IMPLEMENTACIÓN DE UN MÓDULO INTELIGENTE PARA LA DETECCIÓN DE ROBO DE COMBUSTIBLE EN VEHÍCULOS DE CARGA INTEGRADO A LA PLATAFORMA WEB FUELCONTROL

AUTOR
MARTIN LEONARDO LOZADA CORTÉS

PROPUESTA DE GRADO PARA OPTAR AL TÍTULO DE INGENIERO TELEMÁTICO

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

FACULTAD TECNOLÓGICA
INGENIERÍA TELEMÁTICA

Bogotá, D.C., 2025

IMPLEMENTACIÓN DE UN MÓDULO INTELIGENTE PARA LA DETECCIÓN DE ROBO DE COMBUSTIBLE EN VEHÍCULOS DE CARGA INTEGRADO A LA PLATAFORMA WEB FUELCONTROL

AUTOR
MARTIN LEONARDO LOZADA CORTÉS
Modelo de grado: Pasantía

PROPUESTA DE GRADO PARA OPTAR AL TÍTULO DE INGENIERO TELEMÁTICO

DIRECTOR
JORGE EDUARDO HERNÁNDEZ RODRÍGUEZ



UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

FACULTAD TECNOLÓGICA
INGENIERÍA TELEMÁTICA

Bogotá, D.C., 2025

LISTA DE ABREVIATURAS

ACK	Acknowledgement (Acuse de recibo)
ACP	Asociación Colombiana del Petróleo
ACPM	Aceite Combustible Para Motor (Diésel)
API	Application Programming Interface (Interfaz de Programación de Aplicaciones)
AUC	Area Under the Curve (Área Bajo la Curva)
COP	Peso Colombiano (Moneda)
CRISP-DM	Cross-Industry Standard Process for Data Mining
CRM	Customer Relationship Management (Gestión de Relaciones con Clientes)
CSV	Comma-Separated Values (Valores Separados por Comas)
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
ERP	Enterprise Resource Planning (Planificación de Recursos Empresariales)
FPR	False Positive Rate (Tasa de Falsos Positivos)
FN	False Negative (Falso Negativo)
FP	False Positive (Falso Positivo)
GPRS	General Packet Radio Service (Servicio General de Paquetes vía Radio)
GPS	Global Positioning System (Sistema de Posicionamiento Global)
GSM	Global System for Mobile Communications (Sistema Global para Comunicaciones Móviles)

IA	Inteligencia Artificial
IoT	Internet of Things (Internet de las Cosas)
IoU	Intersection over Union (Intersección sobre Unión)
JSON	JavaScript Object Notation
km/h	Kilómetros por hora
LGB	LightGBM (Light Gradient Boosting Machine)
LGBM	LightGBM Classifier
LR	Logistic Regression (Regresión Logística)
MAD	Median Absolute Deviation (Desviación Absoluta Mediana)
MB	Megabyte
ML	Machine Learning (Aprendizaje Automático)
MPI	Materiales y Proyectos Industriales
NaN	Not a Number (No es un número)
NMS	Non-Maximum Suppression (Supresión de No-Máximos)
OBD	On-Board Diagnostics (Diagnóstico a Bordo)
P95	Percentil 95
PIB	Producto Interno Bruto
PR	Precision-Recall (Precisión-Exhaustividad)
PR-AUC	Precision-Recall Area Under Curve
RAM	Random Access Memory (Memoria de Acceso Aleatorio)
REST	Representational State Transfer

RF	Random Forest (Bosque Aleatorio)
ROC	Receiver Operating Characteristic (Característica Operativa del Receptor)
RPM	Revolutions Per Minute (Revoluciones Por Minuto)
SaaS	Software as a Service (Software como Servicio)
SMS	Short Message Service (Servicio de Mensajes Cortos)
TN	True Negative (Verdadero Negativo)
TP	True Positive (Verdadero Positivo)
UTC	Coordinated Universal Time (Tiempo Universal Coordinado)
WGS84	World Geodetic System 1984 (Sistema Geodésico Mundial 1984)
XGB	XGBoost (Extreme Gradient Boosting)
XLSX	Microsoft Excel Open XML Spreadsheet

TABLA DE CONTENIDO

1. OBJETIVOS **7**
1.1 Objetivo General 7
1.2 Objetivos Específicos 7
2. PREGUNTA DE INVESTIGACIÓN **8**
3. INTRODUCCIÓN **9**
4. DEFINICIÓN DEL PROBLEMA **11**
5. PLANTEAMIENTO DEL PROBLEMA **13**
6. MARCO TEÓRICO **15**
6.1. Telemetría vehicular y monitoreo remoto 15
6.2. Sensores de nivel de combustible y fuentes de ruido 15
6.3. Posicionamiento GNSS y detección de estancias 16
6.4. Preprocesamiento robusto de series temporales 16
6.5. Detección y consolidación de eventos 16
6.6. Modelos de anomalía aplicados a robo de combustible 16
6.7 Robo de combustible 17
6.8. Evaluación en contextos desbalanceados 18
6.9. Calibración de puntajes y umbrales operativos 18
6.10. GPS Control e Infraestructura Tecnológica 18
7. METODOLOGÍA **20**
7.1. Enfoque Metodológico Adoptado 20
7.2. CRISP-DM: fases y aplicación en el proyecto 21
8. RESULTADOS **41**
8.1 Contexto Empresarial y Operativo 41
8.2 Recopilación y Caracterización del Dataset 44
8.3. Análisis Exploratorio y Hallazgos Preliminares 48
8.4 Preparación Avanzada de Datos y Generación de Eventos 55
8.5 Partición Temporal y Prevención de Data Leakage 70
8.6 Implementación de Modelos de Machine Learning 74
8.7 Evaluación Integral del Sistema 86
8.8 Integración con Plataforma FuelControl 96
9. CONCLUSIONES **100**
9.2 Respuesta a Preguntas de Investigación 101
9.3 Limitaciones y Trabajo Futuro 102
9.4 Recomendaciones para Investigación Futura 103
9.6 Reflexión Final 104

1. OBJETIVOS

1.1 Objetivo General

Implementar un algoritmo de detección de anomalías para la identificación del robo de combustible en vehículos de carga que debe ser integrado a la plataforma web FuelControl mejorando la gestión operativa de flotas vehiculares.

