Programación de rutas mediante optimización de políticas proximales para la recolección de cajas en la sabana de Bogotá

Juan Camilo Fandiño Coba

Asesor

Rafael Gaitán Ospina

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2025

Resumen

Este proyecto busca optimizar la programación de rutas terrestres para la recolección de cajas de flores de la sabana de Bogotá, con un enfoque en mejorar la eficiencia operativa mediante técnicas avanzadas de aprendizaje automático. En la actualidad, los métodos de planificación utilizados enfrentan desafíos significativos, como el uso subóptimo de la capacidad vehicular, tiempos extensos de recorrido, y la necesidad de múltiples vehículos para cumplir con los requisitos de recolección.

La metodología propuesta integra técnicas de aprendizaje automático y modelos de optimización para abordar el problema de rutas de vehículos (VRP). Los principales indicadores de evaluación incluyen el tiempo promedio requerido para calcular las soluciones, la capacidad promedio de los vehículos utilizada, el tiempo total de las rutas, y el número de vehículos necesarios. A través de simulaciones y pruebas, el proyecto pretende desarrollar una solución que maximice la eficiencia, reduciendo tanto los costos operativos como la complejidad logística. Este enfoque permitirá una toma de decisiones más informada y rápida, adaptándose a las características dinámicas del entorno y las restricciones específicas de la operación en la industria floricultora.

***Palabras clave*:** Optimización, VRP, Machine Learning, PPO, logística.

Abstract

This project aims to optimize the scheduling of ground routes for flower box collection from farms in the Bogotá savanna, using advanced machine learning techniques and optimization models to enhance operational efficiency. Current planning methods face significant challenges, such as suboptimal vehicle capacity utilization, long travel times, and the need for multiple vehicles to meet collection requirements.

The proposed methodology integrates machine learning techniques with vehicle routing problem (VRP) models, focusing on key performance indicators such as average solution computation time, average vehicle capacity utilization, total route time, and the number of vehicles required. Through simulations and testing, the project seeks to develop a solution that maximizes efficiency, reduces operational costs, and simplifies logistics. This approach enables faster and more informed decision-making, adapting to the dynamic conditions and specific constraints of the floriculture industry.

***Keywords*:** Optimization, VRP, Machine Learning, PPO, logistics.

Tabla de Contenido

Lista de Tablas

Lista de Figuras

Introducción

La industria floricultora en Colombia, especialmente en la sabana de Bogotá, enfrenta desafíos logísticos complejos derivados del manejo de productos altamente perecederos, altos volúmenes de producción y restricciones operativas específicas (Vargas Ramírez, Barrera Bello, & Cruz Martín, 2019). Uno de los procesos críticos es la programación de rutas para la recolección de cajas de flores, el cual actualmente se realiza de forma manual, demandando aproximadamente dos horas diarias para su planificación. Esta metodología tradicional no solo es ineficiente en términos de tiempo, sino que también puede conducir a un uso subóptimo de la flota vehicular, mayor número de kilómetros recorridos y dificultades para cumplir con ventanas de tiempo de recolección.

En respuesta a estas limitaciones, este proyecto propone una solución basada en técnicas de Aprendizaje por Refuerzo Proximal (Proximal Policy Optimization, PPO), implementadas en un entorno personalizado desarrollado con la librería Gym de OpenAI. El entorno simula un escenario real compuesto por 30 vehículos y 350 órdenes de recolección, permitiendo modelar con precisión las condiciones operativas. A través del entrenamiento del agente, se busca optimizar múltiples objetivos: minimizar los kilómetros recorridos, maximizar la ocupación de los vehículos y reducir la cantidad de vehículos utilizados.

Uno de los principales aportes del proyecto radica en la eficiencia obtenida durante la etapa de inferencia del modelo entrenado, alcanzando tiempos de respuesta del orden de 30 segundos, lo cual representa una mejora sustancial frente al proceso actual. Además, se evaluaron indicadores clave de desempeño para validar la efectividad de la solución y su viabilidad para ser implementada en la operación real.

Esta investigación no solo demuestra el potencial de aplicar técnicas avanzadas de inteligencia artificial en contextos logísticos tradicionales, sino que también abre la puerta a futuras mejoras e integraciones con sistemas existentes, fomentando la transformación digital de procesos críticos en la cadena de suministro de flores.

Descripción del Problema

En la operación logística de recolección de cajas con flores en las sedes ubicadas en la sabana de Bogotá, uno de los principales retos es la adecuada planificación de las rutas terrestres hacia el aeropuerto para su posterior exportación. Actualmente, los procesos de asignación y secuenciación de rutas presentan altos niveles de ineficiencia debido a que se realizan de manera manual o con herramientas que no consideran de manera integral todas las variables operativas relevantes. Esto se traduce en costos operativos elevados, mayores tiempos de recorrido, sub-utilización de la capacidad vehicular y, en numerosos casos, incumplimiento en las ventanas de tiempo establecidas para las entregas en el aeropuerto.

El incumplimiento de estas ventanas no solo impacta la eficiencia interna, sino que también compromete la calidad del servicio con los diferentes actores de la cadena, afecta la frescura del producto —un factor crítico en el negocio de flores—. Adicionalmente, el tráfico fluctuante de la sabana, las restricciones de acceso a ciertas sedes y las variaciones en el volumen diario de cajas recolectadas incrementan la complejidad de la planificación de rutas.\\

Por tanto, la pregunta que guía este proyecto es:

¿Cómo se puede aplicar el aprendizaje automático para optimizar la planificación de rutas en la recolección de cajas de flores en las sedes de la sabana de Bogotá, con el fin de reducir los costos operativos, minimizar los tiempos de recorrido y mejorar el cumplimiento de las ventanas de tiempo de entrega?

Justificación

En la logística de recolección de flores en las sedes de la sabana de Bogotá, una planificación eficiente de rutas resulta fundamental debido a las características particulares del proceso:

- Producto altamente perecedero: Las flores requieren un manejo ágil y cuidadoso para preservar su calidad y frescura.
- Dispersión geográfica: Las sedes están distribuidas a lo largo de un amplio territorio, lo que aumenta la complejidad en la coordinación de rutas óptimas.
- Ventanas de tiempo estrictas: El cumplimiento riguroso de los horarios de transporte hacia el aeropuerto es esencial para garantizar que los despachos lleguen a tiempo a los mercados internacionales.

Actualmente, la planificación se realiza de manera manual, lo que genera diversas ineficiencias:

- Altos costos operativos: La programación manual demanda tiempo considerable y, en muchos casos, las rutas definidas no son óptimas, lo que incrementa la duración de las operaciones y la cantidad de viajes necesarios.
- Impacto en la calidad del servicio: Los retrasos en la programación afectan la eficiencia en la sede, generando tiempos adicionales de espera para el personal y retrasos documentales.

Frente a estos desafíos, la aplicación de técnicas de aprendizaje automático representa una solución poderosa y eficaz, por las siguientes razones:

Optimización de rutas: Mediante algoritmos evolutivos es posible identificar rutas que minimicen costos operativos, considerando múltiples restricciones como:

- Capacidad de los vehículos
- Ventanas de tiempo por sede
- Distancias entre puntos de recolección y destino
- Agrupación inteligente de sedes

La integración del aprendizaje automático en la logística de recolección no solo permite optimizar los recursos de manera técnica, sino que aporta una ventaja competitiva tangible. Su implementación tendría un impacto directo en:

- La reducción de los costos operativos.
- La mejora de la puntualidad en las entregas.
- La preservación de la calidad del producto en su llegada al aeropuerto.

En este contexto, el uso de técnicas de aprendizaje automático se presenta como una solución viable, escalable y estratégica para enfrentar los retos logísticos actuales, fortaleciendo la competitividad del sector floricultor de la región.

Objetivos

Objetivo General

Construir un modelo de programación de rutas basado en aprendizaje por refuerzo para la recolección de cajas de flores en la sabana de Bogotá, mejorando la eficiencia operativa, el uso de los vehículos disponibles y el cumplimiento de las ventanas de tiempo acordadas.

Objetivos Específicos

Identificar patrones clave que afecten la eficiencia operativa y los tiempos de entrega con base en los datos históricos.

