuous improvement within GPF Corporation

**KeyWords:** Distribution Center, A-Frame, Automated Picking, DMAIC Methodology, Data Mining, Machine Learning, Continuous Improvement.

## TABLA DE CONTENIDO

RESUMEN .....	5
ABSTRACT .....	6
TABLA DE CONTENIDO .....	7
<b>1. Introducción</b> .....	<b>10</b>
<b>2. Revisión de Literatura</b> .....	<b>12</b>
<b>3. Metodología</b> .....	<b>14</b>
<b>3.1. Fase 1: Definir &amp; Entendimiento del negocio</b> .....	<b>17</b>
<b>3.2. Fase 2: Medir &amp; Entendimiento de los datos del negocio</b> .....	<b>17</b>
<b>3.3. Fase 3: Analizar &amp; Exploración de los datos</b> .....	<b>17</b>
<b>5.1. Diagnóstico</b> .....	<b>18</b>
<b>5.1.1. Gemba Walk</b> .....	<b>18</b>
<b>5.1.1.1. Zonas de Picking</b> .....	<b>19</b>
<b>5.1.1.2. El A-Frame</b> .....	<b>19</b>
<b>5.1.1.3. Datos de los A-Frames</b> .....	<b>20</b>
<b>5.2. Validación Cuantitativa del Problema</b> .....	<b>22</b>
<b>5.3. Propuesta de Mejora</b> .....	<b>23</b>
<b>5.3.1. ABC del producto</b> .....	<b>23</b>
<b>5.4. Implementación de Mejora</b> .....	<b>24</b>
<b>5.4.1. Aplicación para Seleccionar Cambios de Ubicación</b> .....	<b>24</b>
<b>5.4.2. Propuesta de Machine Learning</b> .....	<b>25</b>
<b>6. Resultados</b> .....	<b>31</b>
<b>6.1. Evaluar modelos</b> .....	<b>31</b>
<b>6.2. Aumento de Líneas</b> .....	<b>36</b>
7. Discusión .....	37
8. Conclusión .....	38
9. Recomendación .....	39
<b>Referencias</b> .....	<b>41</b>
<b>Anexos</b> .....	<b>46</b>
<b><i>Función 1.1:</i> generar_listas_equidistantes(umbral, uno, max)</b> .....	<b>50</b>
<b><i>Función 1.2:</i> ventana_filtro(df, columna)</b> .....	<b>50</b>

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> <i>Equipos de picking de WAMAS</i> .....	10
<b>Tabla 2.</b> Problemas típicos de análisis de datos empresariales y técnicas para resolverlos....	13
<b>Tabla 3.</b> Sistema y Zonas de picking y sistemas de manejo de inventario .....	19
<b>Tabla 4.</b> Explicación categorías logísticas usadas en los A-Frames .....	21
<b>Tabla 5.</b> Explicación Cuartil Mix.....	22
<b>Tabla 6.</b> Target 1: A-FRAME 1 .....	32
<b>Tabla 7.</b> Target 2: A-FRAME 2 .....	32
<b>Tabla 8.</b> Target 3: Carrusel .....	32
<b>Tabla 9.</b> Target 4: Manual 1 .....	33
<b>Tabla 10.</b> Target 5: Manual 2.....	33
<b>Tabla 11.</b> Target 6: PBT.....	33
<b>Tabla 12.</b> Target 7: Semi-manual.....	34
<b>Tabla 13.</b> Métricas evaluación modelo global .....	34
<b>Tabla 14.</b> AUC de las curvas ROC de acuerdo con cada target-área.....	35
<b>Tabla 15.</b> Resultados por áreas .....	36
<b>Tabla 16.</b> Resultados aumento de Líneas.....	36

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Ciclo QM-CRISP-DM.....	14
<b>Figura 2.</b> Diagrama de la metodología CRISP-DM.....	15
<b>Figura 3.</b> Ilustración del A-Frame.....	20
<b>Figura 4.</b> Matriz de confusión con Random Forest .....	27
<b>Figura 5.</b> Matriz de confusión con K-Nearest Neighbors .....	28
<b>Figura 6.</b> Matriz de confusión con Naive Bayes.....	29
<b>Figura 7.</b> Matriz de confusión con Gradient Boosting Machine.....	31

## 1. Introducción

Corporación GPF es una empresa ecuatoriana del sector del *retail* (CITEC, s.f.), especializada en el área farmacéutica, cuenta con la cadena líder en farmacias Fybeca (Fybeca, s.f.), que en su conjunto suman más de 900 locales, con sus marcas SanaSana y OkiDoki. Pertenece al grupo mexicano FEMSA que opera en diferentes industrias relacionadas al comercio (FEMSA, s.f.). Tiene un centro de distribución denominado: Centro de Logística Avanzada (CLA), ubicado en la parroquia rural de Amaguaña, en Quito. Este CLA fue un proyecto, en el cual la empresa invirtió, buscando automatizar la instalación, integrando el sistema de gestión de almacenes, proporcionado por Oracle, *Retail Warehouse Management System* (RWMS) y con maquinaria de la empresa SSI Schäfer (Ekos, 2017). Esta maquinaria automatizada la conforman equipos destinados al proceso de *picking* de diferentes productos.

En el CLA cada SKU (*Stock Keeping Unit*) es asignado a su zona de *picking* de acuerdo con sus características. Estas zonas están agrupadas en dos sistemas de *picking*, WMS y WAMAS (ambos acrónimos significan *Warehouse Management System*): el primero conformado por la zona PTB (*Pick to Belt*); y el segundo está conformado por equipos automatizados, semiautomatizados, y la zona de *picking* manual:

**Tabla 1.** *Equipos de picking de WAMAS*

WAMAS	A-Frame 1
	A-Frame 2
	Carrusel
	Semi-manual
	Manual 1
	Manual 2

*Nota.* Sophia Nieto, jefe de Productividad y Procesos Logísticos, comunicación personal, 28 de agosto de 2024.

Los A-Frames son los equipos automatizados que despachan pedidos más rápido. Según SSI Schaefer (s.f.), pueden procesar gran cantidad y variedad de productos, completando más pedidos diarios que el picking manual. Funcionan con canales que almacenan productos; al recibir un pedido, un motor los empuja a una banda transportadora. La única intervención manual es el llenado de canales. Los productos de alta demanda deben asignarse a este equipo, y la demanda se mide en líneas promedio diarias, donde más líneas indican mayor demanda

Al analizar los datos entregados, correspondientes a la fecha 20 de septiembre del 2024, se puede calcular la sumatoria de todas las Líneas de los SKU clasificado por zonas, y a partir de eso el porcentaje de las Líneas asignado a cada zona de *picking*. Los resultados muestran que, de todas las Líneas promedio diarias despachadas del CLA, el 34.2% de estas, corresponden a los dos equipos A-Frame. Sophia Nieto, jefe de Productividad y Procesos Logísticos (2024), nos expresó su preocupación, debido a que este dato solía estar, en años pasados, cerca del 40%. Lo que indica un desaprovechamiento reciente de las capacidades del A-Frame, además existen SKUs que deberían ser reubicados: fuera del A-Frame y dentro del A-Frame. Esto lleva a preguntas como: ¿por qué han disminuido las Líneas del A-Frame?, ¿cómo revertir esta situación? por estas dudas, esta tesis plantea una solución en base a los desafíos, que en este caso son:

- Determinar los criterios de selección de los SKUs para asignarlos en los canales de los A-Frames.
- Identificar todos los posibles SKUs candidatos a reubicación.
- Planificar los cambios de ubicación de los SKUs a efectuarse.

Todo esto encaminado a mejorar la asignación de SKUs a los canales de los A-Frames, para aumentar el porcentaje de Líneas asignado en esta zona.

Dado lo expuesto anteriormente, se plantea la aplicación de la metodología para resolver las interrogantes planteadas y una propuesta para cumplir el objetivo deseado. En términos de mejora continua, lo indicado es la aplicación de un DMAIC (*Define, Measure, Analyze, Improve, Control*), ya que, por ejemplo, según Mubarik et al. (2021), la aplicación de DMAIC en el centro de distribución, permitió la mejora en la eficiencia de las operaciones. Ahora, por las particularidades de este caso, y debido a la necesidad de analizar datos históricos de los SKUs proporcionados por la empresa, para la toma de decisiones, se decidió también incluir aspectos de la metodología CRISP-DM (*Cross-Industry Standard Process for Data Mining*), que suele ser aplicada en proyectos de minería de datos (Kregel et al., 2021). De esta manera se busca la integración de las metodologías DMAIC con CRISP-DM, es por esta razón, que se decidió trabajar con una integración entre las dos metodologías, denominada por Kregel como QM-CRISP-DM (*Quality Managment Cross-Industry Standard Process for Data Mining*), para la ejecución de los objetivos de este proyecto.

## **2. Revisión de Literatura**

La literatura existente se centra en la comprensión de datos históricos sobre A-Frames y otros sistemas automáticos de picking (APS). Caputo y Pelagagge (2008) analizaron la demanda y rendimiento de los AOPS (Sistemas Automáticos de Recogida de Pedidos), desarrollando un Sistema de Soporte a la Decisión (DSS) para configurar equipos de manera óptima. Por otro lado, Cuang et al. (2014) compararon políticas de asignación aleatoria y zonificación utilizando herramientas de minería de datos. Chiang et al. (2011) propusieron un enfoque de asignación de almacenamiento basado en minería de datos, conocido como DMSA, que optimiza el uso del espacio en estantes mediante un modelo de programación binaria entera para el Storage Location Assignment Problem (SLAP).

En el contexto de un Centro de Logística automatizado en la industria farmacéutica, se enfrentan a desafíos significativos debido a la alta demanda y la necesidad de precisión en los pedidos (Meller et al., 2008). La metodología DMAIC (Definir, Medir, Analizar, Implementar y Controlar) se presenta como un marco efectivo para mejorar la eficiencia y productividad, basándose en datos estadísticos en lugar de suposiciones (Park et al., 2020). Un estudio en Nigeria mostró que la eficiencia del almacén aumentó del 40% al 70% tras implementar DMAIC (Adeodu et al., 2023). Park et al. (2020) respaldan esta metodología, indicando que proporciona estructura y evita interpretaciones erróneas de grandes volúmenes de datos.

En un estudio de caso sobre Lean Six Sigma en Amore Pacific, Kregel et al. (2021) usaron minería de procesos para analizar datos de sistemas informáticos. Combinando DMAIC con CRISP-DM, lograron reducir la variación en tiempos de espera en un hospital, identificando causas raíz y midiendo el impacto de las soluciones.

En el libro "The Modern Business Data Analyst: A Case Study Introduction into Business Data Analytics with CRISP-DM and R", Jung (2024) ofrece una guía práctica sobre el análisis de datos empresariales, incluyendo una tabla para elegir algoritmos según el tipo de problema y el objetivo del proyecto.

**Tabla 2.** Problemas típicos de análisis de datos empresariales y técnicas para resolverlos

Tipo de problema	Perspectiva de análisis	Técnicas
Encontrar Clusters y valores atípicos	<ul style="list-style-type: none"> <li>¿Podemos agrupar a nuestros clientes en grupos distintos y diferenciados por características de marketing?</li> </ul>	Clustering K- Nearest Neighbour Analysis
Predecir explicar valores categóricos	<ul style="list-style-type: none"> <li>¿Qué productos se suelen comprar juntos?</li> </ul>	Árbol de decisión Regresión logística

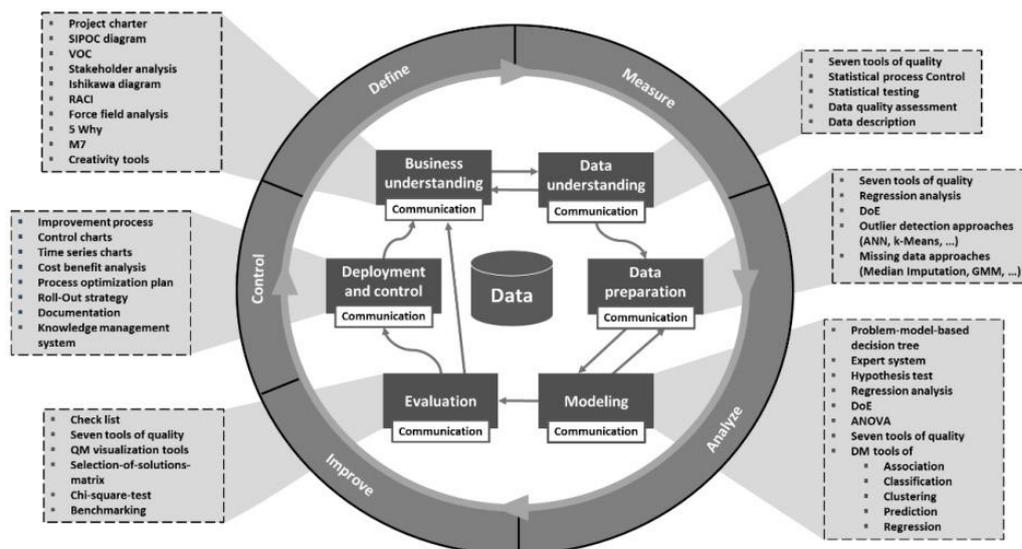
*Nota. Jung, D. (2024). Figura 5.2 [Problemas típicos de análisis de datos empresariales y técnicas para resolverlos.]*

Tufano et al. (2022), en su artículo "Un enfoque de aprendizaje automático para el diseño predictivo de almacenes", apoyan el diseño estratégico de sistemas de almacenamiento mediante clasificadores que predicen la tecnología de almacenamiento (ST) y la estrategia de asignación de almacenamiento (SAS). Utilizan algoritmos como Random Forest, destacando su resistencia al sobreajuste y capacidad para identificar patrones complejos, y Naive Bayes, que muestra robustez ante el ruido sin afectar significativamente el rendimiento.