1.2 Objetivos Específicos

1. Analizar los datos operativos de las flotas vehiculares, incluyendo variables como velocidad, altitud, consumo promedio y registros históricos, para identificar patrones relevantes y diferenciar fluctuaciones normales de pérdidas anómalas.
2. Entrenar un modelo de aprendizaje automático para la detección del robo de combustible, utilizando técnicas adaptadas al análisis de datos en series temporales y condiciones operativas de flotas.
3. Incorporar un módulo de detección de anomalías a la plataforma web FuelControl mediante una API de comunicación en tiempo real con Gurtam, proveedor de información de geolocalización vehicular, generando automáticamente reportes con información de llenado del tanque, índices de consumo y registros de robos de combustible.
4. Validar el modelo de aprendizaje automático entrenado, evaluando su desempeño en la detección del robo de combustible mediante pruebas específicas en series temporales y condiciones operativas de flotas.
5. Realizar pruebas funcionales del sistema en condiciones operativas reales, comparando su precisión y la reducción de falsos positivos frente al algoritmo actual de FuelControl.

2. PREGUNTA DE INVESTIGACIÓN

¿Cuál es el nivel de precisión de un algoritmo avanzado de detección de anomalías para identificar robos de combustible, en comparación con el sistema actual de la plataforma FuelControl?

¿Qué características tienen las anomalías detectadas por el sistema en los datos históricos, y como pueden estas ser utilizadas para optimizar la prevención del robo de combustible?

3. INTRODUCCIÓN

La industria del transporte y la logística es un pilar de la economía colombiana al habilitar el movimiento de bienes y personas. En el país, el modo carretero concentra la mayor parte de la carga doméstica —alrededor de cuatro quintas partes, según estimaciones técnicas recientes—, lo que subraya su centralidad en la matriz logística nacional. (Universidad de los Andes – Giro Zero, 2021) A la vez, la rama de Transporte y almacenamiento aporta cerca del 5 % del PIB (4,9 % en el segundo trimestre de 2025), lo que reafirma su peso en la actividad productiva (DNP, 2025). En línea con esa escala, el Ministerio de Transporte reportó que en 2024 la movilización por carretera superó los 135 millones de toneladas y mostró variaciones positivas interanuales, reflejando la magnitud operativa del sector (MinTransporte, 2022).

Dentro del ámbito logístico, se reconoce como vehículos de carga pesada a camiones rígidos, tractocamiones, semirremolques, volquetas y cabezas tractoras, diseñados para el transporte de mercancías a gran escala sobre rutas interurbanas. En lo que respecta al abastecimiento energético del transporte, los combustibles líquidos derivados del petróleo —especialmente el diésel (ACPM) y la gasolina— siguen siendo predominantes en Colombia. El Informe Económico de Combustibles Líquidos de la Asociación Colombiana del Petróleo (ACP) señala que la demanda de diésel creció un 5.2% en 2024, mientras que el consumo de gasolina (corriente y extra) se contrajo un 5.9% (ACP, 2025); esto evidencia que el diésel sigue siendo el combustible más relevante para el funcionamiento de las flotas pesadas. En paralelo, el sector de combustibles contribuye con alrededor del 1.5% del PIB colombiano y genera decenas de miles de empleos directos (El Colombiano, 2025).

Sin embargo, el sector enfrenta desafíos significativos relacionados con la seguridad y la optimización de recursos, siendo el robo de combustible uno de los problemas más apremiantes, especialmente en vehículos de carga pesada. Este delito no solo genera pérdidas económicas directas estimadas en miles de millones anuales, sino que también impacta la eficiencia operativa, aumenta los costos logísticos y contribuye a la contaminación ambiental.

Aunque existen plataformas de monitoreo vehicular en tiempo real que utilizan tecnologías como GPS, muchas carecen de la sofisticación necesaria para detectar métodos de robo sutiles, como el "hurto por goteo", o para diferenciar las fluctuaciones normales del nivel de combustible (debidas a la inclinación, terreno, vibraciones) de las caídas anómalas que indican un robo. Esta falta de precisión genera falsas alarmas y consume tiempo valioso de los equipos de logística.

Este panorama contextual reafirma la importancia estratégica de diseñar mecanismos más precisos para monitorear el consumo de combustible en flotas pesadas. El robo de combustible —el "hurto por goteo" o caídas anómalas de nivel en ruta— no solo representa pérdidas directas en términos monetarios, sino que incide en la eficiencia operativa, incrementa los costos logísticos y potencia riesgos ambientales.

Ante esta problemática, el análisis y monitoreo de datos, aprovechando tecnologías como el Internet de las Cosas (IoT), Machine Learning (ML) e Inteligencia Artificial (IA), se presentan como herramientas cruciales para la toma de decisiones y la mejora de la gestión de flotas. Este proyecto se enfoca en desarrollar e implementar un sistema que, mediante el análisis de datos históricos y patrones operativos (como velocidad, ubicación GPS, estado de ignición y lecturas de sensores de nivel), permite detectar anomalías indicativas de posibles robos de combustible con mayor precisión. El sistema propuesto busca optimizar el monitoreo, reducir los falsos positivos, anticiparse al problema y generar alertas accionables, contribuyendo así a la seguridad, la reducción de costos y la eficiencia operativa en el sector transporte.

La integración de datos de múltiples fuentes (sensores GPS, nivel de combustible, flujo, vibración, OBD) y el uso de algoritmos avanzados de detección de anomalías son clave para superar las limitaciones de los sistemas actuales. La solución propuesta se integrará con la plataforma FuelControl y proporcionará informes periódicos (semanales) para la gestión, marcando un avance significativo en la prevención del robo de combustible y la optimización de recursos en la logística colombiana.

4. DEFINICIÓN DEL PROBLEMA

En el sector logístico colombiano, el robo de combustible constituye una amenaza multidimensional que va más allá de las pérdidas económicas para afectar la estabilidad operativa y la seguridad del transporte nacional. Según estimaciones del sector gremial de transporte de carga, este ilícito provoca pérdidas significativas anuales, afectando tanto a pequeñas empresas como a grandes corporaciones con flotas de más de cien vehículos (Colfecar, 2024). La problemática se agrava en regiones de intensa movilidad de mercancías, como los corredores Bogotá–Medellín y la Costa Caribe, donde la ausencia de sistemas confiables de detección ha impulsado a las empresas a implementar medidas reactivas (custodias armadas, rutas alternas), lo que puede incrementar los costos logísticos en hasta un 18 % (ITK Soluciones, 2020).

La complejidad de esta problemática radica en las fluctuaciones naturales de los sensores de combustible, causadas por variables operativas como el movimiento del vehículo en terrenos irregulares, cambios bruscos de altitud en zonas montañosas o variaciones térmicas en climas extremos. Estas fluctuaciones, que pueden simular caídas anómalas en el nivel de combustible, derivan en falsas alarmas que saturan a los operadores con alertas irrelevantes. A esto se suma la sofisticación de modalidades delictivas como el "robo por goteo", caracterizado por sustracciones graduales y recurrentes, casi imperceptibles para sistemas basados en umbrales estáticos.