Desarrollar un modelo de optimización de rutas utilizando Aprendizaje por Refuerzo, específicamente con el algoritmo Proximal Policy Optimization (PPO), teniendo en cuenta restricciones de tiempo y capacidad.

Evaluar el impacto del modelo en términos de reducción de costos operativos, mejora de tiempos de recorrido y cumplimiento de ventanas de tiempo, comparándolo con el proceso actual de programación.

Marco de Referencia

En el problema de rutas de vehículos (VRP), tenemos varios vehículos disponibles para visitar un conjunto determinado de vértices, pero cada vértice debe visitarse exactamente una vez. Además, los vehículos tienen una capacidad fija, cada vértice tiene una demanda dada y el objetivo es encontrar la distancia total mínima necesaria para visitar todos los vértices exactamente una vez sujetos a la restricción de capacidad (Kou, Golden, & Bertazzi, 2024).

Este es un problema de optimización combinatoria que ha sido estudiado en matemáticas aplicadas y ciencias de la computación durante décadas. Aunque se han desarrollado numerosos algoritmos exactos y heurísticos para abordarlo, la tarea de ofrecer soluciones rápidas y confiables continúa representando un desafío significativo debido a la complejidad computacional inherente al VRP (Ara, y otros, 2023).

Por su parte, el aprendizaje automático es una rama de la inteligencia artificial que permite que las maquinas puedan adquirir nuevas habilidades y mejorar con le tiempo sin ser necesariamente programado para ello (Mishra & Tyagi, 2022). En este proyecto se usarán diferentes técnicas de Machine learning par abordar el problema de rutas de vehículos con el fin de optimizar el proceso de programación de rutas terrestres.

El aprendizaje por refuerzo, el cual es una rama de aprendizaje automático, se ha utilizado para resolver problemas de ruteo al formular el VRP como un problema de optimización secuencial. Por ejemplo, en (Ara, y otros, 2023) se usaron enfoques basados en redes neuronales recurrentes (RNN) con mecanismos de atención que pueden modelar las decisiones dinámicas requeridas para satisfacer las demandas de los clientes y minimizar los costos totales del ruteo. También existen métodos híbridos como los mostrados en (Shahbazian, Pugliese, Guerriero, & Macrina, 2024) que combinan técnicas de machine learning con

algoritmos tradicionales como la optimización por colonia de hormigas, algoritmos genéticos o métodos de búsqueda local. Otro enfoque asociado al deep learning se encuentra en arquitecturas profundas como Graph Neural Networks (GNN) para representar las relaciones espaciales y de demanda en el VRP. Estas redes aprenden representaciones directamente de la estructura del problema, lo que permite capturar mejor las características complejas de instancias reales del VRP. Estos métodos demuestran cómo el aprendizaje automático está transformando la resolución de problemas de ruteo al incorporar dinámicas más complejas y variabilidad en los datos.

Dentro de los algoritmos más recientes y efectivos de aprendizaje por refuerzo se encuentra el Proximal Policy Optimization (PPO), desarrollado por OpenAI. PPO es un método basado en gradiente de política que mejora la estabilidad del entrenamiento al limitar la magnitud de las actualizaciones de la política, lo cual evita cambios abruptos que podrían desestabilizar el aprendizaje (Schulman, Wolski, Dhariwal, Radford, & Klimov, 2017). Utiliza una función objetivo recortada que garantiza actualizaciones conservadoras, siendo más simple de implementar y computacionalmente más eficiente que otros enfoques como Trust Region Policy Optimization (TRPO). Gracias a estas características, PPO ha sido ampliamente adoptado en tareas complejas como el control robótico y entornos con múltiples restricciones dinámicas. Su aplicabilidad a problemas de ruteo se debe a su capacidad de adaptarse a secuencias de decisiones en entornos no deterministas, optimizando no solo el camino sino también el comportamiento del agente en escenarios logístico (Hugging Face, 2023).

La integración de herramientas de análisis de datos es esencial para abordar problemas logísticos como el VRP. Fuentes como dispositivos GPS, telemetría de vehículos y patrones históricos de demanda permiten alimentar modelos de aprendizaje automático con datos precisos

y actualizados. Estos datos no solo mejoran la calidad de las predicciones, sino que también posibilitan la adaptación de las rutas en tiempo real, una necesidad creciente en entornos logísticos dinámicos (Oyola, Arntzen, & Woodruff, 2016).

El problema de programación de rutas en logística es fundamental debido a su impacto directo en la eficiencia operativa, los costos y la sostenibilidad ambiental de las empresas. La optimización de rutas minimiza las distancias recorridas y, por ende, el consumo de combustible. Esto no solo reduce costos sino también mejora la competitividad de las empresas en un mercado altamente demandante (Bogyrbayeva, Meraliyev, Mustakhov, & Dauletbayev, 2024). Además del ahorro económico, la optimización de rutas logísticas tiene un impacto ambiental significativo al reducir el consumo de combustibles fósiles y las emisiones de CO_2. Esto es especialmente relevante en industrias como la floricultura, donde las prácticas sostenibles pueden fortalecer la relación con clientes que valoran productos de bajo impacto ambiental.

Estudios como los de (Rodríguez & Jaca, 2014) y (Ballou, 2004) han demostrado que mejorar la eficiencia del ruteo puede generar ahorros significativos en los gastos logísticos, los cuales representan entre el 10% y el 20% de los ingresos totales de muchas organizaciones.

Además, se estima que el transporte puede constituir hasta el 50% del total de los costos logísticos, por lo que su optimización tiene un impacto directo en la rentabilidad (Capgemini, University, & eyefortransport, 2017).

La planificación efectiva de rutas asegura la puntualidad en las entregas, un factor crucial para la satisfacción del cliente, esto es especialmente importante en industrias como la de productos perecederos, donde el tiempo es crítico (Ara, y otros, 2023).

Metodología

La metodología se estructura en cuatro fases principales diseñadas para abordar de manera sistemática el análisis, la preparación, el modelado y la implementación de un sistema de optimización de rutas basado en técnicas de aprendizaje automático. Este enfoque asegura una transición lógica desde la comprensión del problema hasta la validación de soluciones en un entorno real.

Fase de Análisis y Comprensión del Problema

En esta etapa inicial, se busca entender profundamente el contexto operativo y los desafíos específicos relacionados con la recolección de cajas de flores en la sabana de Bogotá. Esto incluye la identificación de restricciones logísticas clave, como ventanas de tiempo, capacidad de los vehículos y distancias, las cuales serán fundamentales para el diseño del modelo.

Además, se realiza la recopilación de datos relevantes, incluyendo tiempos de recolección, distancias entre puntos y datos de flota vehicular, asegurando la calidad y representatividad de la información para el análisis posterior. La revisión bibliográfica complementa esta fase, proporcionando una base teórica y técnica para el diseño del modelo mediante la exploración de enfoques existentes en la optimización de rutas.

Fase de Preparación de los Datos

La calidad de los datos es crucial para la construcción de modelos robustos. Esta fase se enfoca en garantizar que la información recopilada sea limpia, coherente y adecuada para su análisis. Se realizan procesos de limpieza, como la eliminación de valores nulos y corrección de errores de formato, y se transforman las variables mediante técnicas de escalamiento y estandarización.

Adicionalmente, los datos son segmentados para identificar patrones clave y organizar las sedes en grupos según su proximidad geográfica u otras características relevantes. Esto facilita el análisis y permite diseñar rutas más eficientes basadas en la agrupación de sedes.

Fase de Modelado

El modelado se centra en la aplicación de técnicas avanzadas para resolver el problema de enrutamiento de vehículos (VRP). En esta fase, se seleccionan y combinan herramientas como algoritmos de clustering para la agrupación de sedes, modelos de regresión para predecir tiempos de recorrido y algoritmos de optimización como los genéticos o búsqueda tabú para generar rutas eficientes.

Los modelos son entrenados y validados utilizando datos históricos, dividiendo la información en conjuntos de entrenamiento y prueba. Finalmente, se simulan escenarios basados en estos modelos, evaluando métricas clave como kilómetros recorridos, tiempo total de rutas y cumplimiento de ventanas de tiempo. Esto permite ajustar los modelos y garantizar su utilidad en condiciones reales.