### **3. Metodología**

Por las limitaciones que se presentan en un centro de distribución automatizado al generar un alto volumen de datos y aumentar la complejidad de los procesos se propone la utilización de las metodologías DMAIC y CRISP-DM. Siguiendo el ejemplo de la metodología usada en la investigación por Rashid Jalloh denominada QM-CRIP-DM, tal y como se muestra en la figura 1, sobre el Ciclo QM-CRISP-DM en el que se muestra como las distintas fases de estas dos metodologías se complementan a la par, y da ejemplos de herramientas que se pueden emplear (2023).

**Figura 1.** *Ciclo QM-CRISP-DM*



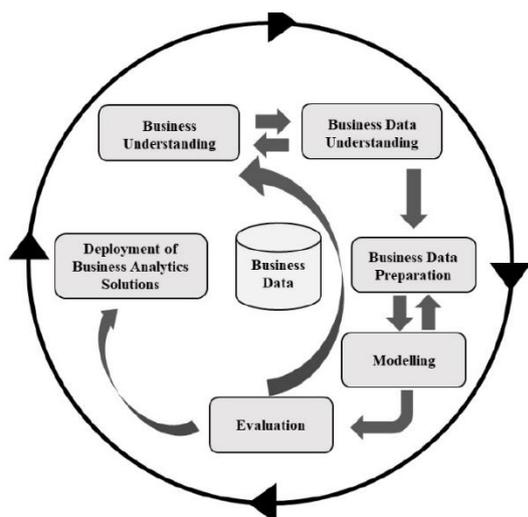
*Nota. De Jalloh, 2023.*

Montgomery (2013) define DMAIC como una metodología estructurada y ampliamente utilizada para mejorar la calidad y los procesos, enfocándose en la resolución de problemas. Sus pasos cruciales son:

- Definir: donde se delimita el problema y el alcance de la mejora.
- Analizar: mediante la información recolectada se obtienen análisis relevantes para identificar causas raíz.
- Implementación: desarrollo y aplicación de posibles soluciones.
- Control: implementar herramientas para asegurar la sostenibilidad de las soluciones propuestas.

Por otro lado, según Azadeh-Fard et al., (2019), plantea que la herramienta de proyectos CRISP-DM es una metodología de estructurada donde sus fases están orientadas en minería de datos, está compuesta por 6 fases.

**Figura 2.** Diagrama de la metodología CRISP-DM



*Nota. De Jung, 2024.*

- **Comprensión del negocio:** donde busca delimitar los objetivos del proyecto, identificar los criterios y herramientas requeridos.
- **Compresión de los Datos:** busca tener un entendimiento de los datos actual, catalogar los atributos que se tienen y determinar la calidad de los datos.
- **Preparación de los datos:** implica la clasificación, limpieza y exploración de los datos, así como la segregación en categorías y creación de sets de entrenamiento, prueba, y validación.
- **Modelado:** selección de herramienta predictiva, prueba del modelo y ajuste de parámetros.
- **Despliegue:** seguimiento y mantenimiento, documentación del modelo y pruebas con el usuario final.

A continuación, se definirán a más detalle las fases de forma complementaria usadas en la metodología QM-CRISP-DM y las herramientas seleccionadas para el análisis de este estudio de caso, en cada fase se utilizarán herramientas combinadas tanto de mejora de la calidad como de minería de datos.

### **3.1. Fase 1: Definir & Entendimiento del negocio**

Para la metodología DMAIC el objetivo principal de la fase definir es identificar la oportunidad del proyecto y validar el impacto de la mejora en los involucrados (Clientes internos y externos) (Montgomery, 2013. p 52). La metodología CRIPS-DM tiene como objetivo identificar el problema comercial con mayor impacto (costo beneficio) y establecer especificación, criterios y recursos que deben asignarse al proyecto (Jung, 2024).

A continuación, se definirán las herramientas seleccionadas para el desarrollo de esta metodología combinada, donde en su ejecución definirá el objetivo del proyecto identificando el problema de mayor impacto.

### **3.2. Fase 2: Medir & Entendimiento de los datos del negocio**

Según Montgomery (2013), la fase de medir tiene como objetivo evaluar y comparar el estado actual del proceso mediante la recopilación de datos e identificación de variables clave de entrada y salida. En la metodología CRISP-DM, Jung (2024) señala que la fase de entendimiento de los datos implica profundizar en el panorama de los datos existentes, su relevancia en el contexto empresarial y del proceso. Esta fase es crucial para la limpieza, formación y preparación de datos para su análisis, abarcando la recopilación, descripción, exploración y verificación de la calidad de los datos. En su ejecución esta fase tomará los datos recolectados y se lo estudiará con las diferentes herramientas tanto de minería de datos como de DMAIC, obteniendo un panorama más exacto de la necesidad y requerimiento del proyecto.

### **3.3. Fase 3: Analizar & Exploración de los datos**

En la metodología DMAIC, se desarrollan y prueban soluciones para las causas raíz de los problemas identificados, utilizando técnicas como diseño de experimentos y simulaciones para generar ideas y evaluar su efectividad. Las soluciones seleccionadas se

implementan en un entorno controlado para medir su impacto, con el fin de optimizar el proceso y reducir la variabilidad, asegurando mejoras sostenibles (Koppel & Chang, 2021).

4. En la fase de modelado, se eligen técnicas adecuadas según el problema, aplicando algoritmos de aprendizaje supervisado, como árboles de decisión o redes neuronales, y ajustando parámetros para optimizar el rendimiento. Se evalúa el modelo con conjuntos de datos de entrenamiento y prueba, midiendo métricas como precisión y AUC-ROC. En la fase de evaluación, se compara el rendimiento del modelo con métricas relevantes, analizando sesgos y limitaciones. Ambas fases son interdependientes y requieren un enfoque iterativo, siendo crucial la calidad del modelo para alcanzar los objetivos del negocio (Jung, 2024).

## 5. Ejecución

### 5.1. Diagnóstico

#### 5.1.1. Gemba Walk.

Usando lo indicado en el *Gemba Walk* se llevó a cabo las inducciones, capacitaciones y clases sobre cómo funcionan diferentes procesos y zonas del CLA, tomando en cuenta las siguientes fases:

- Notificar al equipo: Los coordinadores de cada área estuvieron al tanto del proyecto. Sabían que se buscaba reunir información para implementar un proyecto de mejora dentro de la zona del A-Frame.
- Observación de la ejecución del proceso: Se realizaron preguntas, y observaciones de la forma más respetuosa y siempre dejando en claro que era un espacio para aprender de los operadores.

A continuación, se expone la información recolectada con esta herramienta, enfocada en el cumplimiento del objetivo general:

### 5.1.1.1. Zonas de Picking.

En el Centro de Logística Avanzada (CLA), como ya se mencionó cuenta con una integración de software de Oracle, conocido como *Warehouse Management Systems*. Sin embargo, tienen dos sistemas WMS y WAMAS (ambos significan *Warehouse Management System*), su diferencia está en las zonas de *picking* que los conforman. A continuación, la

**Tabla 3.** Sistema y Zonas de picking y sistemas de manejo de inventario

Sistema	Zona de picking	Características de productos
WAMAS	A-Frames	Cuenta con equipos, denominados A-Frame, que almacena producto en sus canales y mediante un motor por canal y una banda transportadora, hace el <i>picking</i> de forma automatizada
	Semi- Manual	Cuenta con operarios que buscan en percha el producto solicitado y lo depositan en compartimentos que una vez completado el pedido, dejan caer los productos en una banda transportadora que los ubica en la gaveta.
	Manual	Operarios que manualmente llevan los productos solicitados desde la percha a la gaveta de pedido
	Carrusel	Equipo que almacena los productos en gavetas especializadas que al ser solicitadas por un pedido se dirigen automáticamente desde donde están guardadas, usando unos rieles, hasta el operador que hace el <i>picking</i> a la gaveta de pedido.
WMS	PBT ( <i>Pick to Belt</i> )	El pedido se va formando a lo largo de una banda transportadora, los operadores van a las perchas y depositan el producto en la banda.

*Nota. Sophia Nieto, jefe de Productividad y Procesos Logísticos, comunicación personal, 28 de agosto de 2024.*

### 5.1.1.2. El A-Frame

El A-Frame es un equipo de *picking* automatizado cuya eficiencia es de las más altas. Tiene unidades de almacenamiento denominadas canales verticales puestos consecutivamente a lo largo de un marco inclinado con forma de la letra A. Está compuesto con un actuador

electrónico denominado motor, que empuja unidades de producto por cada canal (Boywitz & Boysen, 2019). Debajo del marco existe una banda transportadora donde los productos empujados se dirigen y agrupan por orden a una gaveta que corresponde al pedido en sí, esto ocurre al final de la banda en el punto de llenado, tal como se ve en la Figura 1. Su ventaja intrínseca sobre otras áreas es mantener una alta productividad incluso en periodos de tiempo pico (SSI Schäfer, s.f.). Los A-Frames necesitan de dos procesos importantes para su funcionamiento: el primero es el proceso de abastecimiento, este se encarga de llevar la mercadería desde el inventario hasta la percha y segundo está el *refill* realizado por operadores que llenan los canales de los A-Frames con la mercadería de la percha, evitando que los equipos queden desabastecidos. Finalmente, una vez que se llenan las gavetas, existe una zona de revisión de las gavetas que presentaron productos incompletos en los pedidos realizados por los A-Frames, este proceso se denomina *batch* y los operadores son los encargados de realizar el *picking* manual de múltiples productos de diferentes pedidos simultáneamente para completarlos.

**Figura 3.** *Ilustración del A-Frame*



*Nota. De SSI Schäfer, s.f.*

### 5.1.1.3. Datos de los A-Frames

Sophia Nieto, Jefe de Productividad y Procesos Logísticos, explicó que los datos históricos relacionados a los A-Frames proporcionados por el *Warehouse Management*

*Systems* tienen ciertos indicadores los cuales se definen a continuación (comunicación personal, 28 de agosto de 2024).

Existen una clasificación de los productos llamada Categoría Logística, los clasifica primero por un código correspondiente a la Sección Contable (establecido por la empresa) y segundo por un código que representa si es un alimento, medicamento o suministro, lo cual es clave para cumplir con regulaciones sanitarias impuestas, que implican separar productos en los pedidos. Debido a que según ARCSA (Agencia Nacional de Regulación, Control y Vigilancia Sanitaria, 2023), la separación de los productos es la clave para la eliminación de contaminación cruzada, que sucede cuando agentes físicos o en este caso químicos ponen en riesgo la inocuidad de los productos. En el CLA existen muchas categorías logísticas; sin embargo, la Tabla 2 a continuación explica las tres categorías que se usan en el A-Frame.

**Tabla 4.** *Explicación categorías logísticas usadas en los A-Frames*

Categoría Logística	Sección Contable	Tipo de producto
B&C-ALIM	B&C: Bienestar y consumo	ALIM: Alimentos
B&CP-RESTO	B&CP: Maquillaje y belleza	RESTO: Consumo (Cosméticos, se excluyen alimentos y medicinas)
FARMA-MED	FARMA: Medicina	MED: Medicamentos

*Nota.* De Sophia Nieto, comunicación personal, 28 de agosto de 2024.