Según estudios y análisis del sector, el robo de combustible en flotas de transporte en Colombia se perfila como una amenaza operativa que presiona la seguridad y la eficiencia de las operaciones logísticas. Las revisiones sobre riesgo en el transporte de carga destacan que, en los principales corredores viales del país, los incidentes contra el autotransporte elevan la exposición de las empresas y obligan a reforzar controles internos (Deluque Iguarán, 2018), mientras que el gremio transportador ha señalado incrementos de problemas de inseguridad en diversas zonas que afectan la continuidad del servicio (Colfecar, 2020). En paralelo, la literatura sobre economías ilícitas evidencia que la ilegalidad asociada a combustibles (p. ej., contrabando en fronteras) agrava el entorno operativo y genera externalidades negativas para las compañías formales del sector (Guerrero Burbano, 2022).

En este contexto, se evidencia una brecha tecnológica en la industria: la ausencia de sistemas inteligentes capaces de analizar patrones complejos y generar alertas accionables en tiempo real. La solución no radica únicamente en mejorar la precisión de los sensores, sino en desarrollar herramientas que interpreten datos de manera holística, considerando el entorno operativo de cada flota vehicular. Este avance no solo mitigaría las pérdidas económicas, sino que fortalecería la seguridad en el transporte de carga, un pilar estratégico para la economía nacional.

Estas circunstancias refuerzan la necesidad de soluciones de monitoreo más sofisticadas que integren variables contextuales (historial de rutas, velocidad, posición GPS y señales de tanque) para distinguir eventos genuinos de consumo de las variaciones normales del nivel, disminuyendo falsos positivos y optimizando la gestión de flotas. En la región, investigaciones recientes muestran el potencial de la telemetría e IoT para combinar sensores de nivel con GPS y comunicaciones celulares en la detección y alerta de sustracciones (Flores-Cortez & Gonzales Crespin, 2023), mientras que revisiones sectoriales en Colombia subrayan que la gestión de flotas y la optimización logística son frentes prioritarios para mejorar desempeño y seguridad (Abouarghoub et al., 2021).

5. PLANTEAMIENTO DEL PROBLEMA

El robo de combustible en vehículos de carga representa un desafío crítico para las empresas logísticas en Colombia. Este fenómeno, que afecta tanto a diésel como a gasolina, puede generar incrementos significativos en los costos operativos y está frecuentemente asociado a redes delictivas organizadas (Noticias RCN, 2020). En el contexto nacional, la implementación de tecnologías IoT se ha consolidado como una pieza clave para transformar la gestión de flotas y el seguimiento de activos (LatinPyme, 2020).

El problema se agrava debido a las fluctuaciones naturales en los sensores de combustible, originadas por factores como la inclinación del terreno, el tránsito por carreteras sin pavimentar o las variaciones de temperatura. Por ejemplo, en tractomulas que operan en regiones montañosas, los cambios bruscos de altitud pueden generar mediciones erráticas que se interpretan erróneamente como robos, saturando a los operadores con alertas falsas. Además, modalidades sofisticadas de sustracción —como el "hurto por goteo" o la manipulación externa de sensores— requieren modelos de detección capaces de analizar patrones temporales y correlacionar múltiples variables (velocidad, RPM y geolocalización) para identificar con precisión las anomalías (Pluxee, 2020).

Las limitaciones de los sistemas vigentes radican en tres aspectos clave:

- Algoritmos estáticos: Basados en umbrales fijos de consumo, ignoran el contexto operativo (ej: un camión detenido en pendiente puede mostrar caídas falsas en el nivel de combustible).
- Integración insuficiente: Los sensores no intercambian datos en tiempo real con plataformas como FuelControl, lo que retrasa la generación de alertas.

Datos no contextualizados: No se correlacionan variables críticas como la altitud GPS o el historial de mantenimiento del vehículo, esenciales para descartar falsos positivos. Como consecuencia, las empresas pierden hasta 72 horas valiosas investigando incidentes irrelevantes, mientras los robos reales pasan desapercibidos.

Este proyecto representa una oportunidad para GPS Control de liderar la innovación en la industria del transporte, proporcionando soluciones avanzadas que

incrementen la confianza en los sistemas de monitoreo. Al abordar el problema de manera integral, desde la identificación de las causas hasta la implementación de una solución tecnológica, se espera no solo resolver el problema específico del robo de combustible, sino también sentar las bases para el desarrollo de tecnologías más sofisticadas que puedan aplicarse en otros contextos operativos.

La implementación de un sistema confiable para la detección de robos de combustible es una necesidad urgente y una oportunidad estratégica. La combinación de análisis de datos, algoritmos avanzados y herramientas de integración permitirá no solo mitigar este problema, sino también generar un impacto positivo en la eficiencia operativa de las empresas del sector. Este proyecto no solo beneficiará a GPS Control, sino que también contribuirá al avance tecnológico en el monitoreo de flotas vehiculares, marcando un precedente en la industria.

6. MARCO TEÓRICO

6.1. Telemetría vehicular y monitoreo remoto

La telemática vehicular integra hardware embarcado, GNSS y servicios en la nube para capturar y analizar variables operativas del vehículo. En este proyecto, el nivel de combustible se mide con un sensor capacitivo inalámbrico Escort TD-BLE, instalado directamente en el tanque (perforación Ø 35 mm; posibilidad de recorte de la sonda; sellado), y calibrado con puntos Full/Empty y mediante tarificación por porciones para mapear lecturas a volumen real. El sensor transmite los datos por Bluetooth Low Energy (BLE) al tracker Teltonika FMC920, que dispone de Bluetooth 4.0, GNSS y conectividad celular para enviar los registros al servidor de monitoreo. Con esta arquitectura (sensor→BLE→tracker→nube) se obtiene una serie temporal de nivel que, combinada con ignición, velocidad y ubicación, permite el monitoreo remoto y la detección de anomalías asociadas a posible sustracción de combustible (Escort, 2023).

6.2. Sensores de nivel de combustible y fuentes de ruido

En este proyecto el nivel de combustible se mide con un sensor capacitivo inalámbrico Escort TD-BLE, que estima la altura del líquido a partir de la variación de capacitancia entre electrodos inmersos en el tanque. El propio sensor transmite por Bluetooth Low Energy (BLE) y reporta, además del nivel, temperatura, voltaje de batería y RSSI; su aplicación de configuración permite definir puntos de calibración (vacío/lleno) y ajustar un filtro de suavizado para atenuar saltos breves en la señal durante la marcha. Estos mecanismos (calibración y filtrado) son coherentes con el principio de medición capacitiva, donde la lectura depende del dieléctrico del combustible y de la geometría del sensor, por lo que la calibración en tanque es esencial para mapear la señal a volumen real (Escort, 2023).