Fase de Evaluación

El propósito de esta fase es evaluar el impacto del modelo desarrollado en los principales indicadores logísticos, tales como:

- Reducción en kilómetros recorridos.
- Cumplimiento de ventanas de tiempo.
- Uso eficiente de la capacidad vehicular.

Se analizan resultados bajo escenarios controlados utilizando datos históricos y/o escenarios en vivo, permitiendo identificar áreas de mejora o limitaciones del modelo propuesto.

Ajustes y Conclusiones

En esta última fase, se consolidan los hallazgos y se realizan ajustes necesarios para afinar el modelo o los enfoques empleados. Además, se documentan las conclusiones del proyecto, destacando su contribución potencial a la mejora de las operaciones logísticas y proponiendo líneas futuras de trabajo para abordar retos adicionales o perfeccionar el sistema.

Análisis Exploratorio – Datos Históricos

Descripción de los Datos

Origen de los Datos

Los datos provienen del sistema transaccional empresarial, en el cual se mantiene un histórico de hasta tres meses de las programaciones realizadas. La información asociada especificaciones de la red y la carga de parámetros como lo son las ubicaciones, tiempos, distancias, capacidades y demás también se obtienen de los sistemas transaccionales de la organización.

Variables Disponibles

Las vamos a dividir en dos grupos, variables que son las que cambian día a día y parámetros que son valores fijos.

Dentro de las variables tenemos:

1. Fecha: Hace referencia a la fecha programada de recolección
2. Sede: Son las siglas del nodo generador de carga
3. Placa: Es la placa del vehículo que va a realizar la recolección
4. Cantidad: Es la cantidad de unidades que necesitan recolectarse
5. Orden: Es la agrupación de cierta cantidad que necesita recolectarse y es indivisible

Por su parte los parámetros son:

1. Coordenadas: Hace referencia a la ubicación geográfica de la sede
2. Grupo: Hace referencia a la agrupación propia del sistema
3. Distancia: Es la distancia en metros entre las diferentes ubicaciones
4. Tiempo: Es el tiempo en segundos entre las diferentes ubicaciones
5. Capacidad: Es la capacidad en unidades de un vehículo

6. Tipo vehículo: Es la tipología del vehículo

7. Parqueadero: Hace referencia al parqueadero en el cual inicia el vehículo antes de la programación

Volumen de la Muestra

En este caso se tomó una muestra de 30 ejercicios de días continuos, lo que se representa con 15.949 registros.

Limpieza de Datos

En este componente se realizó un ajuste en la información histórica dado que se observaron los siguientes escenarios:

- Ordenes solo de Bogotá: Los datos provienen de diferentes zonas en donde se realiza el proceso de recolección, este proyecto está enmarcado en la zona específica de Bogotá por lo cual solo se deben dejar los nodos asociados a esta región
- Ordenes sin recolección: Existen algunos registros de ordenes que no tienen una placa asociada para el proceso de recolección, esto se debe a algunas ordenes que no se programan directamente por la organización o que se movieron de día.

Análisis Exploratorio de Datos

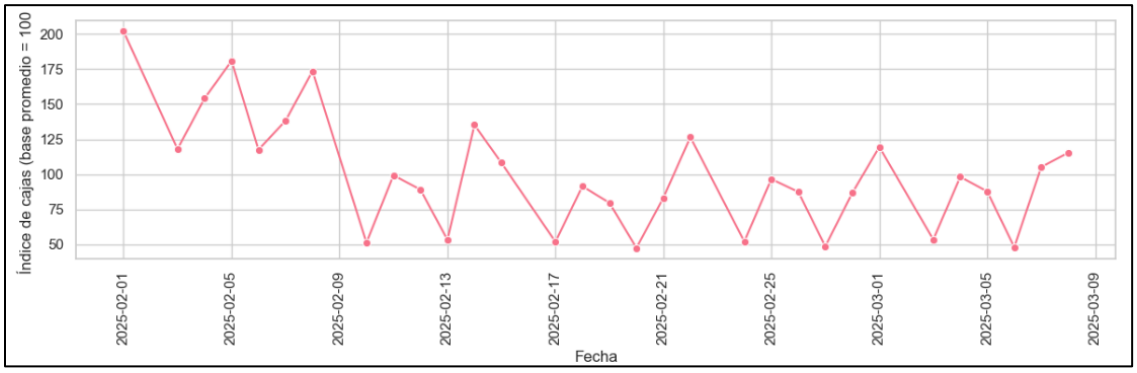
Este análisis se dividió en tres grupos con el fin de entender mejor los datos:

Distribución y Comportamiento de los Datos

En la **Figura 1 se** graficó la cantidad de cajas que se programaron para cada día de la muestra de datos disponible. Para proteger la confidencialidad de la información operativa, no se muestran los valores absolutos de cajas, sino que se utiliza un índice relativo calculado con base en el promedio del periodo analizado. Esta representación permite identificar patrones y tendencias sin revelar datos sensibles.

Figura 1

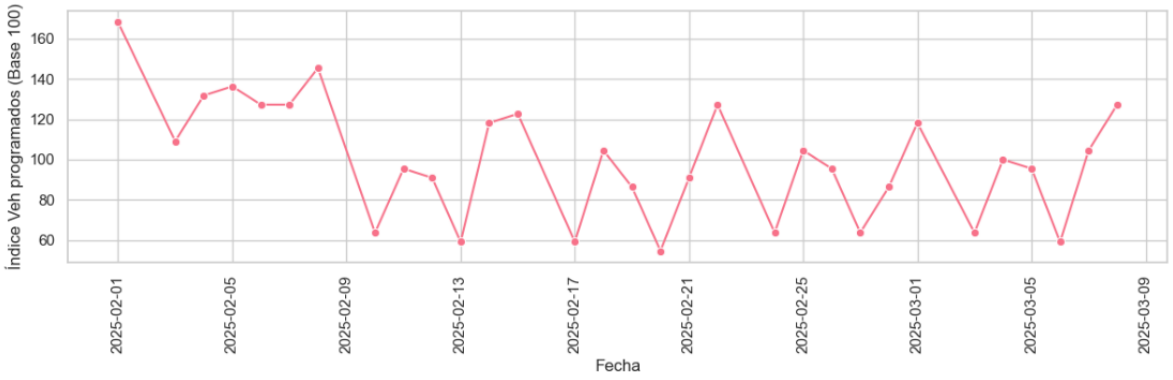
Evolución Relativa de Cajas Recolectadas a lo Largo del Tiempo



De aquí se puede inferir que el volumen de cajas suele ser estable en el tiempo. También se observa una concentración mayor de cajas en los primeros días de la muestra, lo cual hace referencia a un espacio de tiempo en el cual se manejan volúmenes más altos por la cercanía a la celebración de San Valentín en Estados Unidos.

Figura 2

Evolución Relativa de Vehículos Programados por Fecha



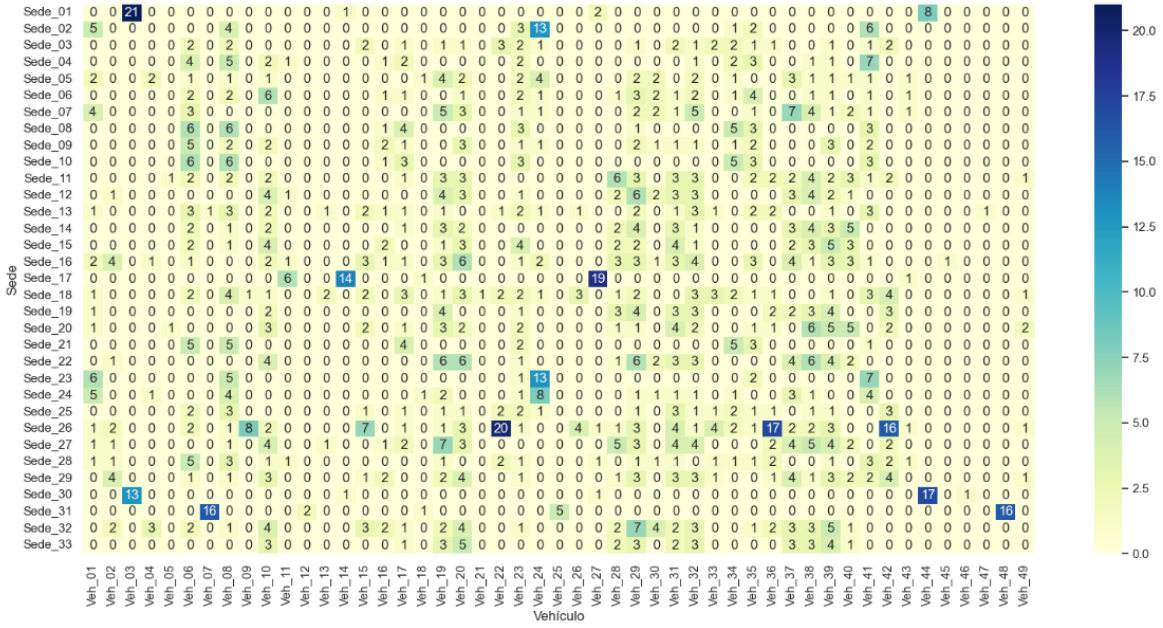
En la **Figura 2** se observa la evolución de la cantidad de vehículos programados a lo largo del periodo analizado. Se identifica una tendencia general con una media estable, aunque al