Otras maneras internas de la empresa de clasificar los productos, como QMIX (Cuartil Mix) esta clasificación se basa en dos grandes criterios: focos comerciales y análisis de cuartiles. Primero clasifica en tres focos comerciales clave: OAI (Capacidad de Ahorro imperdible), AC (Alternativa conveniente) e IMS (Productos más vendidos del mercado). Los productos que se clasificaron como alguna de esas tres categorías pertenecen a los productos más estratégicos e importantes para la empresa. Segundo, los SKU restantes se someten a un análisis de cuartiles: Q1, Q2, Q3 y Q4, esto permite identificar como se desempeñan estos productos en función de la demanda. Al sumar los productos de los focos comerciales y

cuartiles, se obtiene el total de los productos que maneja la empresa. A continuación, se presenta una tabla del orden de prioridad de los productos:

**Tabla 5.** *Explicación Cuartil Mix*

Etiquetas de QMIX	
Focos comerciales	OAI (Capacidad de Ahorro Imperdible)
	AC (Alternativa Conveniente)
	IMS (Productos más vendidos del mercado)
Análisis Cuartil	Q1 (Cuartil 1)
	Q2 (Cuartil 2)
	Q3 (Cuartil 3)
	Q4 (Cuartil 4)

*Nota. De Sophia Nieto, comunicación personal, 28 de agosto de 2024.*

## 5.2. Validación Cuantitativa del Problema

El objetivo es mejorar la asignación de SKUs a los A-Frames en el Centro de Distribución de Corporación GPF, ya que el porcentaje de Líneas asignadas a los A-Frames ha disminuido del 40% al 34.2%. Es necesario validar esta información con datos proporcionados por la empresa, que consisten en dos bases de datos.

- Base de Cambios de Ubicación: 15 mil registros
- Base de Despachos: 2,99 millones de registros
- Base de Dimensiones: 190 mil registros

### 5.2.1. Análisis Exploratorio de Datos.

Se uso un Análisis Exploratorio de Datos (EDA), con el objetivo de entender que información acompaña estas bases, que relaciones existen entre los datos mostrados, y preparar los datos para validar la existencia del problema. A continuación, se explica lo realizado con más detalle, esto se aplicó mediante uso de la librería Pandas en el lenguaje Python, que tuvo el objetivo de eliminar datos repetidos y estructurar los datos de una mejor manera con el fin de trabajar los mismos más adelante:

- Eliminación de 74 columnas de Base de Cambios de Ubicación, correspondientes a información no relevante para este estudio como: información redundante, información sobre proveedores, información de otras áreas.
- Normalización de la Base de Datos Despacho, se aplicó una normalización de la segunda forma (2NF), donde se eliminó la redundancia parcial, sabiendo que cumple con los requerimientos de la primera forma (1NF) (eliminación de valores repetidos. Esta segunda forma todos los atributos no clave deben depender completamente de la clave principal, la cual es en este caso la columna “Item”.
- Eliminación de registros correspondientes a SKUs descatalogados.

Se identificó ciertos atributos de la base de datos que nos servirán los cuales son, algunos fueron obtenidas en tres meses y otras en 15 días, esto se debe a que los datos fueron entregados de esta manera:

- Líneas Promedio Diarias: Obtenida a partir de los últimos 3 meses. Y por cada SKU.
- Unidades Promedio Diarias: Obtenida a partir de los últimos 3 meses. Y por cada SKU.  
Las unidades son la cantidad de productos de un mismo SKU pedido a lo largo del día.
- Líneas Canceladas Promedio diarias: Obtenida a partir de los últimos 15 días. Y por cada SKU.
- Rotación: Obtenida a partir de la base de datos de despachos, obtenida a partir de datos de un mes.

### 5.3. Propuesta de Mejora

#### 5.3.1. ABC del producto

Para cumplir el objetivo de identificar todos los posibles SKUs candidatos a estar en el A-Frame, se hizo primero una clasificación de productos ABC, tomando en cuenta los atributos “Líneas promedio 3 Meses” y “Unidades promedio 3 Meses” de cada Ítem.

Se estableció para “Líneas promedio 3 Meses” clasificarlos en las siguientes clases:

- Clase A: De 0 a percentil 80 de líneas
- Clase B: De percentil 80 de líneas a 100 líneas
- Clase C: De 100 al máximo disponible

Se estableció para “Unidades promedio 3 Meses” clasificarlos de manera similar. La clasificación se hizo en base a un análisis de Pareto, debido a la no uniformidad de los datos. Para ambos casos las clases A, B y C representan en conjunto el 20 % de los SKUs que tienen el 80% de Unidades o Líneas más frecuentes. Esta es una manera común en manejo de inventarios de segmentar productos.

#### 5.4. Implementación de Mejora

##### 5.4.1. Aplicación para Seleccionar Cambios de Ubicación

Dado que el objetivo busca aumentar las Líneas dentro del A-Frame, y que existe un proceso manual de seleccionar los cambios de ubicación de SKUs, se desarrolló una aplicación en Python, que permite emular esta operación de cambios de ubicación. Las funcionalidades de la aplicación son, en el Anexo 3 se explica a detalle la interfaz:

- Juntar las bases de datos, anteriormente mencionadas, y aplicar la respectiva limpieza de datos.
- Desplegar un menú que permita hacer un filtrado personalizado, tomando en cuenta factores como: Líneas Promedio, Unidades promedio, Rotación, Qmix, etc.
- Un botón que permita identificar en base a criterios recopilados en el Gemba Walk, los SKUs que deberían entrar al A-Frame y salir del A-frame.
- Reasignar de acuerdo con la prioridad de aumentar más líneas con los SKUs que entren y disminuir menos Líneas con los SKUs que salen. Crear una lista de máximo 100 cambios de ubicación semanales.

De esta manera, según Sophia Nieto, este proceso automatizado por la aplicación, se hace manualmente, y se puede decir que se divide en dos partes:

- Descargar las bases de datos del sistema, cuya duración de tiempo promedio estimado es media hora.
- Ejecución de crear los cambios de ubicación, cuya duración de tiempo promedio estimado es una hora y media hora.

Este proceso no solo tiene la ventaja de cumplir los objetivos del proyecto sino también de intervenir en la parte de ejecutar los cambios de ubicación. Ya que actualmente la aplicación tarda 2 minutos en promedio en recrear el proceso de hora y media.

#### **5.4.2. Propuesta de Machine Learning**

Se ha desarrollado un algoritmo basado en especificaciones operativas y la experiencia de los operadores. Sin embargo, la literatura sugiere que los algoritmos de aprendizaje supervisado pueden identificar productos y predecir su asignación de manera más estructurada, reduciendo la dependencia del juicio subjetivo. Esto proporciona un proceso de selección más flexible y adaptable. Además, se destaca la efectividad de estos algoritmos en la asignación de recursos para mejorar la eficiencia operativa. Por lo tanto, se optará por utilizar cuatro algoritmos para abordar esta problemática: *Random Forest*, *K-Nearest Neighbors (KNN)*, *Gradient Boosting Machines (GBM)*, *Naive Bayes*.

Para identificar el mejor algoritmo, se analizarán los modelos con métricas específicas, enfocándose en minimizar los falsos positivos (ver anexo 5). Un falso positivo ocurre cuando un producto se asigna incorrectamente a un área (SEMI, A-Frames, PBT, etc.), lo que va en contra del objetivo de optimizar la asignación y mejorar la eficiencia operativa. Esto provoca una ocupación inadecuada de espacios y subutilización de recursos.

### *Definición de métricas*

- *Recall o Sensibilidad:* establece la proporción entre el número de predicciones positivas entre todos los verdaderos positivos.
- *Precisión:* muestra la proporción entre el número total de predicciones verdaderas positivas sobre el número total de predicción positivas del modelo.
- *Accuracy:* Indica el número de predicciones acertadas en proporción a todas las predicciones realizada por el modelo.

(Jung, 2024)

Es destacable mencionar que todas las variables respuesta de los siguientes modelos de aprendizaje supervisado son variables categóricas por lo que se tiene una matriz de 0-1, lo que convierte a los modelos en modelos de multi-clasificación.

### **Random Forest**

El algoritmo Random Forest se basa en la creación de múltiples árboles de decisión a partir de muestras aleatorias del conjunto de datos original. Primero, se generan subconjuntos de datos utilizando el método de "bootstrap". Luego, para cada subconjunto, se construye un árbol de decisión, eligiendo aleatoriamente un conjunto de características en cada nodo para evitar la correlación entre árboles. Las predicciones se realizan mediante la votación de los árboles en clasificación o tomando el promedio en regresión (Tufano, et al. 2022).

### *Implicaciones dentro del estudio*

En la predicción del área de almacenamiento en un centro de distribución de FEMSA, el algoritmo Random Forest analiza características como rotación, costo, altura, peso, volumen y categoría de los productos. Entrena múltiples árboles de decisión con subconjuntos aleatorios de datos y evalúa combinaciones de características para clasificar los

productos. El modelo predice el área de almacenamiento óptima mediante la votación de los árboles, optimizando así la asignación de productos. Como algoritmo multi-clasificador, la matriz de confusión muestra el porcentaje de casos por clase, reflejando la clasificación global por área.

Tras ejecutar el modelo de aprendizaje automático Random Forest, se obtuvieron los resultados que dieron lugar a la matriz de confusión (ver Anexo 5 para más detalles sobre su interpretación).

**Figura 4.** *Matriz de confusión con Random Forest*

		Predicciones del modelo	
		Negativo	Positivo
Valor Real	Negativo	Verdadero negativo 85,41%	Error tipo 1 Falso positivo 0,92%
	Positivo	Error tipo 2 Falso Negativo 1,72%	Verdadero Positivo 11,95%

*Nota. Elaborado por Acosta et al, (2024).*

La matriz proporciona una visión del rendimiento del algoritmo al clasificar los productos en las áreas de almacenamiento, tenemos un porcentaje de valores verdaderos de 97.36%, teniendo un porcentaje de error 2.64%.

### ***K-Nearest Neighbors (KNN)***

El método K-Nearest Neighbors (K-NN) inicia dividiendo el conjunto de datos de entrenamiento en  $N$  *clusters* mediante un método de agrupamiento. Se define una curva promedio para cada *cluster* que refleja su comportamiento. Al recibir una nueva observación, K-NN clasifica analizando los  $k$  vecinos más cercanos, utilizando una función de distancia, generalmente euclidiana. Predice la clase o valor de la observación según la mayoría de las etiquetas (clasificación) o el promedio de valores (regresión), lo que lo hace flexible y efectivo para diversas aplicaciones (Violetto & Noro, 2019).

### Implicaciones dentro del estudio

Para el presente estudio, el algoritmo *k-Nearest Neighbors (kNN)* analizaría las características de los productos, como rotación, costo, peso, volumen y categoría logística, etc. Al recibir una nueva observación, *kNN* calcularía la distancia entre esta observación y todos los productos en el conjunto de datos de entrenamiento para identificar los *k* vecinos más cercanos. Luego, clasificaría la nueva observación en función de la mayoría de las etiquetas de estos vecinos, determinando así el área de almacenamiento más adecuada para el producto en función de sus características, lo que optimizaría la organización del almacén y mejoraría la eficiencia operativa.

Una vez corrido el modelo de aprendizaje automatizado de *K-Nearest Neighbors*, se obtuvieron los resultados con los cuales se realizaron la siguiente matriz de confusión, (Anexo 5, para más información de la lectura de una matriz de confusión.)

**Figura 5.** Matriz de confusión con *K-Nearest Neighbors*

		Predicciones del modelo	
		Negativo	Positivo
Valor Real	Negativo	Verdadero negativo 83.80%	Error tipo 1 Falso positivo 2.53%
	Positivo	Error tipo 2 Falso Negativo 2.91%	Verdadero Positivo 10.75%

*Nota. Elaborado por Acosta et al, (2024).*

La matriz proporciona una visión del rendimiento del algoritmo al clasificar los productos en las áreas de almacenamiento, tenemos un porcentaje de valores verdaderos de 94.53%, teniendo un porcentaje de error 5.44%.

### Naive Bayes

Naive Bayes es un algoritmo de clasificación que utiliza el teorema de Bayes para calcular la probabilidad de un evento. Asume la independencia de las características,

simplificando los cálculos. Para clasificar un nuevo ejemplo, calcula la probabilidad de pertenencia a cada clase usando las características y la fórmula de Bayes. Las probabilidades a priori se estiman de la frecuencia en el conjunto de entrenamiento, y las de las características se multiplican bajo la suposición de independencia. Se elige la clase que maximiza la probabilidad calculada (Tufano et al., 2022).