Las principales fuentes de ruido para este tipo de medición son el “sloshing” (oleaje del combustible) producido por aceleración/frenado/curvas o irregularidades de la vía, y las vibraciones del vehículo; ambos fenómenos generan oscilaciones rápidas del nivel aparente. La literatura sobre tanques de combustible en automoción recomienda combinar suavizado de señal y análisis en tramos estacionarios (vehículo detenido) para

distinguir variaciones reales de volumen de fluctuaciones por movimiento del fluido (Wachowski et al., 2018).

6.3. Posicionamiento GNSS y detección de estancias

La precisión de posición/velocidad GNSS depende de factores como geometría satelital (HDOP), multitrayectoria y entorno. En condiciones típicas, la velocidad GNSS presenta errores muy bajos; sin embargo, la posición puede degradarse por HDOP y entorno urbano, lo que aconseja combinar velocidad/ignición y clustering espacial para identificar tramos estacionarios ("stay points") y separar conducción de detenciones al analizar nivel de combustible. Métodos de detección de estancias y clustering denso como DBSCAN son estándar para agrupar puntos en proximidad espacial/temporal y filtrar ruido de trayectorias (GPS.gov, 2024).

6.4. Preprocesamiento robusto de series temporales

Para atenuar outliers por sloshing/sensado, se recomiendan estimadores robustos como la Mediana de las Desviaciones Absolutas (MAD), que conserva robustez ante distribuciones no normales y proporciona una escala resistente para la detección de valores atípicos. La literatura estadística aplicada (NIST/ITL) discute su relación con la desviación estándar y su escalado (MADN) (NIST/ITL, 2016).

Complementariamente, la detección de cambios en series (change-point detection) permite localizar rupturas de nivel asociadas a eventos (p. ej., descensos abruptos con el vehículo detenido). Revisiones recientes encuadran estas técnicas en términos de función de costo, método de búsqueda y restricciones de número de cambios (Truong et al., 2020).

6.5. Detección y consolidación de eventos

Tras localizar candidatos por cambio de nivel en spans estacionarios, es habitual consolidar detecciones superpuestas mediante Non-Maximum Suppression (NMS), técnica clásica de visión por computador adaptada aquí a ventanas temporales: se conserva el candidato con mayor puntuación y se suprimen vecinos redundantes dentro de una tolerancia temporal. Esto reduce duplicados y mejora la trazabilidad de eventos (Neubeck & Van Gool, 2006).

6.6. Modelos de anomalía aplicados a robo de combustible

Cuando faltan etiquetas exhaustivas, los métodos no supervisados como Isolation Forest (IF) son adecuados para detectar patrones raros ("few and different") en

características derivadas (pendiente de nivel, duración del span, relación con ignición/velocidad, distancia a surtidores habituales, etc.). IF aísla anomalías construyendo árboles aleatorios que separan rápidamente instancias atípicas y ha sido ampliamente referenciado para detección de anomalías en datos tabulares (Liu et al, 2008).

6.7 Robo de combustible

El robo de combustible en flotas de transporte es un problema operativo y económico de alto impacto: afecta la eficiencia, distorsiona los costos de viaje y puede comprometer la seguridad del vehículo y la cadena logística. Debido a que las sustracciones ocurren bajo contextos variados—paradas programadas, trayectos en marcha, estaciones de servicio o incluso mediante manipulación del sistema de medición—su detección requiere integrar telemetría (GPS, nivel de combustible), reglas operativas y análisis de anomalías. Para enmarcar el análisis y estandarizar el vocabulario, la siguiente tabla clasifica las principales modalidades de robo de combustible, ofreciendo una descripción breve y las fuentes donde cada modalidad se documenta, lo que servirá como base para diseñar las reglas de negocio y los modelos de detección en este proyecto.

Tabla 1

Modalidades de robo de combustible

Modalidad	Descripción breve	Indicadores/ejemplos	Fuente
Sifoneo (“ordeña”/“popoteo”) desde el tanque	Extracción por manguera o bomba con el motor apagado; práctica extendida en flotas.	Caídas bruscas del nivel con el vehículo detenido; vocabulario operativo de “ordeña/popoteo”.	Devi, Murki, Hanmanthu, Ram, & Kamran (2023); Limón Bazán (2022); Capcha de la Cruz & Chaico Pecho (2017)
Perforación/forzamiento del tanque (“drilling”, “burglary”)	Daño directo al tanque o tapas para acceder al combustible.	Orificios o roturas; evidencias físicas y lecturas anómalas.	Devi et al. (2023)
Manipulación del indicador/sensor de combustible	Alteración del flotante/gauge para ocultar el volumen real o generar falsas alarmas.	Lecturas imposibles (p. ej., “vacío” con tanque lleno), cambios escalonados; daños en mirillas.	Devi et al. (2023); Prabhu, Deepa, & Nithya (2022)

Corte/derivación de líneas y conexiones (“bypass”)	Intervención de mangueras o “sight measures” para extraer o desviar combustible.	Líneas cortadas, mirillas aplastadas, derrames asociados.	Prabhu et al. (2022)
Fraude/colusión en estaciones de servicio y recibos	Diferencia entre lo pagado y lo realmente cargado; manipulación documental.	Recibos que no coinciden con volumen real; participación del encargado o personal interno.	Khuwaja, Ali, Mihalca, & Țarcă (2019); Udhaya Kumar, Vishnukanth, Deepak Kumar, Nandalal, & Sathishkumar (2021)
Robo durante el tránsito (en ruta)	Sustracción en movimiento o en paradas no autorizadas.	Pérdidas sistemáticas en trayectos; discrepancias vs. consumo esperado.	Devi et al. (2023)
Tomas clandestinas en ductos (“huachicoleo”)	Perforación de oleoductos/red de hidrocarburos para extraer combustible.	Presencia de “tomas” y robo a camiones cisterna; contexto México.	Limón Bazán (2022)

Nota. Elaboración propia con información extraída de múltiples artículos académicos.

6.8. Evaluación en contextos desbalanceados

El robo de combustible es raro respecto al volumen total de registros; por ello, la curva Precision–Recall (PR) es más informativa que la ROC al evaluar clasificadores en datasets desbalanceados. La literatura demuestra que PR refleja mejor la utilidad operativa al penalizar falsos positivos en escenarios de baja prevalencia. Por tanto, se recomienda optimizar umbrales con base en PR/F1 y reportar precision@k para priorización operativa (Saito & Rehmsmeier, 2015).