inicio del periodo se registra una mayor cantidad de vehículos programados esto en línea con la tendencia de las cajas vistas en la **Figura 1**.

Con base en la evolución relativa de vehículos programados, se puede observar que la capacidad operativa presenta fluctuaciones a lo largo del tiempo, reflejando posibles ajustes en la planificación según la demanda. Estas variaciones son consistentes con la dinámica observada en el volumen de cajas recolectadas, sugiriendo que la operación se adapta de forma flexible a las necesidades diarias.

Figura 3

Cantidad Vehiculos Programados por Fecha

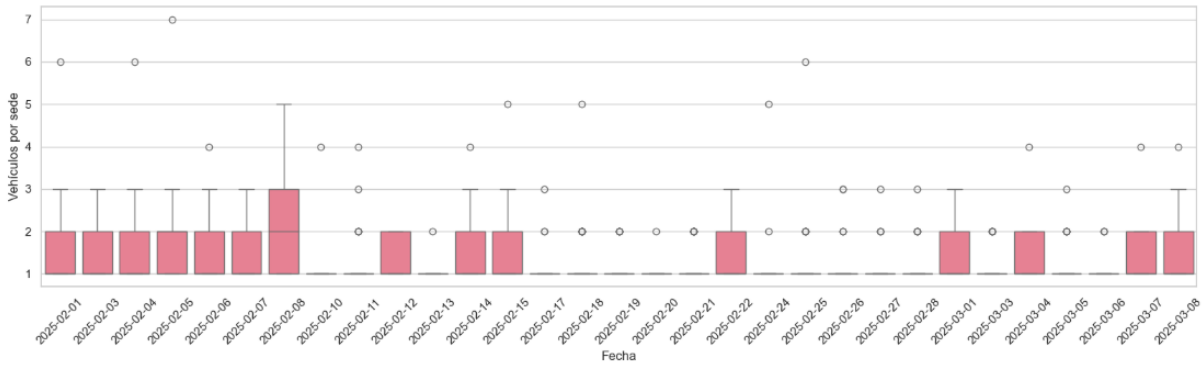


En la **Figura 3** se presenta un conteo del número de días en los que cada vehículo visitó una sede específica. Se observa una alta concentración de viajes en ciertos puntos, lo cual se explica, en varios casos, porque algunas sedes pertenecen a un grupo que solo puede ser atendido por un transportador específico, generando así una asociación directa.

Otro de los valores elevados se debe a que una de las sedes concentra el mayor volumen de despacho dentro de la organización, lo que provoca que los vehículos de mayor capacidad disponibles sean programados de forma recurrente para realizar recolecciones en dicha ubicación.

Figura 4

Distribución Diaria de Vehiculos Distintos por Finca

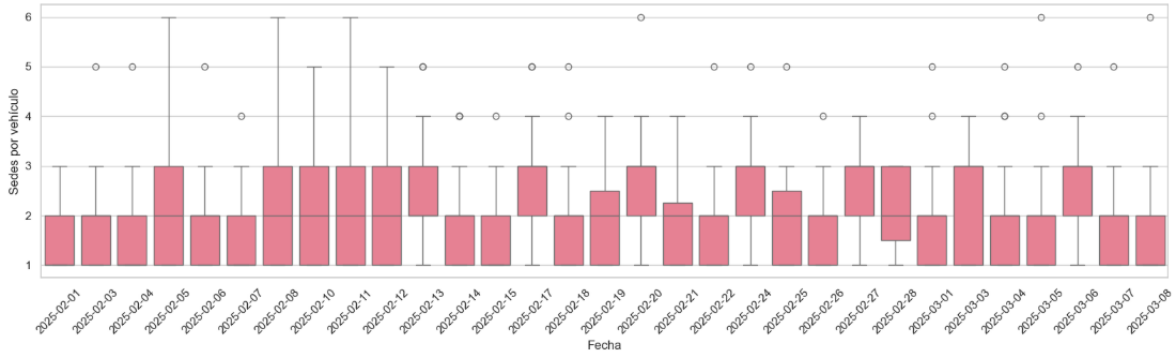


En la **Figura 4** se realizó una visualización tipo diagrama de caja y bigotes para visualizar que tantos vehículos visitan una misma sede en el mismo día de operación, esto es importante dado que uno de los indicadores de medición para la viabilidad de una programación de rutas es la cantidad de veces que una sede necesita atender un vehículo dado que esto influye en tiempos, seguridad y calidad de la flor.

En este caso se observa que la media de vehículos en una sede es de entre 1 y 2, sin embargo, en el gráfico se observa que hay datos aislados con 6 y 7 vehículos en una sede, estos hacen referencia a la sede mas grande en volumen dado que por el tamaño del despacho es imposible que se organice en pocos vehículos.

Figura 5

Distribución Diaria de Fincas Distintas por Vehiculo



Otro indicador clave en la programación de rutas es la cantidad de sedes que visita un vehículo, lo cual es diferente al punto anterior. Esto lo podemos ver en la **Figura 5**. Este punto es relevante en términos de que se busca que un mismo vehículo visite la menor cantidad de sedes con el fin de evitar trasladar la flor innecesariamente generando aperturas que impacten la cadena de frío lo que implica afectar la calidad de la flor. Como se ve en la imagen, el promedio se encuentra entre 2 y 3 sedes visitadas por vehículo, aquí también se presentan algunos datos aislados con 5 o 6 sedes visitadas, estos hacen referencia a que siempre en cada programación hay una cantidad de sedes con un volumen de despacho muy bajo o sobrante de otro vehículo para el cual no vale la pena dedicar un vehículo completo, entonces se organiza un recorrido que visite dichas sedes y recoja todo en un viaje largo. Normalmente solo es un vehículo es cual tiene esta cantidad de viajes.