### *Implicaciones en el estudio*

Al recibir una nueva observación, el algoritmo calcularía la probabilidad de que el producto pertenezca a cada área de almacenamiento, basándose en las características del producto y utilizando la suposición de independencia entre estas características. Finalmente, asignaría el producto al área con la mayor probabilidad, optimizando así la organización del almacén y mejorando la eficiencia en la gestión del inventario.

Tras ejecutar el modelo de Naive Bayes, se obtuvieron resultados que dieron lugar a la matriz de confusión (ver Anexo 5 para más detalles).

**Figura 6.** *Matriz de confusión con Naive Bayes*

		Predicciones del modelo	
		Negativo	Positivo
Valor Real	Negativo	Verdadero negativo 76.36%	Error tipo 1 Falso positivo 9.35%
	Positivo	Error tipo 2 Falso Negativo 1.91%	Verdadero Positivo 12.36%

*Nota. Elaborado por Acosta et al, (2024).*

La matriz proporciona una visión del rendimiento del algoritmo al clasificar los productos en las áreas de almacenamiento, tenemos un porcentaje de valores verdaderos de 88.72%, teniendo un porcentaje de error 11.26%.

## **Gradient Boosting Machine**

Gradient Boosting es un método de aprendizaje automático que mejora la precisión mediante la construcción secuencial de modelos. Comienza con un modelo simple, como un árbol de decisión, que predice los datos de entrenamiento. Luego, se calculan los errores y se construye un nuevo modelo para predecir estos errores. Las predicciones se combinan con las del modelo anterior, ajustadas por un factor de aprendizaje. Este proceso se repite hasta alcanzar la precisión deseada o completar un número de iteraciones, buscando minimizar una función de pérdida que mide la discrepancia entre predicciones y valores reales (Kalusivalingam et al., s.f.).

#### *Implicaciones en el modelo*

El algoritmo Gradient Boosting Machine (GBM) utilizaría las características de los productos, como rotación, costo, altura, peso, volumen y categoría logística, para realizar predicciones precisas. Comenzaría con un modelo inicial simple que predice el área de almacenamiento y, a partir de las discrepancias entre las predicciones y los valores reales, calcularía los errores. Luego, construiría nuevos árboles secuencialmente, cada uno enfocado en corregir los errores del modelo anterior. Las predicciones se actualizarían iterativamente, ajustándose con un factor en este caso las características de cada producto, hasta que el modelo minimizara la función de pérdida y alcanzara un nivel óptimo de precisión, optimizando así la asignación de productos en el almacén.

Una vez corrido el modelo de aprendizaje automatizado de *Gradient Boosting Machine*, se obtuvieron los resultados con los cuales se realizaron la siguiente matriz de confusión, (Anexo 5, para más información de la lectura de una matriz de confusión.)

**Figura 7. Matriz de confusión con Gradient Boosting Machine**

		Predicciones del modelo	
		Negativo	Positivo
Valor Real	Negativo	Verdadero negativo 84.58%	Error tipo 1 Falso positivo 1.13%
	Positivo	Error tipo 2 Falso Negativo 1.26%	Verdadero Positivo 13.01%

*Nota. Elaborado por Acosta et al, (2024).*

La matriz muestra el rendimiento del algoritmo en la clasificación de productos en almacenamiento, con un 97.59% de valores verdaderos y un 2.39% de error.

## 6. Resultados

### 6.1. Evaluar modelos

La evaluación del modelo es una fase crítica dentro de la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) principalmente porque permite validar los resultados, asegurando que el modelo cumpla con los objetivos establecidos en la fase de comprensión del negocio, lo que garantiza su relevancia y utilidad para las necesidades empresariales (Jung, 2024, p.28). Durante la evaluación se utilizan métricas específicas, como precisión, sensibilidad y especificidad, para medir el rendimiento del modelo y determinar su idoneidad para la implementación (Jung, 2024, p.31).

A continuación, se presentará 7 tablas cada una correspondiente a un área, cada tabla presentará las métricas específicas anteriormente definidas. Ya que se definió anteriormente, el error que se pretende minimizar con mayor impacto es el falso positivo, la métrica que mide el que tan correctos son los datos predichos, es la sensibilidad; por este motivo se concluirá a partir de esta medida

**Tabla 6.** *Target 1: A-FRAME 1*

	Random Forest	KNN	GBM	Naive Bayes
Acurracy	0.9928	0.9353	0.9952	0.9041
Sensibilidad	0.9928	0.9353	0.9952	0.9041
Precisión	0.9928	0.9349	0.9952	0.9210

*Nota. Elaborado por Acosta et al, (2024).*

Las métricas de evaluación indican que el modelo GBM es el mejor, con una puntuación de 0.9952. Esto significa que, para el área A-Frame 1, GBM predice correctamente qué productos deben estar en esta área.

**Tabla 7.** *Target 2: A-FRAME 2*

	Random Forest	KNN	GBM	Naive Bayes
Acurracy	0.9904	0.9472	1	0.8609
Sensibilidad	0.9904	0.9472	1	0.8609
Precisión	0.9905	0.9531	1	0.9360

*Nota. Elaborado por Acosta et al, (2024).*

Estudiando los algoritmos desde las métricas de evaluación se puede concluir que el modelo de GBM es el mejor puntuado con un 1. Es decir que para el área de A- Frame 2, el modelo que mejor predice que productos deben estar en esta área de forma correcta es GBM.

**Tabla 8.** *Target 3: Carrusel*

	Random Forest	KNN	GBM	Naive Bayes
Acurracy	0.9329	0.9017	0.9257	0.8297
Sensibilidad	0.9329	0.9017	0.9257	0.8297
Precisión	0.9336	0.9024	0.9293	0.8739

*Nota. Elaborado por Acosta et al, (2024).*

El análisis de las métricas de evaluación concluye que Random Forest es el mejor algoritmo, con una puntuación de 1. Esto indica que, para el área de Carrusel, Random Forest predice correctamente los productos que deben estar en esta área

**Tabla 9.** *Target 4: Manual 1*

	Random Forest	KNN	GBM	Naive Bayes
Acurracy	0.9376	0.9089	0.9424	0.7842
Sensibilidad	0.9376	0.9089	0.9424	0.7842
Precisión	0.9364	0.9064	0.9414	0.8378

*Nota. Elaborado por Acosta et al, (2024).*

Estudiando los algoritmos desde las métricas de evaluación se puede concluir que el modelo de GBM es el mejor puntuado con un 0.9424. Es decir que para el área de Manual 1, el modelo que mejor predice que productos deben estar en esta área de forma correcta es GBM.

**Tabla 10.** *Target 5: Manual 2*

	Random Forest	KNN	GBM	Naive Bayes
Acurracy	1	1	1	0.9904
Sensibilidad	1	1	1	0.9904
Precisión	1	1	1	10.000

*Nota. Elaborado por Acosta et al, (2024).*

Estudiando los algoritmos desde las métricas de evaluación se puede concluir que los mejores modelos son Random Forest, K-NN y GBM teniendo una puntuación igualitaria de 1. Es decir, que para el área de Manual 2, los mejores modelos que predicen que predicen de forma más correcta son los modelos anteriormente mencionados.

**Tabla 11.** *Target 6: PBT*

	Random Forest	KNN	GBM	Naive Bayes
Acurracy	0.9616	0.9448	0.9712	0.8585
Sensibilidad	0.9616	0.9448	0.9712	0.8585
Precisión	0.9590	0.9406	0.9698	0.9032

*Nota. Elaborado por Acosta et al, (2024).*

Estudiando los algoritmos desde las métricas de evaluación se puede concluir que el modelo de GBM es el mejor puntuado con un 0.9712. Es decir que para el área de PBT, el modelo que mejor predice que productos deben estar en esta área de forma correcta es GBM.

**Tabla 12.** *Target 7: Semi-manual*

SEMI- MANUAL				
	Random Forest	KNN	GBM	Naive Bayes
Acurracy	0.9760	0.9736	0.9976	0.9832
Sensibilidad	0.9760	0.9736	0.9976	0.9832
Precisión	0.9766	0.9687	0.9976	0.9817

*Nota. Elaborado por Acosta et al, (2024).*

Estudiando los algoritmos desde las métricas de evaluación se puede concluir que el modelo de GBM es el mejor puntuado con un 0.9976. Es decir que para el área de Semi, el modelo que mejor predice que productos deben estar en esta área de forma correcta es GBM.

**Tabla 13.** *Métricas evaluación modelo global*

METRICAS GLOBALES				
Métrica	Random Forest	KNN	GBM	Naive Bayes
Acurracy	0.9702	0.9445	0.9760	0.8873
Sensibilidad	0.9702	0.9445	0.9760	0.8873
Precisión	0.9698	0.9437	0.9762	0.9219

*Nota. Elaborado por Acosta et al, (2024).*

Las métricas presentadas evalúan globalmente los modelos según los resultados del set de prueba y su comparación con el set de validación. La tabla 13 muestra que el modelo Gradient Boosting Machine tiene los mejores valores: 0.9760 en accuracy, sensibilidad y precisión, superando a los demás modelos.

En todas las áreas, excepto en el Carrusel, el modelo con la mayor sensibilidad (Recall) es el Gradient Boosting Machine (GBM). Este algoritmo es autoajutable, lo que es crucial en este estudio debido a la naturaleza dinámica de los datos. Al aprender de sus errores y optimizar recursos en función de la función de pérdida, el GBM se convierte en un algoritmo flexible. Sin embargo, esto no garantiza que siempre sea el de mejor rendimiento. A medida que los datos de entrenamiento mejoren, se espera que otros algoritmos también aprendan y su rendimiento varíe y mejore.

### *Evaluación de los modelos a través de Curvas ROC*

Las curvas ROC (Receiver Operating Characteristic) son gráficas que evalúan el rendimiento de modelos de clasificación, especialmente binaria. Representan la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) a diferentes umbrales, mostrando el trade-off entre sensibilidad y especificidad. Un modelo ideal tendría TPR de 1 y FPR de 0, lo que se refleja en una curva cercana al punto (0,1) (Aznar, 2024).

Una de las métricas clave de la curva ROC es el Área Bajo la Curva (AUC). Un AUC de 1 indica un modelo perfecto, mientras que 0.5 sugiere una clasificación aleatoria. Valores entre 0.5 y 1 indican capacidad de discriminación, siendo mejores los que se acercan a 1 (Aznar, 2024, p. 162).

A continuación, se presenta una tabla con el AUC de las curvas ROC para cada target. Para visualizar los gráficos de las curvas ROC, consulte el Anexo 7.

**Tabla 14.** *AUC de las curvas ROC de acuerdo con cada target-área*

Modelo	Random Forest	KNN	GMB	Naive Bayes
A-FRAM1	1	0,96	1	0,98
A-FRME 2	1	0,91	1	0,96
CARRUSEL	0,98	0,94	0,99	0,92
MANUAL 1	0,98	0,9	0,98	0,86
MANUAL 2	1	1	1	0,99
PBT	0,98	0,84	0,97	0,69
SEMI-MANUAL	0,99	0,74	1	0,93
Promedio	0,990	0,899	0,991	0,890

*Nota. Elaborado por Acosta et al, (2024).*

Al igual que con las métricas de evaluación, el modelo con mejor rendimiento promedio es el Gradient Boosting Machine, aunque los resultados varían por áreas.

**Tabla 15.** Resultados por áreas

Área	Mejor modelo -Análisis Curvas ROC
A-FRAME 1	Random Forest
A-FRME 2	Random Forest
CARRUSEL	GBM
MANUAL 1	GMB/ Random Forest
MANUAL 2	Random Forest/KNN/GBM
PBT	Random Forest
SEMI-MANUAL	Random Forest

*Nota. Elaborado por Acosta et al, (2024).*

## 6.2. Aumento de Líneas

A partir de la implementación de la automatización del proceso actual, y la aplicación de los modelos de *Machine Learning*. Se obtuvo resultados, enfocados en cuantas Líneas aumentaría el A-Frame semanalmente con los datos obtenidos.