6.9. Calibración de puntajes y umbrales operativos

Para que una puntuación de evento sea interpretable como probabilidad y permita fijar umbrales por patrón (p. ej., descensos cortos vs. extendidos), se recomienda calibración: Platt scaling (ajuste sigmoidal) o isotonic regression, según la forma de distorsión. Estudios comparativos muestran ventajas de la regresión isotónica cuando la relación no es sigmoide (Platt, 1999).

6.10. GPS Control e Infraestructura Tecnológica

GPS Control es una empresa líder en el monitoreo satelital de vehículos y flotas, ofreciendo soluciones tecnológicas que permiten gestionar activos en tiempo real. Sus servicios incluyen, entre otros, notificaciones de alarmas, control de velocidad, geocercas

y monitoreo de temperatura y combustible, además de integrarse con sistemas ERP y aplicaciones móviles para facilitar la gestión integral de flotas (GPS Control, 2025).

En cuanto a su infraestructura tecnológica, GPS Control utiliza el ERP Odoo—reconocido por su flexibilidad y capacidad para centralizar funciones empresariales—junto con bases de datos PostgreSQL y su extensión PostGIS para el manejo de datos geoespaciales, lo que optimiza la gestión de geocercas y trayectorias vehiculares (Odoo S.A., 2025).

7. METODOLOGÍA

7.1. Enfoque Metodológico Adoptado

El presente proyecto adoptó la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) como marco estructural para el desarrollo del sistema de detección de robo de combustible. Esta metodología, ampliamente reconocida en la industria de analítica de datos y aprendizaje automático, fue establecida por un consorcio de empresas líderes en minería de datos a finales de los años 90 y se ha consolidado como el estándar de facto para proyectos de ciencia de datos (Chapman et al., 2000; Shearer, 2000).

La elección de CRISP-DM se fundamenta en varias características que resultan especialmente pertinentes para este proyecto. En primer lugar, su naturaleza iterativa permite refinar continuamente los resultados mediante ciclos de retroalimentación entre las diferentes fases, lo cual es crucial cuando se trabaja con datos de sensores vehiculares que presentan alta variabilidad y ruido inherente (Wirth & Hipp, 2000). En segundo lugar, CRISP-DM proporciona una estructura sistemática que facilita la trazabilidad de decisiones técnicas y la reproducibilidad de resultados, aspectos fundamentales para la validación científica del trabajo (Martínez-Plumed et al., 2019).

A diferencia de metodologías ágiles como Scrum, orientadas principalmente a la gestión de proyectos de software, CRISP-DM está específicamente diseñada para guiar el proceso analítico desde la comprensión del problema de negocio hasta el despliegue operativo de modelos predictivos (Schröer et al., 2021). Esta orientación resulta más apropiada para documentar científicamente un pipeline completo de datos que incluye preprocesamiento, ingeniería de características, modelado estadístico y validación experimental.

El modelo CRISP-DM organiza el proceso de minería de datos en seis fases interconectadas: (1) Comprensión del negocio, (2) Comprensión de los datos, (3) Preparación de los datos, (4) Modelado, (5) Evaluación, y (6) Despliegue (Chapman et al., 2000). Cada fase produce entregables específicos que sirven como insumos para las fases subsiguientes, permitiendo una gestión estructurada del proyecto y facilitando la

comunicación entre equipos técnicos y stakeholders de negocio (Azevedo & Santos, 2008).

Es importante destacar que, siguiendo las mejores prácticas de presentación científica, este documento separa claramente la descripción metodológica (contenida en este capítulo) de los resultados y hallazgos obtenidos, los cuales se presentan en el capítulo de Análisis y Resultados. Esta separación permite evaluar independientemente el rigor del proceso seguido y la validez de las conclusiones alcanzadas (IBM Corporation, 2021).

7.2. CRISP-DM: fases y aplicación en el proyecto

La Tabla 2 presenta la correspondencia entre las fases estándar de CRISP-DM y su aplicación específica en el contexto de este proyecto de detección de robo de combustible para GPS Control, empresa integradora de soluciones telemáticas, en colaboración con su cliente MPI (operador de flota) y el fabricante de sensores Escort.

Tabla 2

Aplicación de la Metodología CRISP-DM al Proyecto de Detección de Robo de Combustible

Fase CRISP-DM	Aplicación en este proyecto (MPI / GPS Control / Escort)
Entendimiento del negocio	Revisión del portafolio de GPS Control y su operación; reuniones con soporte técnico (configuración de equipos y sensores) y con personal administrativo de MPI para metas y restricciones; exploración de problemáticas frecuentes: lecturas ruidosas y falsas extracciones (Capcha de la Cruz & Chaico Pecho, 2017; Khuwaja et al., 2019).
Comprensión de los datos	Inventario de fuentes (registros exportados de plataforma), variables (timestamp, lat/lon, velocidad, ignición, nivel por tanque) y evaluación de calidad (faltantes, duplicados, saltos imposibles); segmentación de viajes y paradas.
Preparación de los datos	Limpieza, estandarización y alineación temporal; suavizado/estimación del nivel para mitigar ruido cinemático sin perder caídas reales; construcción de variables (tasa de caída, estabilidad en parada, indicadores de reabastecimiento, contexto GPS) (Prabhu et al., 2022; Khuwaja et al., 2019).
Modelado	Baseline de reglas (umbral de caída en parada/ignición OFF + persistencia mínima + exclusión de reabastecimiento) y modelos de anomalía (p. ej., Isolation Forest, DBSCAN); (Devi et al., 2023; Udhaya Kumar et al., 2021).
Evaluación	Métricas por evento (precisión, exhaustividad, F1) y por vehículo/zona; validación

	temporal (split por ventanas/vehículos) y análisis de sensibilidad a filtros/umbrales.
Despliegue	Integración de reglas/modelos en plataforma: alerta georreferenciada, bitácora y proceso de recalibración por sensor/ruta; envío de notificaciones (Udhaya Kumar et al., 2021).

Nota. Elaboración propia basada en Chapman et al. (2000) y adaptada al contexto específico del proyecto.

7.2.1. Entendimiento del Negocio

Esta fase inicial se desarrolló mediante un enfoque combinado de entrevistas semiestructuradas y revisión documental extensiva. El proceso comenzó con el análisis del portafolio de servicios de GPS Control, identificando el monitoreo de rutas y combustible para flotas de carga como servicio central, con potencial de extensión a otros segmentos vehiculares (motocicletas, automóviles particulares, transporte de pasajeros).
Se realizaron múltiples sesiones de trabajo con diferentes stakeholders:

Personal técnico de GPS Control. Responsables de la instalación, configuración y soporte de dispositivos telemáticos. Estas sesiones permitieron documentar procedimientos de instalación, parámetros de configuración típicos y problemáticas técnicas recurrentes reportadas por clientes.