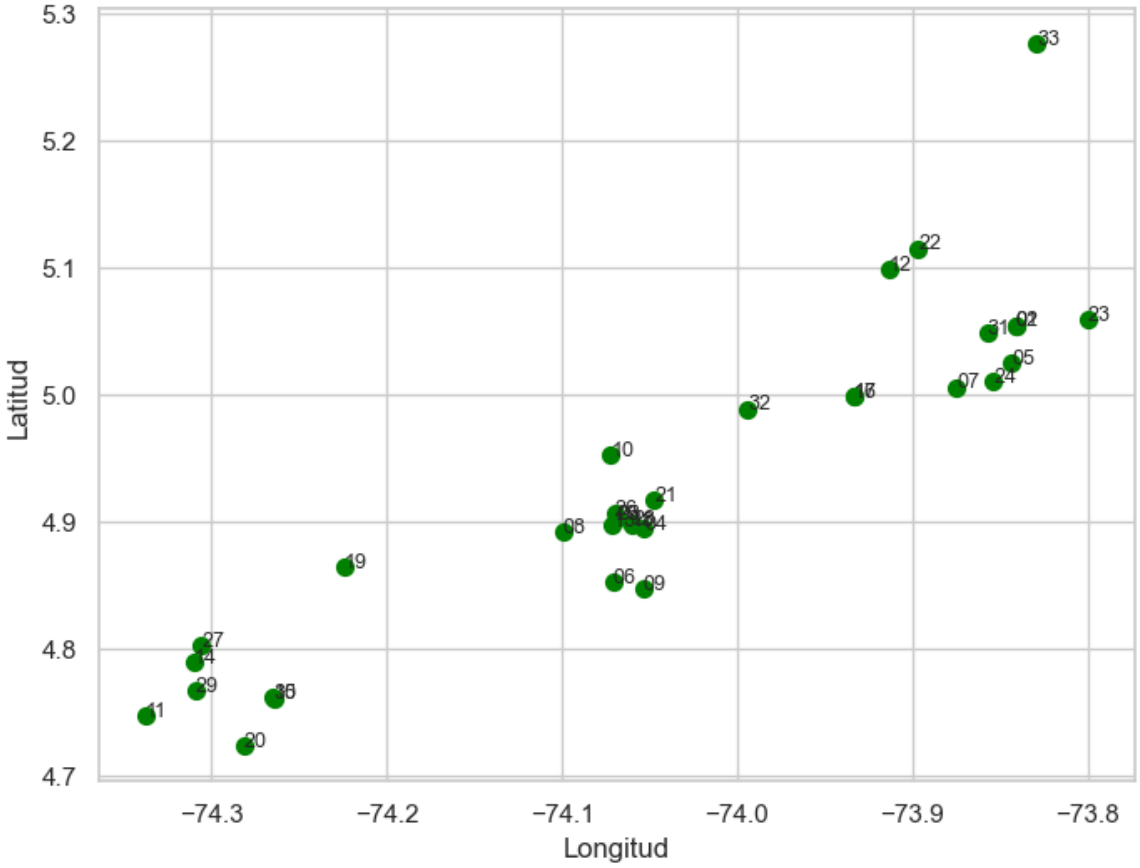
Análisis Espacial

En este caso se realizó una ilustración de la ubicación de las sedes que se puede ver en la **Figura 6**. Acá es relevante ver como parece existir una distribución en zonas de las sedes, en el contacto que se ha tenido con el equipo programador si se tiene en cuenta dicha zonificación para

programar sedes por cercanía a fin de disminuir los tiempos de ruta que se generan. También es importante recalcar que existe una sede que pareciera estar fuera de las agrupación, esta sede es la 33, entendiendo un poco mas el contexto en el que se realiza la programación de rutas para dicha sede existe una regla definida y es que siempre se envía un solo vehículo por el volumen total del despacho.

Figura 6

Ubicación de las Sedes



Modelo de Optimización de Rutas con PPO

En este capítulo presentamos la aplicación de técnicas de aprendizaje por refuerzo, específicamente el algoritmo Proximal Policy Optimization (PPO), al problema de optimización de rutas en logística de transporte de flores. El objetivo es diseñar un agente capaz de generar rutas eficientes que minimicen distancia recorrida y maximicen la ocupación de los vehículos.

Formulación del Problema como Entorno de RL

En esta sección detallamos cómo el problema de optimización de rutas se adapta a un entorno de aprendizaje por refuerzo.

Representación del Estado

Cada estado representa el contexto actual del sistema logístico. Se define como un vector de características que incluye:

- La ubicación actual de todos los vehículos.
- Las capacidades actuales de todos los vehículos.
- La ubicación de todos los pedidos.
- Las cantidades de todos los pedidos.
- Información sobre los pedidos asignados.

Este diseño permite que el agente tome decisiones basadas en una visión completa y actualizada de la operación.

Espacio de Acciones

El agente debe decidir el siguiente paso operativo. El espacio de acciones está compuesto por:

- Selección de un vehículo disponible
- Elección de recoger un pedido que no ha sido recolectado.

- Consideraciones de factibilidad: sólo se habilitan acciones que respeten la capacidad del vehículo.

Las acciones son representadas como índices sobre los nodos disponibles, permitiendo una interpretación discreta adecuada para PPO.

Función de Recompensa

La función de recompensa guía el aprendizaje del agente hacia rutas óptimas. Está diseñada con los siguientes componentes:

- Recompensa negativa proporcional a la distancia recorrida: incentiva trayectorias cortas.
- Bonificación por alta ocupación de vehículos: fomenta el uso eficiente de la capacidad.
- Penalización por pedidos no recolectados: finalizar una ejecución con pedidos sin asignar castiga la solución
- Penalización incremental por cantidad de vehículos en la misma sede: fomentar que los vehículos recojan la mayor cantidad de pedidos en la misma sede.

Esta estructura busca balancear objetivos operativos múltiples, alineándose con las prioridades del negocio de transporte de flores.

Arquitectura de la Solución

La solución fue desarrollada utilizando Python, estructurada en torno a un entorno personalizado de aprendizaje por refuerzo y entrenada utilizando el algoritmo Proximal Policy Optimization (PPO) de la librería stable-baselines3.

A continuación, se describe la arquitectura a nivel de módulos y librerías principales:

Entorno Personalizado en Gym

Se creó una clase OrderAssignmentEnv, que hereda de gym.Env, definiendo:

Espacios de observación: Un spaces.Dict que incluye:

- Capacidades disponibles de los vehículos.
- Capacidades utilizadas.
- Volúmenes de las órdenes.
- Estado de asignación de las órdenes.
- Espacio de acciones: Una acción discreta que representa asignar una orden a un vehículo (MAX_VEHICLES × MAX_ORDERS combinaciones posibles).

Lógica de recompensa

- Incentiva asignar pedidos, maximizar la ocupación del vehículo y minimizar distancia recorrida.
- Penaliza asignaciones inválidas, exceso de capacidad y visitar demasiadas fincas.

Adicionalmente, el entorno maneja:

- Padding de datos para manejar un tamaño fijo de vehículos y órdenes (MAX_VEHICLES=30, MAX_ORDERS=350).
- Visualización de métricas como distancias recorridas, uso de capacidad y distribución de órdenes por finca usando matplotlib.

Obtención de Datos Históricos

La función get_history_data_oneday extrae:

- Matriz de distancias.
- Detalles de órdenes
- Asignaciones de órdenes a sedes.

- Información de la flota de vehículos.
- Ubicación inicial de los vehículos.

Estos datos son utilizados para inicializar el entorno de entrenamiento.

Algoritmo de Aprendizaje

El entrenamiento se realiza utilizando, algoritmo PPO (stable-baselines3), configurado con:

- Política de entrada múltiple (MultiInputPolicy) adaptada a las observaciones definidas.
- make_vec_env para vectorizar el entorno.
- configure para establecer la carpeta de logs (stdout + tensorboard).

Entrenamiento, se entrenó el agente durante 50,000 pasos. Guardado, el modelo resultante se guarda para su posterior uso bajo el nombre order_assignment_ppoV2.

Bibliotecas Principales Utilizadas

- Gymnasium (gymnasium) para construir el entorno personalizado.
- Stable Baselines 3 (stable-baselines3) para implementar PPO.
- Numpy para manipulación eficiente de arreglos y matrices.
- Matplotlib para graficar métricas del entorno.

Experimentación y Resultados

En esta sección se presentan los resultados obtenidos durante la fase de experimentación del modelo de asignación de órdenes para la flota de vehículos. Se utilizaron 31 días de datos, de los cuales se evaluaron los resultados de 10 escenarios seleccionados. Cada escenario representó un día distinto en el que las condiciones de la flota y los pedidos podían variar. El proceso de experimentación se llevó a cabo de manera progresiva, evaluando continuamente el desempeño del modelo con cada escenario y ajustando parámetros según fuese necesario.

Configuración del Modelo

El modelo de asignación de órdenes fue entrenado utilizando el algoritmo Proximal Policy Optimization (PPO) de la librería Stable-Baselines3. Este modelo es un algoritmo de refuerzo que permite optimizar políticas en entornos complejos y dinámicos, como el de la asignación de órdenes a vehículos.