La Líneas aumentadas son únicamente calculadas usando solo 100 posibles cambios de ubicación. Como del modelo se obtiene dos archivos .csv: SKUs que entran y SKUs que salen. Y se tiene la cantidad de Líneas promedio por cada SKU en estos archivos. Se puede sumar las Líneas que salen y las que entran. A partir del dato inicial de Líneas dentro del A-Frame, este es restado por las Líneas que salen y sumado las Líneas que entran. Y a la vez se saca los puntos porcentuales que aumento respecto al total de Líneas del CLA. Solo con esos 100 cambios.

*Resultados aumento de Líneas en diferentes métodos*

**Tabla 16.** Resultados aumento de Líneas

Método	Líneas aumentadas en A-Frame	Puntos porcentuales aumentados: Líneas del A-Frame respecto a las del CLA
Modelo Actual	485	0,39%
Random Forest	234	0,18%
KNN	22	0,01%
Gradient Boost	256	0,21%
Naive Bayes	250	0,20%

*Nota. Elaborado por Acosta et al, (2024).*

## 7. Discusión

Como se concluyó en el análisis de aumento de líneas, el modelo que más incrementa el número de líneas salientes en el A-Frame es el algoritmo automatizado. Es importante mencionar esto, ya que, aunque en la literatura se encuentra que existen aumentos significativos en la eficiencia operativa con el uso de algoritmos de aprendizaje supervisado, al compararlo con un algoritmo basado en los criterios de los operadores, este último resulta ser el que más aumenta la eficiencia. Esto nos lleva a conclusiones como que, si bien metodologías de minería de datos como los algoritmos de aprendizaje supervisado tienen un gran poder y potencial computacional, la experiencia de los operarios es un factor significativo e importante para considerar, y que, a largo plazo, puede aportar de manera complementaria a los modelos. Sin embargo, esto comprendería una inducción y capacitación en el uso de la herramienta en los operarios.

Sabemos que el principal objetivo que inició este estudio fue el aumento de líneas, lo que se demostró en la Tabla 17 en el A-Frame. Sin embargo, el desarrollo de un modelo de asignación de productos según sus características, utilizando modelos de aprendizaje supervisado, abre el camino a aplicaciones del proyecto que van mucho más allá de la simple asignación de productos a A-frames. Comprender el funcionamiento de los modelos permite identificar variables significativas para la toma de decisiones. A largo plazo, esto puede usarse para simular cambios en criterios de asignación en áreas de almacenamiento, nuevos productos o eliminación de capacidad. Este desarrollo tiene un gran potencial para aplicaciones en proyectos dentro de la empresa.

Como se observó en el análisis de la evaluación de los distintos modelos empleados en este proyecto, las conclusiones fueron que, aunque el algoritmo de GBM presentó los mejores resultados tanto en las métricas como en la evaluación de AUC-ROC, no se puede afirmar

que este sea el mejor para todas las iteraciones futuras. Esto se debe a la naturaleza dinámica de los datos y a la hiperparametrización de los modelos. Los datos de entrenamiento son solo una instantánea en el tiempo, la cual cambia con el paso del tiempo. Al elegir únicamente un modelo, perderíamos las ventajas de la evaluación y comparación de otros algoritmos. De la misma forma, cada modelo tiene sus propios parámetros establecidos que, en caso de sobreajuste, pueden ser ajustados para mejorar el rendimiento. Este proceso se itera tantas veces como sea necesario hasta aumentar la métrica de sensibilidad, ya que nuestro interés es reducir los falsos positivos. Esto implica que cada predicción tiene métricas y rendimientos diferentes, y que, con el tiempo, los modelos pueden aprender de sus propias iteraciones.

## 8. Conclusión

Este proyecto tuvo como principal objetivo el aumento de las Líneas de la zona de A-Frames. Se plantea la implementación de una app (aplicación) que haga los filtrados de selección de cambios de ubicación que conducen a aumentar ese objetivo. Usando dos perspectivas de sistemas: La automatización del proceso actual, facilitando filtrados y limpieza de datos, y la perspectiva de usar el *Machine Learning* para que un modelo entrenado indique los destinos óptimos de los SKUs, con el mismo fin de aumentar las Líneas que están asignadas a los dos Equipos A-Frame. Como se muestra en la Tabla 16, en este caso el mejor método es el proceso actual, ya que con este aumentaron las Líneas 0,39 puntos porcentuales, con la limitante de realizar 100 cambios correspondientes a los cambios posibles en cada semana. Es menester recalcar que el objetivo de esta aplicación es que sea algo dinámico y estos outputs del listado de 100 cambios se hagan cada semana actualizando los datos. Para este proyecto se evaluó únicamente con los datos descargados de un único instante de tiempo. En una implementación en operaciones, esta aplicación puede que de un mayor aumento con modelos de *Machine Learning*. Por esto mismo es difícil estimar en cuanto tiempo se tardaría en aumentar a un 40% el porcentaje de Líneas. Si se hace una

proyección estimando que se tiene el mismo aumento semanalmente, de 0,39%, se alcanzaría el objetivo en 15 semanas. Sin embargo, este dato requiere de la ejecución de una fase de Implementar y Controlar en el DMAIC dentro de la operación, para así incluso hacer pronósticos de este tiempo.

Para alcanzar este objetivo se cumplió con la estandarización de criterios de selección de SKUs aptos para estar o no en el A-Frame, usando los filtros internos dentro de la aplicación desarrollada en Python. Para esto se necesitó la creación de un ABC del producto en base a las Líneas y Unidades despachadas, y el cálculo de la rotación de cada SKUs, en base a los despachos. También, la misma aplicación identifica todos los posibles SKUs aptos para estar en el A-Frame y planifica como hacer los cambios, tomando en cuenta que los canales desocupados por los SKUs que salen del A-Frame deben ser los mismos canales que serán usados para los SKUs que entran al A-Frame.

Por último, el modelo se validó, ya que se envió la lista de los SKUs que debían entrar al A-Frame, resultante de crear la aplicación. Esta lista tenía 6 SKUs, de los cuales 4 fueron finalmente colocados. No permite sacar datos concluyentes, sobre la efectividad, pero puede llegar a ser un paso inicial para el proyecto más allá de este trabajo de titulación.

## 9. Recomendación

Para la implementación de este proyecto en operaciones va mucho más allá de lo presentado en este informe por lo que se recomienda las siguientes implementaciones y tomar en cuenta algunas ideas:

- Es necesario la implementación de llamadas SQL con el WMS hacia la aplicación para que la parte del proceso de bajada de datos sea automática. Y que todo el proceso mase de tener una duración de dos horas y media a ser cuestión de minutos.

- Es necesario mejorar el set de entrenamiento de los modelos. Debido a que con la iteración de datos probada el modelo automatizado aumentaba las líneas en mayor cantidad que cualquier otro modelo de *Machine Learning*.
- Se puede con más datos y outputs del modelo que se tengan semanalmente, llevar un control de que tanto se ha alcanzado el objetivo de este proyecto. Y se puede pronosticar a mejor detalle cuando se alcance este.

## Referencias

- Adeodu, A., Maladzhi, R., Kana-Kana Katumba, M. G., & Daniyan, I. (2023). Development of an improvement framework for warehouse processes using lean six sigma (DMAIC) approach. A case of third party logistics (3PL) services. *Heliyon*, 9(4). <https://doi.org/10.1016/j.heliyon.2023.e14915>
- Alemán S., (s.f). Gemba – Kaisen una herramienta eficaz para lograr mejoras continuas en los procesos de producción. <https://orcid.org/0009-0006-5665-1656>
- Augusto Guimarey López, F., Luz Hernández Monsalve, L., & Humberto Vasquez Coronado, M. (2021). *Mejora de la productividad empleando la metodología DMAIC* (Vol. 8, Issue 2). <https://orcid.org/0000-0002-2616-6653>
- Azadeh-Fard, N., Megahed, F. M., & Pakdil, F. (2019). Variations of length of stay: a case study using control charts in the CRISP-DM framework. In *International Journal of Production Research, Journal of Manufacturing Technology Management* (Vol. 11, Issue 3).
- Bartholdi, J., & Hackman, S. T. (2005). *Warehouse & distribution science*. Disponible en: [www.tli.gatech.edu/whscience/book/wh-sci.pdf](http://www.tli.gatech.edu/whscience/book/wh-sci.pdf)
- Caputo, A. C., & Pelagagge, P. M. (2006). Management criteria of automated order picking systems in high-rotation high-volume distribution centers. *Industrial Management & Data Systems*, 106(9), 1359–1383. <https://doi.org/10.1108/02635570610712627>
- CAT. (2021). Very narrow aisle (VNA) applications: Everything you need to know. (n.d.). *Cat® Lift Trucks EAME*. <https://www.catlifttruck.com/blog/very-narrow-aisle-vna-applications-everything-you-need-know>

- Chen, M.-C., Huang, C.-L., Wu, H.-P., Hsu, M.-F., & Hsu, F.-H. (2005). A Data Mining Technique to Grouping Customer Orders in Warehouse Management System. En A. Abraham, Y. Dote, T. Furuhashi, M. Köppen, A. Ohuchi, & Y. Ohsawa (Eds.), *Soft Computing as Transdisciplinary Science and Technology* (pp. 1063–1070). Springer Berlin Heidelberg.
- Chiang, D. M.-H., Lin, C.-P., & Chen, M.-C. (2011). The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres. *Enterprise Information Systems*, 5(2), 219–234.  
<https://doi.org/10.1080/17517575.2010.537784>
- Chuang, Y.-F., Chia, S.-H., & Wong, J.-Y. (n.d.). Enhancing order-picking efficiency through data mining and assignment approaches.
- CITEC. (s.f.). Corporación GPF. *CITEC*. <https://citec.com.ec/Directorio/corporacion-gpf-grupo-fybeca/>
- David Boywitz, S. S., & Boysen, N. (2019). Sequencing of picking orders to facilitate the replenishment of A-Frame systems. *IIE Transactions*, 51(4), 368–381.  
<https://doi.org/10.1080/24725854.2018.1513672>
- Ekos. (22 de marzo de 2017). Centro de Logística Avanzada (CLA) - Provefarma Corporación GPF. *Revista Ekos*. <https://ekosnegocios.com/articulo/visitamos-a-centro-de-logistica-avanzada-cla-provefarma-corporacion-gpf>
- Femsa. (s.f.). *Quiénes somos*. Femsa. <https://www.femsa.com/es/acerca-de-femsa/quienes-somos/>
- Fybeca. (s.f.). *Quiénes somos*. Fybeca. <https://www.fybeca.com/acerca-de-nosotros/quienes-somos.html>

- Hunter, J. D., et al. (2023). *Matplotlib: Visualization with Python*. matplotlib.org.  
<https://matplotlib.org/>
- Inviarobotics. (2022, July 14). Order Fulfillment: a Guide to Throughput Metrics. *inVia Robotics*. <https://inviarobotics.com/blog/order-fulfillment-guide-throughput-metrics/>
- Jalloh, R. (2023). *Synthesizing a predictive analytics methodology for quality management*.  
<https://aaltodoc.aalto.fi/items/0b3f004a-ff1c-4495-a6a5-f27409d81027>
- Jung, D. (2024). *The Modern Business Data Analyst A Case Study Introduction into Business Data Analytics with CRISP-DM and R*.
- Kim, W. R., Kim, N. U., & Chang, Y. S. (2016). A study on the control of an automatic picking system (APS) in a warehouse. *Assembly Automation*, 36, 388–397.  
<https://api.semanticscholar.org/CorpusID:62954702>
- Koppel, S., & Chang, S. (2021). MDAIC – a Six Sigma implementation strategy in big data environments. *International Journal of Lean Six Sigma*, 12(2), 432–449.  
<https://doi.org/10.1108/IJLSS-12-2019-0123>
- Kregel, I., Stemann, D., Koch, J., & Coners, A. (2021). Process Mining for Six Sigma: Utilising Digital Traces. *Computers and Industrial Engineering*, 153.  
<https://doi.org/10.1016/j.cie.2020.107083>
- Li, Y., Méndez-Mediavilla, F. A., Temponi, C., Kim, J., & Jimenez, J. A. (2021). A heuristic storage location assignment based on frequent itemset classes to improve order picking operations. <https://doi.org/10.3390/app11041839>
- Liu, C. M. (1999). Clustering techniques for stock location and order-picking in a distribution center. *Computers & Operations Research*, 26(10–11), 989–1002.  
[https://doi.org/10.1016/S0305-0548\(99\)00026-X](https://doi.org/10.1016/S0305-0548(99)00026-X)