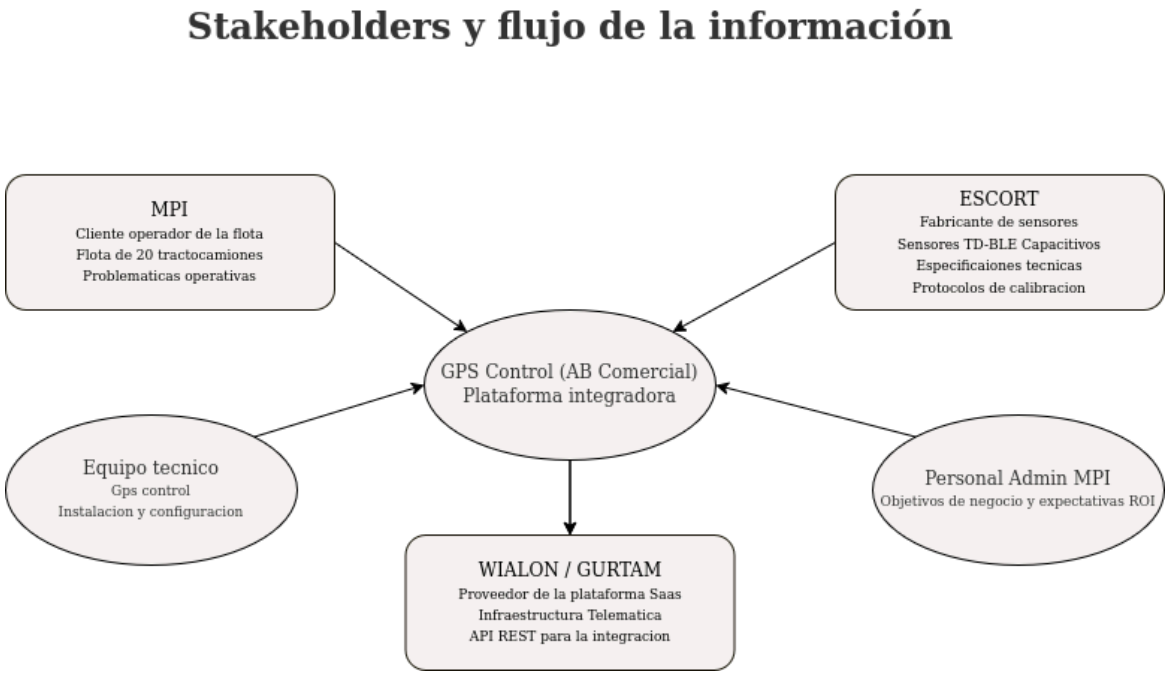
Personal administrativo de MPI. El personal administrativo de la empresa cliente participó en varias reuniones que resultaron fundamentales para comprender los objetivos operativos, restricciones presupuestarias, prioridades de negocio y expectativas de retorno de inversión del sistema de detección.

Técnicos de Escort. El fabricante de sensores de nivel de combustible proporcionó información técnica detallada sobre especificaciones de sensores, parámetros de calibración óptimos, recomendaciones de instalación física y limitaciones tecnológicas inherentes a la medición de nivel en tanques vehiculares.
Como resultado de este levantamiento, se identificaron problemáticas críticas que debían ser abordadas por el sistema:

- Lecturas ruidosas durante la marcha del vehículo, causadas por el movimiento del combustible dentro del tanque
- Configuraciones subóptimas de sensores que generan desviaciones sistemáticas en las mediciones

- Falsos positivos en la detección de extracciones, principalmente debidos a la no diferenciación entre fluctuaciones normales y caídas anómalas de nivel
- Falta de integración entre variables operativas (velocidad, ubicación, estado de ignición) para contextualizar los cambios de nivel de combustible.

Figura 1.
Diagrama de stakeholders y flujo de información en la fase de comprensión del negocio



7.2.2. Comprensión de los Datos

La fase de comprensión de datos se estructuró en torno a cuatro actividades principales: inventario de fuentes, caracterización de variables, evaluación de calidad y análisis exploratorio.

Inventario de fuentes de datos: Se recopilaron registros históricos de telemetría correspondientes a cuatro vehículos de carga pesada (tractocamiones) durante un período de seis meses (noviembre 2024 - abril 2025). El dataset resultante contiene 733,022 observaciones distribuidas en ocho archivos mensuales, cada uno en formato CSV o XLSX exportado desde la plataforma Wialon/Gurtam.

Caracterización de variables: El diccionario de datos (ver Tabla 3) documenta las variables clave contenidas en los registros de telemetría. Estas incluyen identificadores únicos de vehículo, marcas temporales con resolución de segundos, coordenadas geográficas en sistema WGS84, nivel de combustible medido por sensor capacitivo, velocidad instantánea derivada de GPS y estado de ignición del motor.

Evaluación de calidad de datos. Se implementó un protocolo sistemático de evaluación que identificó:

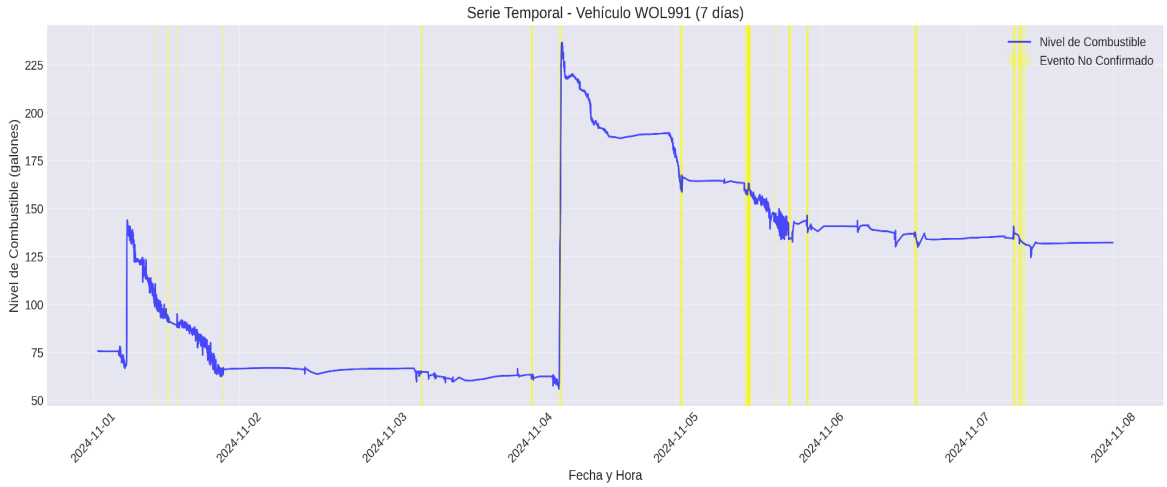
- 13 coordenadas GPS inválidas (fuera de rangos geográficos plausibles: latitud ∉ [-90, 90] o longitud ∉ [-180, 180])
- 2,949 valores faltantes en lecturas de nivel de combustible (0.4% del total de registros)
- 172 brechas temporales superiores a 1 hora entre registros consecutivos
- Valores duplicados exactos en 47 registros, indicativos de retransmisiones