Parámetros de Entrenamiento

Los parámetros utilizados para entrenar el modelo fueron los siguientes:

- Escenarios Evaluados: 10 días seleccionados de un total de 31.
- Pasos de Entrenamiento por Escenario: 50,000 pasos.
- Tiempo de Entrenamiento por Escenario: Aproximadamente 5 minutos.
- Tiempo de Inferencia por Escenario: Aproximadamente 30 segundos.

El entrenamiento fue realizado de forma progresiva, es decir, el modelo fue entrenado con los datos de cada día de forma secuencial, incorporando la experiencia de los escenarios previos para optimizar la asignación de pedidos en el siguiente día.

Desempeño del Modelo

Evaluación de Métricas de Desempeño

Las métricas clave utilizadas para evaluar el desempeño del modelo fueron distancia recorrida y porcentaje de capacidad utilizada de los vehículos. Estas métricas son fundamentales para entender la eficiencia del modelo en cuanto a la minimización de costos operativos y la optimización de los recursos disponibles (vehículos).

Figura 7

Métricas de Desempeño

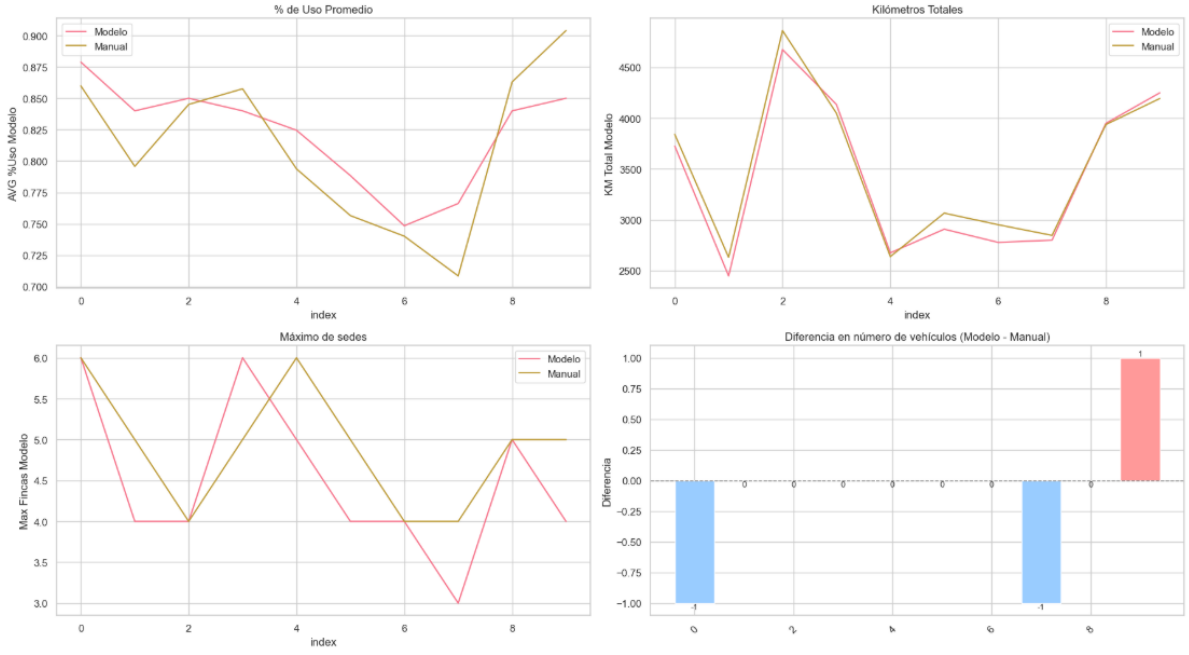


Tabla 1

Comparativo Solución Modelo PPO y Manual

	Modelo PPO				Modelo PPO			
Fecha	Vehiculos	AVG %Uso	KM Total	Max Fincas	Vehiculos	AVG %Uso	KM Total	Max Fincas

		Modelo PPO				Modelo PPO		
20250211	-1	88%	3590	5	-	86%	3841	6
20250213	-	82%	2284	5	-	80%	2631	5
20250215	-	82%	5585	3	-	85%	4862	4
20250218	-	82%	3204	4	-	86%	4049	5
20250220	-	82%	1732	7	-	79%	2638	6
20250224	-	79%	1813	6	-	76%	3066	5
20250227	-	75%	1731	4	-	74%	2952	4
20250303	-1	77%	1667	4	-	71%	2847	4
20250304	-	79%	3536	4	-	86%	3938	5
20250307	+1	82%	4394	5	-	90%	4194	5

- Distancia Recorrida: El modelo mostró un desempeño en términos de distancia recorrida comparable al enfoque de programación manual utilizado previamente. La distancia recorrida es crucial, ya que se busca minimizar los costos asociados al transporte de las órdenes desde las fincas hasta los clientes.

- Porcentaje de Uso de Vehículos: El porcentaje de uso de la capacidad de los vehículos también mostró una consistencia con los resultados de la programación manual. Esta métrica es importante para maximizar la eficiencia del uso de los vehículos, evitando el sub-utilizado o sobrecargado de los mismos.

- Número de vehículos usados: El modelo PPO logró en dos escenarios utilizar un vehículo menos respecto a los utilizados en la solución manual, solamente en un escenario tuvo que usar un vehículo de mas, esto es un resultado exitoso dado que en la gran mayoría de escenarios logra usar la misma o menor flota.

En resumen, el modelo logró desempeñar sus tareas con eficacia en cuanto a las dos métricas principales de evaluación, con un rendimiento similar al enfoque manual.

Indice de Eficiencia

El índice de eficiencia se calcula utilizando la siguiente fórmula para cada métrica, comparando el modelo Manual con el modelo PPO. Se puede calcular para las métricas de KM Total y AVG \%Uso de la siguiente manera:

Índice de Mejora en KM Total

$$IndiceMejoraKMTotal = \frac{PromedioKMManual - PromedioKMPPO}{PromedioKMManual} x100$$

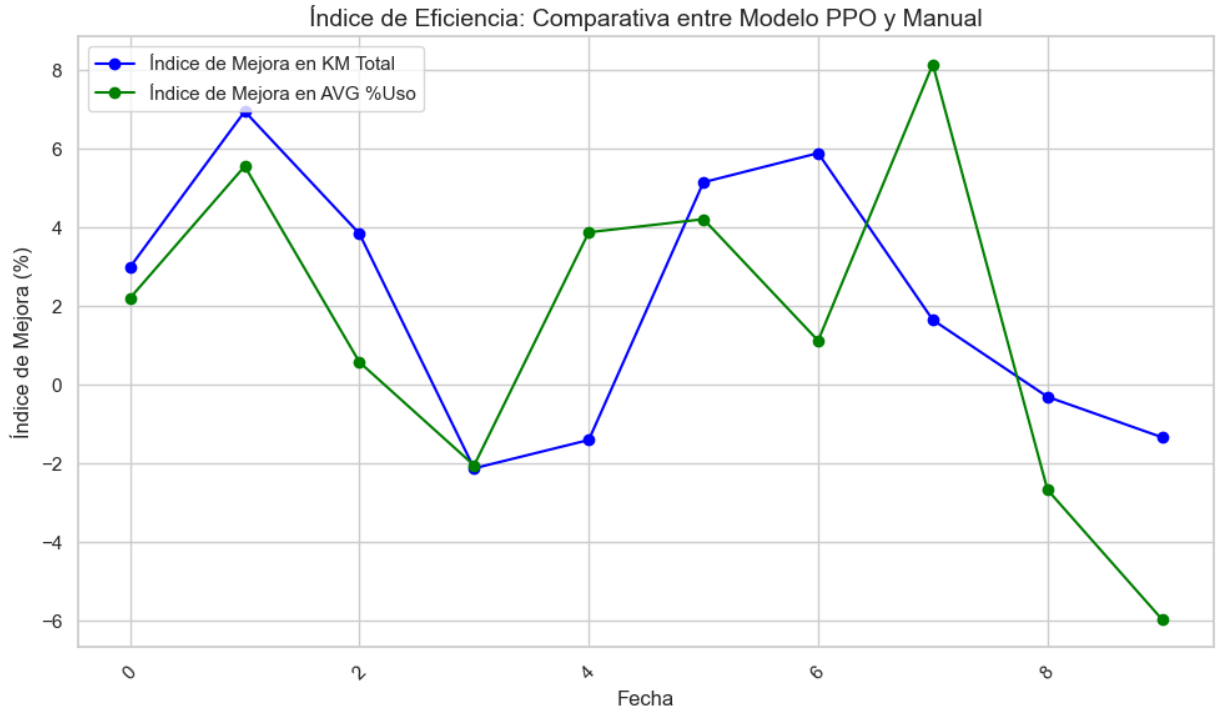
Índice de Mejora en AVG %Uso

$$IndiceMejoraAVG\%Uso = \frac{Promedio\%UsoPPO - Promedio\%UsoManual}{Promedio\%UsoManual} x100$$

Este índice proporciona el porcentaje de mejora en cada una de las métricas, donde un valor positivo indica que el modelo PPO es más eficiente que el modelo manual.