- Liu, P., Wu, Y., & Xu, N. (2009). Assigning SKUs to multiple automated-picking areas over multiple periods. *2009 IEEE International Conference on Automation and Logistics*, 50–55. <https://doi.org/10.1109/ICAL.2009.5262982>
- Meller, R., & Pazour, J. (2008). A heuristic for SKU assignment and allocation in an A-Frame system. *IIE Annual Conference and Expo 2008*.
- Montgomery, D. C., & Wiley, J. (n.d.). *Sixth Edition Introduction to Statistical Quality Control*.
- Mubarik, M., Rawi, D., & Khan, S. (2022). A DMAIC approach for warehouse storage and order picking process improvement. *International Journal of Productivity and Quality Management*, 1(1). <https://doi.org/10.1504/IJPQM.2021.10042369>
- Pandas Development Team. (2023). *pandas: Python Data Analysis Library*. [pandas.pydata.org](https://pandas.pydata.org). <https://pandas.pydata.org/>
- Pang, K.-W., & Chan, H.-L. (2017). Data mining-based algorithm for storage location assignment in a randomised warehouse. *International Journal of Production Research*, 55(14), 4035–4052. <https://doi.org/10.1080/00207543.2016.1244615>
- Park, S. H., Dhalgaard-Park, S. M., & Kim, D. C. (2020). New paradigm of lean six sigma in the 4th industrial revolution era. *Quality Innovation Prosperity*, 24(1), 1–16. <https://doi.org/10.12776/QIP.V24I1.1430>
- Pazour, J. A., & Meller, R. D. (2011). An analytical model for A-frame system design. *IIE Transactions*, 43(10), 739–752. <https://doi.org/10.1080/0740817X.2010.549099>
- Project Management Institute. (2017). *Guía de los Fundamentos para la Dirección de Proyectos (Guía del PMBOK)* (6ta ed.). Pennsylvania, USA: Project Management

Institute. Recuperado de [https://topodata.com/wp-content/uploads/2019/10/GUIA\\_PMBok.pdf](https://topodata.com/wp-content/uploads/2019/10/GUIA_PMBok.pdf)

Radhi, M., Ryan Hamonangan Sitompul, D., Hamonangan Sinurat, S., & Indra, E. (2021). ANALISIS BIG DATA DENGAN METODE EXPLORATORY DATA ANALYSIS (EDA) DAN METODE VISUALISASI MENGGUNAKAN JUPYTER NOTEBOOK. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima*, 4(2).

D. Santos. (2022). Proyect charter: qué es, cómo crearlo y ejemplos. Campus Virtual UCI <https://omeka.campusuci2.com/biblioteca/files/original/55d75aeb690cdabf85521c8fc b502220.pdf>

Ramos. S. (2018). *Gestion de los interesados en un proyecto de implantación de ERP*. Universidad Técnica Superior de Valencia

S. A. Jernigan. (2004). *Multi-Tier Inventory Systems With Space Constraints*. PhD thesis, Georgia Institute of Technology, Atlanta, GA.

SSI Schäfer. (s.f.). *A-Frame*. <https://www.ssi-schaefer.com/es-es/productos/soluciones-de-picking/picking-automatico/a-frame-1215096>

SSI Schäfer. (s.f.). *A-Frame*. <https://www.ssi-schaefer.com/resource/blob/6040/fc06e2f912a8b87f7d77c080e90248e5/brochure-a-frame-en--dam-download-en-1685--data.pdf>

Yancuta. F.(2020). *Marco teórico para diagnóstico y propuesta de mejora en una línea de producción de galletas de una empresa de consumo masivo empleando herramientas de la filosofía lean manufacturing*. Pontificia universidad católica del Perú facultad de ciencias e ingeniería.

## Anexos

### Anexo 1: Explicación herramientas de DMAIC para la fase de medir y entendimiento del negocio

- **5 ¿Por qué?**

Con información obtenida a través de Sophia Nieto (2024) se tiene el siguiente diagrama de 5 ¿Por qué?, con el objetivo de encontrar la causa raíz del problema y definirlo:

**Tabla Anexo 1.1**

<b>Pregunta</b>	<b>Respuesta</b>
<b>¿Por qué ha disminuido el porcentaje de líneas despachadas por los A-Frames?</b>	Porque ha disminuido la cantidad asignada de SKUs en los A-Frames
<b>¿Por qué ha disminuido la cantidad asignada de SKUs en los A-Frames?</b>	Porque mediante el proceso de cambios de ubicación, a lo largo del tiempo los SKUs que estaban en los A-Frames han sido reubicados
<b>¿Por qué mediante el proceso de cambios de ubicación, a lo largo del tiempo los SKUs que estaban en los A-Frames han sido reubicados?</b>	Porque el proceso de cambios de ubicación obedece a criterios subjetivos de los operadores
<b>¿Por qué el proceso de cambios de ubicación obedece a criterios subjetivos de los operadores?</b>	Porque hace falta una estandarización de los criterios y una organización de los SKUs asignados
<b>¿Por qué hace falta una estandarización de los criterios y una organización de los SKUs asignados?</b>	Porque un SKU mal ubicado, según sus características, donde no corresponde. No permite aprovechar las ventajas de los equipos de picking.

Elaborado por Acosta et al, (2024). 5 ¿Por qué?

La causa raíz obtenida a través de 5 ¿Por qué? es: Existe una falta de estandarización y precisión al momento de asignar los SKUs a sus ubicaciones. En base a esto definimos el objetivo de este proyecto: Mejorar la asignación de SKUs a los canales de los A-Frames.

- **Project Charter**

En base a lo antes mencionado en esta parte de diagnóstico junto con la información recolectada en el Gemba Walk se tiene el siguiente *Project Charter*:

### Tabla Anexo 1.2

#### *Acta de constitución del proyecto (Project Charter)*

Nombre del proyecto: Mejorar la asignación de SKUs a los canales de los A-Frames en el Centro de Distribución de Corporación GPF					
Planteamiento del problema	Justificación y beneficios				
Actualmente, ha disminuido la cantidad de líneas despachadas por los equipos de <i>picking</i> automático A-Frames, como resultado de los cambios de ubicación de los productos, que se realizan semanalmente. Esto ha resultado en el uso ineficiente de los A-Frames en el <i>picking</i> de los productos de mayor rotación.	Vale la pena realizar este proyecto debido a que si los A-Frames tienen los productos adecuados la empresa mejora del uso de sus equipos y de la preparación de los pedidos. Es importante realizarlo ahora debido a que Corporación GPF puede llegar a aumentar la eficiencia de sus operaciones logísticas. El proyecto encaja con la visión de la empresa de aprovechar al máximo las ventajas de sus equipos automatizados dentro de su centro de logística avanzada.				
Objetivos	Cronograma				
Mejorar la asignación de SKUs a los canales de los A-Frames	Fase		Fecha		
	Visitas		agosto - septiembre		
	Limpieza		septiembre		
	EDA		septiembre		
	Desarrollar modelo		septiembre - octubre		
	Análisis de resultados		octubre		
	Pruebas y correcciones		octubre - diciembre		
	Alcance	Miembros del equipo			
1er paso de proceso	Recorrido general del centro	Rol	Nombre	Cargo	% de tiempo

	logístico avanzado				
Último paso del proceso	Presentar la propuesta de proyecto al jefe de productividad y procesos logísticos	Miembro del grupo	Camilo Acosta	Ingeniero Industrial	100
Alcance	Mejorar la eficiencia de los equipos de <i>picking</i> automático A-Frames	Miembro del grupo	Rocío Castro	Ingeniera Industrial	100
Fuera de alcance	Recursos no aprobados por parte del gerente de corporación GPF	Miembro del grupo	Sabrina Cárdenas	Ingeniera Industrial	100

Elaborado por Acosta et al, (2024). Project Charter

- **Matriz de análisis de FODA**

### Tabla Anexo 1.3

Matriz de Decisiones estratégicas FODA

Amenazas	Fortalezas
<ul style="list-style-type: none"> <li>• Demandas variables por picos semanales.</li> <li>• Canales encerrados.</li> <li>• Limitada capacidad de número de cambios de ubicación por semana.</li> <li>• Riesgo de contaminación cruzada</li> <li>• Cortes de luz: Impiden operatividad del A-Frame.</li> </ul>	<ul style="list-style-type: none"> <li>• Contar con una base de datos estructurados con información de los SKUs. (morfología, línea, unidades, nivel del servicio)</li> <li>• Apertura de admirativos en implementación de proyectos de mejora con colaboración de tesisas.</li> <li>• Sistemas de hardware y software son robustos en cuanto a validez de los datos.</li> </ul>
Debilidades	Oportunidades
<ul style="list-style-type: none"> <li>• No existe un proceso de selección de SKUs establecido para asignar en los canales del A-Frame.</li> </ul>	<ul style="list-style-type: none"> <li>• Aumento del porcentaje del tiempo del canal lleno.</li> <li>• Reducción de Reprocesamiento de órdenes. (Batch)</li> </ul>

<ul style="list-style-type: none"> <li>Proceso de cambios de ubicación depende del recurso limitado del Departamento del Cambio de Ubicaciones.</li> </ul>	<ul style="list-style-type: none"> <li>Optimización de SKUs según criterios.</li> </ul>
--	---

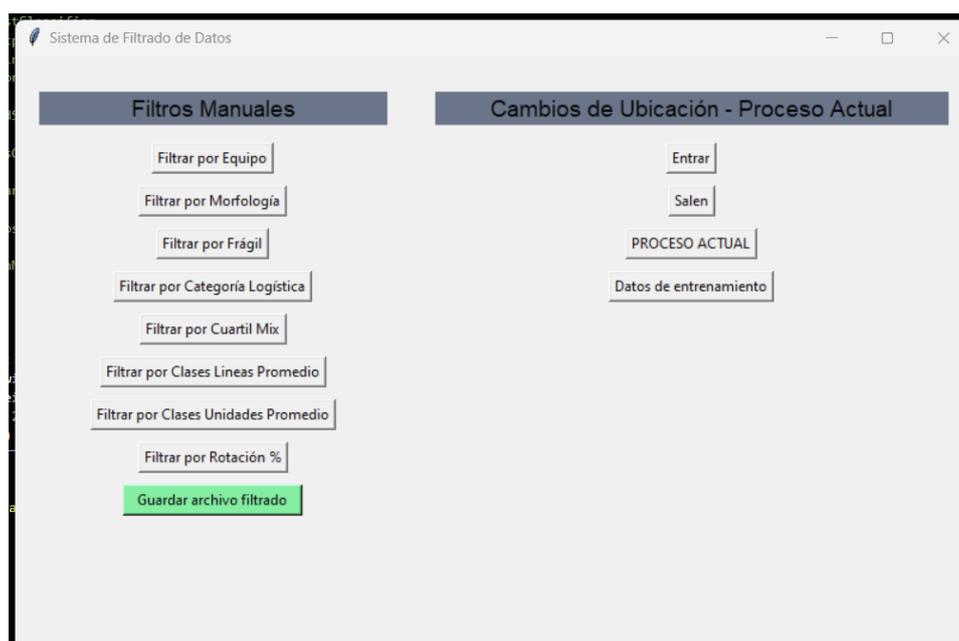
Elaborado por Acosta et al, (2024). *Matriz FODA enfocado en proceso de picking automático A-Frame [Matriz]*

## Anexo 2: Explicación Aplicación proceso actual

La aplicación consta de dos partes, a lado izquierdo existen los Filtros Manuales. Donde cada botón muestra un factor o característica de filtrado de los SKUs del CLA. Una vez hecho el filtrado especializado o particular el botón verde permite guardar el archivo filtrado en un archivo .csv

La parte de la derecha tiene los siguientes botones, donde el procesado actual se ejecuta automáticamente:

- Entrar: Devuelve un archivo .csv, con todos los posibles SKUs que entran.
- Salen: Devuelve un archivo .csv, con todos los posibles SKUs que salen.
- PROCESO ACTUAL: Devuelve dos archivos .csv, uno con la lista de los que salen del A-Frame, otro con los que entran al A-Frame. Pero la suma de los SKUs seleccionados en ambos archivos es 100, cumpliendo la limitación de los 100 cambios de ubicación efectuados.
- Datos de entrenamiento: Permite filtrar la base de datos con SKUs ideales para hacer el entrenamiento posterior de *Machine Learning*.



## Anexo 3: Documentación del código

**Función 1.1:** `generar_listas_equidistantes(umbral, uno, max)`

**Objetivo general:** generar etiquetas (label) y límites (limites) para definir clases o rangos de unidades de tres valores y entrada. Y facilita hacer el proceso de hacer el ABC del producto.