Análisis exploratorio. Mediante visualizaciones de series temporales se caracterizaron:

- Patrones de frecuencia de muestreo variable: 1 mensaje/hora con vehículo estacionado e ignición apagada; 1 mensaje/minuto estacionado con ignición encendida; frecuencia creciente con velocidad en marcha (hasta 1 mensaje/segundo a alta velocidad)
- Ruido característico de lecturas de combustible: fluctuaciones de alta frecuencia (±0.5-2.0 galones) durante marcha en carreteras irregulares
- Eventos de reabastecimiento: incrementos abruptos de nivel (típicamente >15 galones en <5 minutos)
- Patrones de consumo normal: decremento gradual correlacionado con velocidad y tiempo de operación.

Figura 2.

Muestreo de segmentación de conducción



Nota. Serie temporal de nivel de combustible mostrando ruido, consumo normal, reabastecimiento y posible extracción

Tabla 3

Diccionario de Datos del Sistema Telemático

Variable	Tipo de Dato	Descripción	Unidades	Ejemplo	Notas
vehicle_id	String	Identificador único del vehículo	--	GQU478	Placa o código interno
timestamp	DateTime	Fecha y hora de captura del registro	--	2024-11-15 14:30:25	UTC, formato ISO 8601
latitude	Float	Coordenada geográfica de latitud	Grados decimales	4.6097	Sistema WGS84 (EPSG:4326)
longitude	Float	Coordenada geográfica de longitud	Grados decimales	-74.0821	Sistema WGS84 (EPSG:4326)
total_fuel_gal	Float	Nivel total de combustible en tanques	Galones	85.3	Suma de tanques izquierdo + derecho

speed_kmh	Float	Velocidad instantánea del vehículo	km/h	62.4	Derivada de GPS (no velocímetro)
ignition	Integer	Estado de ignición del motor	--	1	0 = apagado, 1 = encendido
stationary_drain	Integer	Etiqueta de evento de robo (variable objetivo)	--	1	0 = normal, 1 = robo confirmado

Nota Elaboración propia. Las coordenadas geográficas utilizan el datum WGS84. El nivel de combustible es medido por sensor capacitivo Escort con precisión nominal de ±1%. La variable stationary_drain fue etiquetada manualmente mediante inspección experta.

7.2.3. Preparación de los Datos

La preparación de datos constituyó la fase más intensiva del proyecto, representando aproximadamente el 60% del esfuerzo total de desarrollo. El proceso se estructuró en cuatro subprocesos principales: limpieza y consolidación, segmentación de trayectorias, cálculo de umbrales adaptativos y generación de eventos candidatos.

Limpieza y consolidación. Los ocho archivos mensuales fueron integrados en un único dataset mediante un proceso de concatenación que preservó el orden cronológico y eliminó 47 registros duplicados exactos. Se implementaron las siguientes transformaciones:

- Estandarización de formatos de fecha-hora a UTC con zona horaria explícita
- Validación y corrección de coordenadas GPS (13 registros con coordenadas inválidas fueron marcados como faltantes)
- Imputación controlada de valores faltantes: 2,949 registros con nivel de combustible ausente fueron tratados mediante interpolación lineal temporal, limitada a gaps de máximo 3 puntos consecutivos y aplicada únicamente dentro de segmentos estacionarios
- Conversión de tipos de datos: velocidad y nivel de combustible a float64, estado de ignición a booleano.

Segmentación de trayectorias. Para facilitar el análisis contextual, los registros fueron segmentados en períodos estacionarios y en movimiento mediante los siguientes criterios:

- Período estacionario: Velocidad ≤ 1 km/h durante al menos 3 minutos consecutivos
- Cambio de modo: Transiciones entre estados estacionario/movimiento o cambios de estado de ignición fueron identificadas como límites de segmento
- Manejo de gaps temporales: Brechas >5 minutos en modo estacionario con ignición encendida, o >120 minutos con ignición apagada, fueron tratadas como separadores de segmentos independientes

Esta segmentación resultó en la identificación de 2,547 segmentos estacionarios, de los cuales 1,832 cumplieron el criterio de duración mínima (≥3 minutos) para análisis posterior.

Cálculo de umbrales adaptativos. Reconociendo que diferentes vehículos y condiciones operativas presentan niveles de ruido distintos, se implementó un sistema de umbrales adaptativos basado en la Desviación Absoluta Mediana (MAD), una medida robusta de dispersión menos sensible a valores atípicos que la desviación estándar (Leys et al., 2013).

Para cada combinación de (vehículo, estado_operativo), donde estado_operativo ∈ {estacionario_con_ignición, estacionario_sin_ignición}:

1. Se calculó la serie de diferencias de nivel: Δfuel[i] = fuel[i] - fuel[i-1]
2. Se computó la MAD: MAD = 1.4826 × median(|Δfuel - median(Δfuel)|)
3. Se definieron umbrales de detección:
    - threshold_step = max(3.0 × MAD, 1.0 galones) # Umbral para cambio individual
    - threshold_cumulative = max(4.0 × MAD, 3.5 galones) # Umbral para cambio acumulado

El factor 1.4826 es un término de consistencia que hace que MAD sea comparable con la desviación estándar bajo normalidad (Rousseeuw & Croux, 1993).