Figura 8

Indice Eficiencia



La **Figura 8** presenta las mejoras porcentuales obtenidas según las métricas definidas. Se observa que en seis escenarios el modelo PPO muestra un desempeño superior, evidenciado por incrementos tanto en KMTotal como en el promedio de uso (KMTotal > 0 y AVG Uso > 0). En un escenario, el rendimiento es comparable al de la referencia, dado que KMTotal disminuye mientras que el promedio de uso mejora (KMTotal < 0 y AVG Uso > 0). Finalmente, en tres escenarios, el modelo no logra superar la solución base, presentando disminuciones en ambas métricas (KMTotal < 0 y AVG Uso < 0).

Inferencia Rápida

Una de las grandes ventajas del modelo es su capacidad de inferencia rápida. La inferencia en el modelo duró aproximadamente 30 segundos por escenario, lo cual es una mejora significativa en comparación con el proceso manual actual, que es mucho más lento (alrededor de 2 horas) y propenso a errores. Esta rapidez en la toma de decisiones es especialmente valiosa en entornos dinámicos, donde los tiempos de respuesta son cruciales para una operación eficiente.

Ajuste y Mejoras

Durante la experimentación, se identificaron ciertos ajustes necesarios para garantizar que el modelo pudiera adaptarse a las variaciones en los datos de un día a otro. En particular, el modelo tuvo que ser ajustado para manejar el cambio en la cantidad de vehículos y pedidos de un día a otro sin comprometer su estabilidad.

Ajuste de Datos de Entrada: A medida que se cambiaban los días, la cantidad de vehículos y pedidos variaba significativamente. Esto presentó un desafío, ya que el modelo no podía funcionar correctamente si los datos de entrada no estaban alineados con las expectativas

del modelo, dado que se entrenó bajo un conjunto de datos con un número fijo de vehículos y órdenes.

Para abordar este problema, se implementaron ajustes en el modelo que permitieron manejar de manera adecuada las variaciones en los datos. Estos ajustes aseguraron que el modelo pudiera adaptarse dinámicamente a las nuevas condiciones de cada día sin que se produjeran errores de ejecución o pérdidas de rendimiento.

Inicio de Pruebas con el Equipo Operativo

Una vez finalizada la fase de experimentación y evaluación del modelo, se dio inicio a las pruebas en tiempo real con el equipo operativo. Aunque no se llevó a cabo una implementación completa del modelo en un entorno de producción, se realizaron pruebas preliminares en condiciones controladas con un conjunto reducido de datos y vehículos.

Estas pruebas tuvieron como objetivo observar cómo el modelo PPO se comportaba bajo las condiciones reales de operación, identificando posibles desafíos en la implementación y ajustando los parámetros para asegurar una transición exitosa. Durante este proceso, se identificaron varias métricas adicionales, como la interacción con el personal operativo y la capacidad de los sistemas de gestión para integrar las predicciones del modelo en tiempo real. Esto es importante dado que al momento representar una solución entregada por el modelo puede llevar hasta 30 minutos reflejarla en los sistemas de información.

A lo largo de las pruebas, se recopiló retroalimentación del equipo operativo sobre la usabilidad y la efectividad de las recomendaciones generadas por el modelo. Principalmente por la necesidad de fijar algunas rutas que simpre se hacen de la misma forma, esto restringe el espacio de trabajo para la optimización pero es necesario para acoplarse a la operación.

Aunque no se obtuvo una implementación total, los resultados obtenidos en estas pruebas iniciales indican que el modelo tiene un gran potencial para ser implementado en producción, mejorando significativamente la asignación de órdenes y optimizando el uso de la flota de vehículos.

Conclusiones

El modelo de asignación de órdenes utilizando el algoritmo Proximal Policy Optimization (PPO) ha demostrado ser una herramienta eficaz para optimizar la asignación de pedidos a vehículos en un entorno dinámico. Durante la fase de experimentación y pruebas, se observaron varias ventajas clave del modelo en comparación con los métodos manuales:

Eficiencia en la Asignación de Órdenes: El modelo PPO ha mostrado ser capaz de asignar órdenes de manera eficiente, alcanzando un desempeño comparable o superior al enfoque manual. Con 30 vehículos y 350 órdenes como entorno fijo, el modelo pudo gestionar la asignación de manera precisa y rápida, optimizando el uso de la flota y minimizando la distancia recorrida, uno de los objetivos clave del proceso.

Rapidez en la Inferencia: Uno de los aspectos más destacables de este modelo es la rapidez en la inferencia, con un tiempo promedio de 30 segundos por escenario. Esta mejora significativa frente al proceso manual, que toma aproximadamente 2 horas, es un cambio fundamental en la operativa diaria, permitiendo una toma de decisiones mucho más ágil y precisa. Esta reducción de tiempo tiene un impacto directo en la productividad y eficiencia operativa.

Evaluación de Indicadores: Las métricas de desempeño evaluadas, como el porcentaje de uso de los vehículos y la distancia recorrida, han mostrado una mejora consistente, destacándose en varios escenarios del análisis comparativo. Los índices de mejora obtenidos reflejan una optimización de los recursos disponibles y una reducción de costos asociados a la operación, lo que confirma la efectividad del modelo PPO en un entorno real.

Adaptabilidad del Modelo: Durante la experimentación, el modelo mostró una gran capacidad de adaptación a los cambios en las condiciones diarias, como las fluctuaciones en el

número de vehículos disponibles y las órdenes asignadas. Esta capacidad para ajustarse rápidamente a nuevas situaciones es crucial en un entorno operativo como el de asignación de órdenes a vehículos, donde la dinámica puede cambiar en cualquier momento.

Potencial para Integración en Producción: Si bien las pruebas con el equipo operativo aún están en curso, los resultados obtenidos en esta fase son altamente prometedores. La integración de este modelo en el entorno de producción podría llevarse a cabo de manera exitosa con mejoras incrementales en la interfaz de usuario y ajustes en los flujos de trabajo actuales. La automatización de la asignación de órdenes mediante el modelo PPO podría reducir significativamente el esfuerzo manual, liberando tiempo para tareas de mayor valor.

Recomendaciones y Trabajos Futuros

Integración con los Sistemas Transaccionales: Para optimizar aún más el proceso y evitar la digitación manual en las herramientas actuales, se recomienda integrar el modelo de asignación de órdenes con los sistemas transaccionales de la organización. Esta integración permitiría una carga automática de los datos, eliminando la necesidad de plantillas de carga manuales y reduciendo la probabilidad de errores humanos. Además, se facilitaría una mayor rapidez en la actualización de los datos y una sincronización más eficiente con otros sistemas de la empresa.

Mejora en el Modelo con Rutas Prefijadas: Una mejora adicional en el modelo sería la incorporación de rutas prefijadas para los vehículos. Aunque el modelo PPO optimiza la asignación de órdenes de manera eficiente, el uso de rutas predefinidas podría aportar una mayor precisión en la planificación de las rutas, especialmente en zonas con características geográficas específicas o donde existen restricciones de acceso. Esta mejora podría contribuir a una reducción aún mayor en la distancia recorrida y en el tiempo de entrega de las órdenes.

Expansión a Otras Operaciones en Otras Zonas del País: Una vez validado el modelo en el entorno actual, se recomienda explorar su expansión a otras operaciones de la empresa en diferentes zonas del país. La adaptación del modelo a nuevas áreas geográficas permitiría optimizar la asignación de vehículos en una escala más amplia, aprovechando la flexibilidad del modelo PPO para ajustarse a diversas condiciones operativas. La expansión a otras zonas también podría implicar la adaptación de los parámetros del modelo a nuevas realidades, como el tipo de vehículos disponibles o las características de las rutas.