#### Parámetros de entrada:

- `umbral`: es un valor que se utilizará como el límite inferior de la segunda clase.
- `uno`: representa el límite superior de la segunda clase y el límite inferior de la tercera.
- `max`: es el valor máximo y define el límite superior de la última clase.

#### Variables de salida:

- `límites`: permiten entender y delimitar las clases

#### Desglose de la función:

- Construcción de la lista de límites (lista)

```
lista = [0, umbral, uno, max]
```

Se crea una lista de cuatro elementos, el primer valor es 0, el siguiente valor es (`umbral`), ingresado por el usuario, seguido del valor (`uno`) y (`max`). Esta lista se usará para crear los intervalos según el orden.

- Generación de etiquetas (label):

```
label = [f"{lista[i]}-{lista[i+1]}" for i in range(len(lista) - 1)]
```

Se crea una lista de etiquetas que describen los intervalos entre los elementos consecutivos de la lista (`lista`) por ejemplo, si `lista` es `[0, 10, 20, 30]`, las etiquetas generadas serán `["0-10", "10-20", "20-30"]`.

#### Código de función:

```
# Función para generar listas equidistantes
def generar_listas_equidistantes(umbral, uno, max):
    lista = [0, umbral, uno, max]
    label = [f"{lista[i]}-{lista[i+1]}" for i in range(len(lista) - 1)]
    limites = lista
    return label, limites
```

**Función 1.2:** `ventana_filtro(df, columna)`

**Objetivo general:** Crear una ventana de filtrado de opciones en una interfaz gráfica usando la librería (`tkinter`), esto permitirá al usuario seleccionar opciones específicas de una columna en un *DataFrame* para filtrar datos. Vamos a revisar cada parte de esta función.

#### Parámetros de entrada:

1. `df`: es un *DataFrame* de `pandas` que contiene los datos de donde se extraerán las opciones de filtrado.
2. `columna`: es el nombre de la columna en el *DataFrame* (`df`) sobre la cual se quiere aplicar el filtro.

#### Variables de salida:

- `getattr(ventana_filtro, 'df_filtrado', df)`: Dataframe que será usado en el contexto del menú de botones, con el filtro respectivo, por ejemplo: filtro de Equipo, filtro de Qmix, etc.

### Desglose de la función:

Creación de la ventana emergente (top):

```
top = tk.Toplevel()
top.title(f"Filtrar por {columna}")
```

Se crea una ventana emergente (Toplevel) que será independiente de la ventana principal. El título de la ventana será "Filtrar por [columna]", donde [columna] representa el nombre de la columna que se está filtrando.

Extracción de opciones únicas:

```
opciones = df[columna].dropna().unique().tolist()
se obtienen los valores únicos de la columna que especifico el usuario, es decir, sin valores únicos, y se convierten en una lista llamada (opciones.) Estas serán
```

Creación de checkboxes:

```
checks = {}
for opcion in opciones:
    var = tk.BooleanVar()
    chk = tk.Checkbutton(top, text=opcion, variable=var)
    chk.pack(anchor="w")
    checks[opcion] = var
```

Se crea un nuevo diccionario vacío en el que almacenara las variables del checkbutton

- Un Checkbutton, que representa una casilla de verificación con el texto de la opción y que está vinculada a var para que el estado de cada checkbox se pueda leer después.
- Cada Checkbutton se agrega a la ventana (top) con `pack(anchor="w")`, que alinea las casillas a la izquierda.

La variable var se almacena en el diccionario checks con la opción como clave, lo que permite recuperar el estado de cada checkbox al final.

### Código de función:

```
# Ventana de filtrado de opciones
def ventana_filtro(df, columna):
    top = tk.Toplevel()
    top.title(f"Filtrar por {columna}")

    # Opciones únicas de la columna
    opciones = df[columna].dropna().unique().tolist()

    # Crear lista de checkboxes para cada opción
    checks = {}
```

```

for opcion in opciones:
    var = tk.BooleanVar()
    chk = tk.Checkbutton(top, text=opcion, variable=var)
    chk.pack(anchor="w")
    checks[opcion] = var

# Función de aplicar selección
def aplicar_seleccion():
    seleccionadas = [opcion for opcion, var in checks.items() if var.get()]
    if seleccionadas:
        ventana_filtro.df_filtrado = aplicar_filtro(df, columna,
seleccionadas)
        top.destroy()
    else:
        messagebox.showwarning("Advertencia", "Debes seleccionar al menos una
opción.")
    # Botón para aplicar el filtro
    btn_aplicar = tk.Button(top, text="Aplicar", command=aplicar_seleccion)
    btn_aplicar.pack()
    # Esperar a que la ventana de filtro se cierre
    top.wait_window()
    return getattr(ventana_filtro, 'df_filtrado', df)

```

### **Función 1.3:** filtro\_rotacion(df)

**Objetivo general:** Permite obtener la rotación usando la base de Despachos y la une al dataframe final.

#### **Parámetros de entrada:**

**df:** el dataframe filtrado previo

#### **Variables de salida:**

**df:** el dataframe con la columna de rotación

#### **Código de función:**

```

# Aplicar filtro por rotación
def filtro_rotacion(df):
    top = tk.Toplevel()
    top.title("Filtrar por Rotación %")

```

```

    lbl_inicio = tk.Label(top, text="Ingrese el valor inicial (entre 0 y
100):")
    lbl_inicio.pack()
    entry_inicio = tk.Entry(top)
    entry_inicio.pack()

    lbl_fin = tk.Label(top, text="Ingrese el valor final (entre el valor
inicial y 100):")
    lbl_fin.pack()
    entry_fin = tk.Entry(top)
    entry_fin.pack()

def aplicar_rango():
    try:
        inicio = int(entry_inicio.get())
        fin = int(entry_fin.get())
        if 0 <= inicio <= fin <= 100:
            ventana_filtro.df_filtrado = df[(df['Rotacion %'] >=
inicio) & (df['Rotacion %'] <= fin)]
            top.destroy()
        else:
            messagebox.showerror("Error", "Valores fuera de rango.")
    except ValueError:
        messagebox.showerror("Error", "Entrada inválida. Ingrese
números enteros.")

    btn_aplicar = tk.Button(top, text="Aplicar", command=aplicar_rango)
    btn_aplicar.pack()
    top.wait_window()

return getattr(ventana_filtro, 'df_filtrado', df)

```

#### **Anexo 4. Tabla: Problemas típicos de análisis de datos empresariales y técnicas para resolverlos.**

Tipo de problema	Perspectiva de análisis	Técnicas
------------------	-------------------------	----------

Comparar decisiones comerciales	<ul style="list-style-type: none"> <li>• ¿Una nueva etiqueta de whisky aumenta las ventas?</li> <li>• ¿Existe alguna diferencia entre dos campañas de clientes en la tienda online?</li> </ul>	Pruebas de hipótesis Pruebas A/B
Encontrar Clústers y valores atípicos	<ul style="list-style-type: none"> <li>• ¿Cuál es el whisky más parecido al Junflivet?</li> <li>• ¿Podemos agrupar a nuestros clientes en grupos distintos y diferenciados por características de marketing?</li> </ul>	Clustering Nearest Neighbour Analysis
Encontrar reglas y relaciones	<ul style="list-style-type: none"> <li>• Si un cliente compra nuestro whisky Junflivet de 10 años ¿qué whisky comprará a continuación?</li> <li>• ¿Qué productos se suelen comprar juntos?</li> </ul>	Análisis de asociación
Predecir explicar valores categóricos	<ul style="list-style-type: none"> <li>• ¿El cliente de nuestra tienda es solvente o no?</li> <li>• ¿Qué productos se suelen comprar juntos?</li> </ul>	Árbol de decisión Regresión logística
Predecir y explicar valores numéricos	<ul style="list-style-type: none"> <li>• ¿Cuánto aumenta la venta en una nueva etiqueta?</li> <li>• ¿Comprar la fluencia de diferentes factores en el volumen de las ventas?</li> </ul>	Análisis de regresión
Predecir y analizar los acontecimientos	<ul style="list-style-type: none"> <li>• ¿Cómo evolucionará el número de ventas de nuestros productos?</li> <li>• ¿Qué desarrollos futuros son probables?</li> </ul>	Series de previsión temporales

Recuperado de: Jung, D. (2024). Figura 5.2 [Problemas típicos de análisis de datos empresariales y técnicas para resolverlos.]

### Anexo 5: Matriz de confusión

		Predicciones del modelo	
		Negativo	Positivo
Valor Real	Negativo	Verdadero negativo VN	Error tipo 1 Falso positivo FP
	Positivo	Error tipo 2 Falso Negativo FN	Verdadero Positivo VP

Recuperado de: Jung, D. (2024). Figura 5.12 [Matriz de confusión para calcular el rendimiento de los modelos de clasificación.]

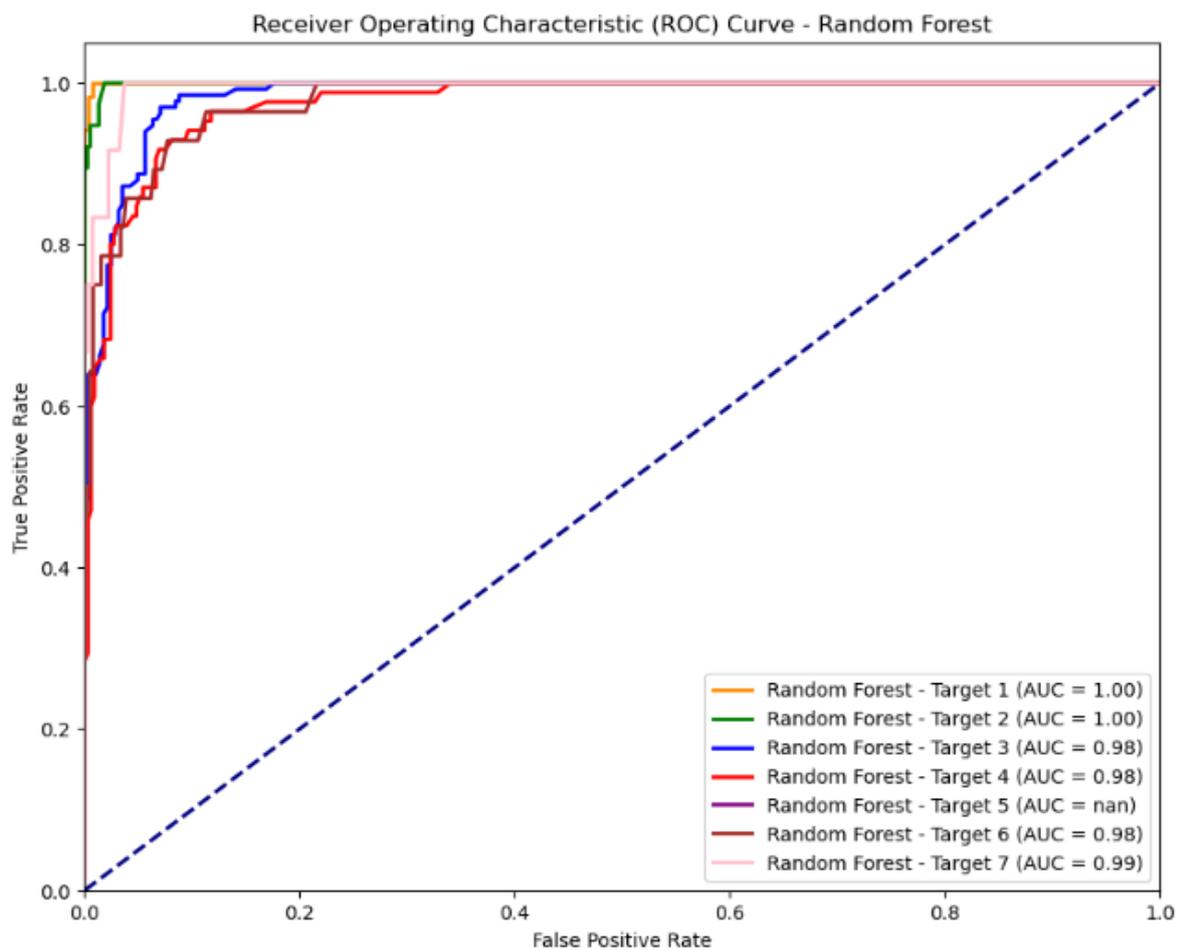
Donde:

- Verdaderos positivos (VP): el número de casos que el modelo predijo correctamente la clase, fueron positivos según la verdad fundamental.
- Verdaderos negativos (VN): el número de casos donde el modelo predijo correctamente la clase negativa y fueron verdad.
- Falsos positivos (FP): el número de casos donde el modelo predijo la clase positiva, pero en realidad fueron negativos
- Falsos negativos (FN): el número de casos donde el modelo predijo la clase negativa, pero en realidad fueron positivos

(Jung, 2024 p.181)

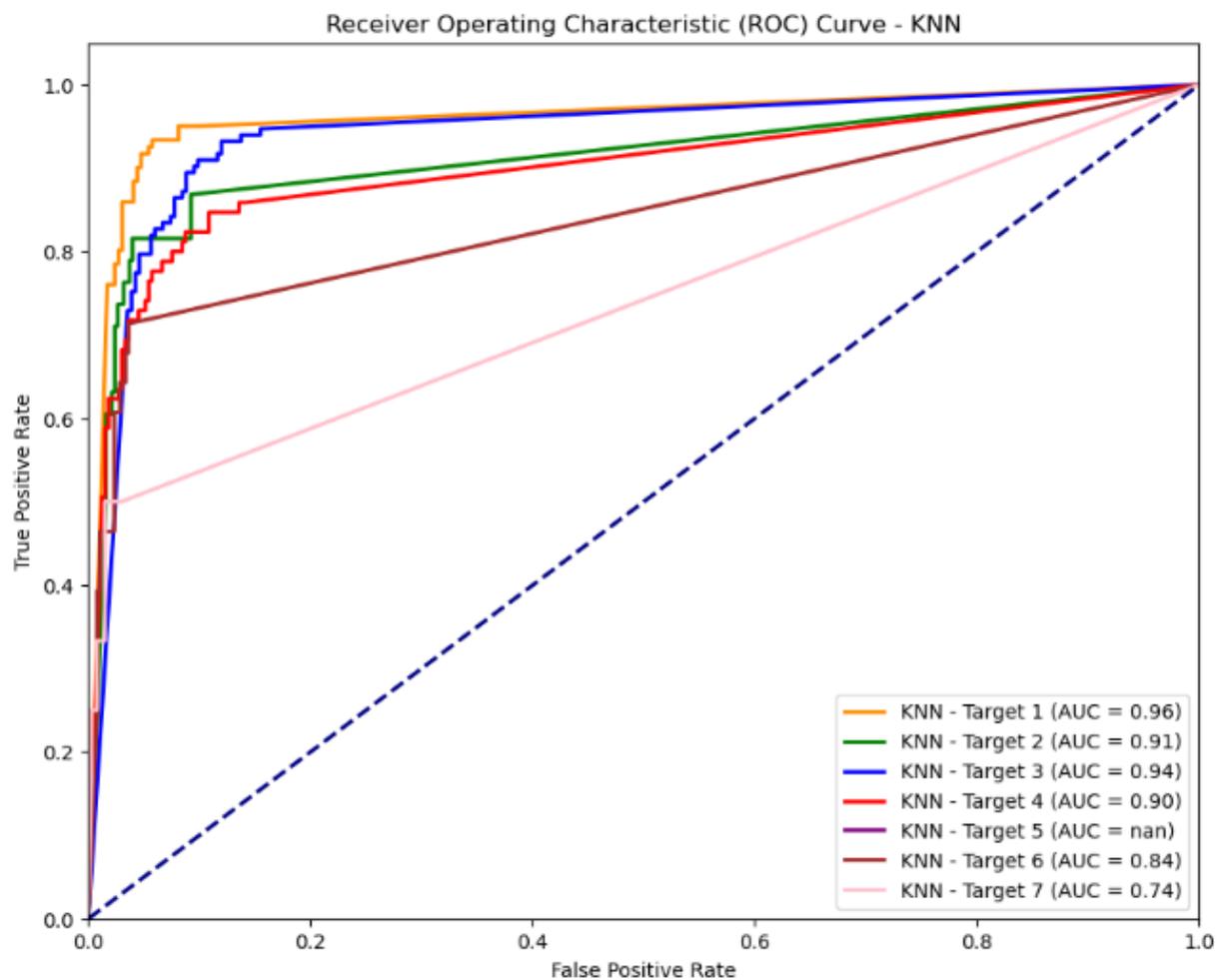
## Anexo 6: Gráficos Curvas ROC

### Gráfico ROC- Random Forest



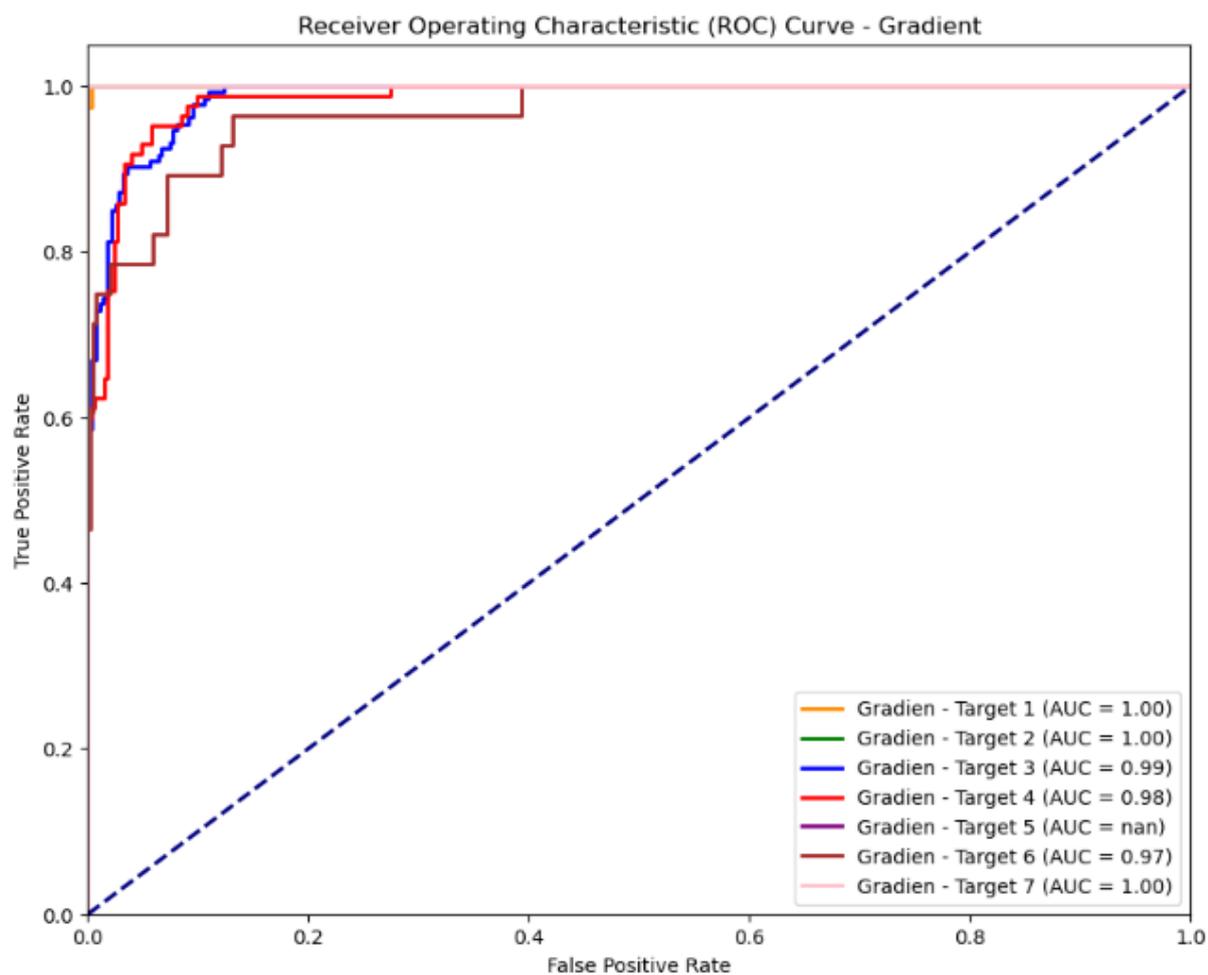
*Nota. Elaborado por Acosta et al, (2024). Gráfico ROC- Random Forest. Usando Matplotlib.*

### Gráfico ROC- KNN



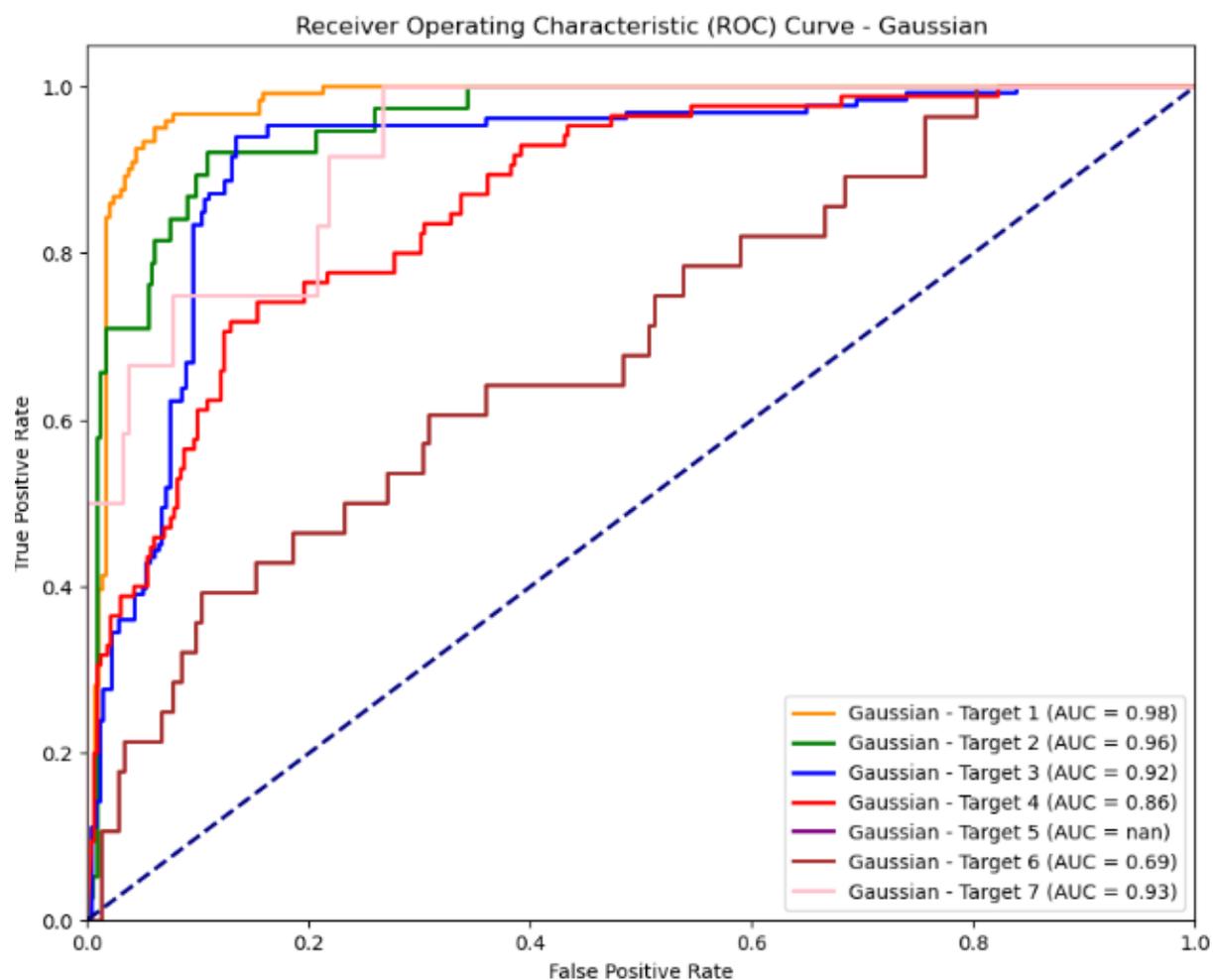
*Nota. Elaborado por Acosta et al, (2024). Gráfico ROC- KNN. Usando Matplotlib*

### **Gráfico ROC- Gradient Boosting Machine**



*Nota. Elaborado por Acosta et al, (2024). Gráfico ROC- GBM. Usando Matplotlib*

**Gráfico ROC- Naive Bayes**



*Nota. Elaborado por Acosta et al, (2024). Gráfico ROC- Naive Bayes. Usando Matplotlib.*