Restricción de Vehículos en sedes específicas: Otra mejora en el modelo sería la implementación de restricciones de vehículos que no pueden ingresar a ciertas sedes debido a

características de acceso, tamaño o capacidad. La inclusión de estas restricciones en el modelo permitiría una asignación de órdenes aún más precisa y ajustada a las necesidades reales de las operaciones. Este tipo de restricciones podría integrarse como una parte clave del proceso de optimización, mejorando la eficiencia en el uso de la flota y garantizando el cumplimiento de las condiciones operativas.

Con estas recomendaciones y la visión para el trabajo futuro, el modelo de asignación de órdenes tiene un gran potencial para ser una solución eficiente y escalable en el ámbito de la gestión de flotas y distribución de productos.

Referencias

Al-Khazraji, H. (2022). Comparative Study of Whale Optimization Algorithm and Flower Pollination Algorithm to Solve Workers Assignment Problem. *International Journal of Production Management and Engineering, 10*. doi:10.4995/ijpme.2022.16736

Ara, S., Akib, M. M., Oion, M. S., Shohel, M. H., Ridoy, M. N., Kabita, F. A., & Shahiduzzaman, M. (2023). Vehicle Routing Problem Solving Using Reinforcement Learning. *2023 26th International Conference on Computer and Information Technology (ICCIT).* doi:10.1109/ICCIT60459.2023.10441644

Bai, R., Chen, X., Chen, Z.-L., Cui, T., Gong, S., He, W., . . . Zhang, H. (2023). Analytics and machine learning in vehicle routing research. *International Journal of Production Research, 61*, 4–30. doi:10.1080/00207543.2021.2013566

Ballou, R. H. (2004). *Business Logistics/Supply Chain Management* (5th ed.). Upper Saddle River, NJ: Pearson Education.

Bhandari, U. (2022). Solving Vehicle Routing Problem using Machine Learning based clustering and TSP Cluster Redistribution. *International Journal of Research and Review, 9*. doi:10.52403/ijrr.20221041

Bogyrbayeva, A., Meraliyev, M., Mustakhov, T., & Dauletbayev, B. (2024). Machine Learning to Solve Vehicle Routing Problems: A Survey. *IEEE Transactions on Intelligent Transportation Systems, 25*. doi:10.1109/TITS.2023.3334976

Boumpa, E., Tsoukas, V., Chioktour, V., Kalafati, M., Spathoulas, G., Kakarountas, A., . . . Malindretos, G. (2022). A Review of the Vehicle Routing Problem and the Current Routing Services in Smart Cities. *Analytics, 2*. doi:10.3390/analytics2010001

Capgemini, University, P. S., & eyefortransport. (2017). The 2017 Third-Party Logistics Study: The State of Logistics Outsourcing. *The 2017 Third-Party Logistics Study: The State of Logistics Outsourcing*. Retrieved from https://3plstudy.com

Czuba, P., & Pierzchała, D. (2021, February). Machine Learning methods for solving Vehicle Routing Problems. *Sustainable Economic Development and Advancing Education Excellence in the Era of Global Pandemic.*

Hugging Face. (2023). Deep Reinforcement Learning: Proximal Policy Optimization (PPO). *Deep Reinforcement Learning: Proximal Policy Optimization (PPO)*.

Kasarda, J. (2016). Logistics Is about Competitiveness and More. *Logistics, 1*, 1. doi:10.3390/logistics1010001

Kostrikov, I., Yarats, D., Fergus, R., Ma, T., & Nachum, O. (2020). Image as a Plan: Model-based Hierarchical Reinforcement Learning for Visual Navigation. *arXiv preprint arXiv:2006.13205*. Retrieved from https://arxiv.org/abs/2006.13205

Kou, S., Golden, B., & Bertazzi, L. (2024). An improved model for estimating optimal VRP solution values. *Optimization Letters, 18*. doi:10.1007/s11590-023-02082-w

Margineanu, G. A. (2024). Big Data–Análisis de tráfico y optimización de rutas con machine learning. *Big Data–Análisis de tráfico y optimización de rutas con machine learning*.

Mazar, M., Belgherri, H., Saouabe, A., & Salihou, F. (2023). Integrating Artificial Intelligence and Optimization Techniques for Efficient Delivery. *2023 International Conference on Decision Aid Sciences and Applications (DASA).* doi:10.1109/DASA59624.2023.10286743

Mishra, S., & Tyagi, A. K. (2022). Emerging Trends and Techniques in Machine Learning and Internet of Things-Based Cloud Applications. In *Handbook of Research of Internet of Things and Cyber-Physical Systems.* doi:10.1201/9781003277323-7

OpenAI. (2018). OpenAI Baselines: PPO. *OpenAI Baselines: PPO*.

Ou, S., Ismail, Z. H., & Sariff, N. (2024). Hybrid Genetic Algorithms for Order Assignment and Batching in Picking System: A Systematic Literature Review. *IEEE Access, 12*. doi:10.1109/ACCESS.2024.3357689

Oyola, J., Arntzen, H., & Woodruff, D. L. (2016). The stochastic vehicle routing problem, a literature review. *EURO Journal on Transportation and Logistics*.

Pugliese, L. D., Ferone, D., Festa, P., Guerriero, F., & Macrina, G. (2023). Combining variable neighborhood search and machine learning to solve the vehicle routing problem with crowd-shipping. *Optimization Letters, 17*. doi:10.1007/s11590-021-01833-x

Revanna, J. K., & Al-Nakash, N. Y. (2023). Tensor Flow Model with Hybrid Optimization Algorithm for Solving Vehicle Routing Problem. In *Lecture Notes in Networks and Systems* (Vol. 672). doi:10.1007/978-981-99-1624-5_8

Rodríguez, R., & Jaca, C. (2014). *Gestión logística y cadena de suministro.* Madrid, España: ESIC Editorial.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*. Retrieved from https://arxiv.org/abs/1707.06347

Shahbazian, R., Pugliese, L. D., Guerriero, F., & Macrina, G. (2024). Integrating Machine Learning Into Vehicle Routing Problem: Methods and Applications. *IEEE Access, 12*, 93087–93115. doi:10.1109/ACCESS.2024.3422479

Syed, A. A., Akhnoukh, K., Kaltenhaeuser, B., & Bogenberger, K. (2019). Neural network based large neighborhood search algorithm for ride hailing services. In *Lecture Notes in Computer Science* (Vol. 11804). doi:10.1007/978-3-030-30241-2_49

Vamsi, V. S., Telukuntla, Y. R., Kumar, P. S., & Gutjahr, G. (2023). Comparison of Attention Mechanisms in Machine Learning Models for Vehicle Routing Problems. In *Lecture Notes in Electrical Engineering* (Vol. 998). doi:10.1007/978-981-99-0047-3_53

Vargas Ramírez, E. C., Barrera Bello, N. A., & Cruz Martín, J. S. (2019). Perspectivas de mejora en la cadena logística exportadora de las rosas en la Sabana de Bogotá para el periodo 2015–2018. *Perspectivas de mejora en la cadena logística exportadora de las rosas en la Sabana de Bogotá para el periodo 2015–2018*. Retrieved from https://repository.ucc.edu.co/entities/publication/231c75e1-b201-49e4-8bd7-127cfdf0b7a3

Wang, Q., & Hao, Y. (2023). Routing optimization with Monte Carlo Tree Search-based multi-agent reinforcement learning. *Applied Intelligence, 53*. doi:10.1007/s10489-023-04881-1

Zhang, J. (2021). An Improved Genetic Algorithm for Vehicle Routing Problem. In *Advances in Intelligent Systems and Computing* (Vol. 1282, pp. 163–169). doi:10.1007/978-3-030-62743-0_23

Zhou, J., Wu, Y., Song, W., Cao, Z., & Zhang, J. (2023). Towards Omni-generalizable Neural Methods for Vehicle Routing Problems. *Proceedings of Machine Learning Research, 202*.