Generación de eventos candidatos. Utilizando los umbrales adaptativos, se implementó un detector de ventana deslizante que identificó 1,491 eventos candidatos de extracción. El algoritmo opera de la siguiente manera:

Para cada segmento estacionario:

1. Extraer subserie de niveles de combustible dentro del segmento
2. Para cada ventana temporal [t_inicio, t_fin] dentro del segmento:
    - Calcular caída acumulada: drop = fuel[t_inicio] - fuel[t_fin]
    - Identificar el mayor cambio negativo individual: min_step = min(Δfuel[t_inicio:t_fin])
    - Si (drop > threshold_cumulative) O (min_step < -threshold_step):
        - Registrar evento candidato con sus atributos (tiempo inicio/fin, magnitud, contexto operativo)
        - Aplicar Non-Maximum Suppression (NMS) para evitar detecciones duplicadas de un mismo evento físico

Filtrado de plausibilidad física. Se descartaron 127 eventos candidatos (8.5% del total) que violaban restricciones físicas:

- Tasa de caída >2.0 galones/minuto (excede capacidad típica de extracción por sifón)
- Cambio individual >50 galones (indica fallo de sensor más que evento real)

Figura 3.
Pipeline de preparación de datos

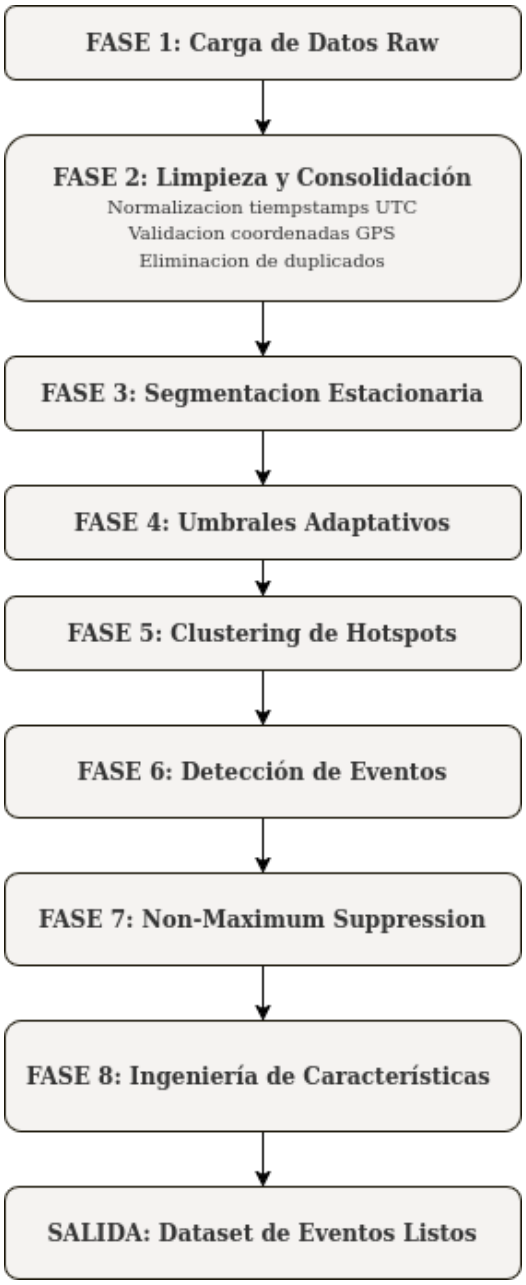


Tabla 4

Variables Diseñadas para la Detección de Eventos de Robo

Variable	Tipo	Descripción	Cálculo/Lógica	Relevancia
time_diff_s	Float	Tiempo transcurrido desde registro anterior	timestamp[i] - timestamp[i-1]	Normalización temporal
fuel_diff	Float	Cambio en nivel de combustible	total_fuel[i] - total_fuel[i-1]	Señal primaria de detección
is_stationary	Boolean	Indicador de vehículo detenido	speed_kmh < 1.0	Condición H1 (vehículo estacionario)
ignition_off	Boolean	Indicador de motor apagado	ignition == 0	Condición H2 (ignición apagada)
drop_gal	Float	Magnitud acumulada de caída en evento	sum(fuel_diff[t_start:t_end]) para fuel_diff < 0	Severidad del evento
min_step_gal	Float	Mayor caída individual en evento	min(fuel_diff[t_start:t_end])	Detección de extracciones rápidas
duration_min	Float	Duración temporal del evento	(t_end - t_start) / 60	Clasificación de patrones
n_negative_steps	Integer	Cantidad de mediciones con caída	count(fuel_diff < threshold_noise)	Persistencia del patrón
median_dt_s	Float	Mediana de intervalos entre mediciones	median(time_diff_s[t_start:t_end])	Control de calidad temporal
is_hotspot	Boolean	Ubicación geográfica recurrente	cluster_id >= 0 (DBSCAN)	Factor de riesgo contextual
cluster_count	Integer	Frecuencia de eventos en la ubicación	count(events en cluster)	Severidad de hotspot
pattern	Categorical	Clasificación del tipo de evento	Reglas basadas en duración y estado	Segmentación para modelado

hod_sin, hod_cos	Float	Componentes cíclicas de hora del día	sin(2π×hour/24), cos(2π×hour/24)	Patrones temporales
is_weekend	Boolean	Indicador de fin de semana	weekday >= 5	Patrones temporales
is_night	Boolean	Indicador de horario nocturno	(hour >= 22) OR (hour <= 5)	Factor de riesgo temporal
rate_gpm	Float	Tasa de caída de combustible	drop_gal / duration_min	Filtro de plausibilidad
pct_ign_on	Float	Proporción del evento con ignición encendida	mean(ignition[t_start:t_end])	Contexto operativo
median_speed	Float	Velocidad mediana durante evento	median(speed_kmh[t_start:t_end])	Validación de estado estacionario

Nota. Elaboración propia. Las variables relacionadas con tiempo cíclico (hod_sin, hod_cos) se calculan para evitar discontinuidad artificial en la frontera 23:59→00:00. El clustering espacial (DBSCAN) utiliza distancia haversine con ε=60m y min_samples=12.

7.2.4. Modelado

La estrategia de modelado adoptó un enfoque híbrido que combina reglas de negocio interpretables con técnicas de aprendizaje automático, buscando balancear precisión predictiva con explicabilidad operativa. Se implementaron tres niveles de detección:

Nivel 1 - Modelo baseline basado en reglas. Sistema de reglas determinísticas derivadas del conocimiento experto y análisis exploratorio:

```
REGLA 1 (Caída estacionaria - H1):
  SI (fuel_diff_smoothed < threshold_stationary) Y (is_stationary == TRUE)

  ENTONCES marcar_como_sospechoso

REGLA 2 (Caída con ignición apagada - H2):

  SI (fuel_diff_smoothed < threshold_ignition_off) Y (ignition_off == TRUE)
```

```
  ENTONCES marcar_como_sospechoso

REGLA 3 (Condición combinada de alto riesgo - H1 ∧ H2):

  SI REGLA1 Y REGLA2

  ENTONCES marcar_como_alto_riesgo

FILTROS DE EXCLUSIÓN:

  - Excluir si evento coincide temporalmente con reabastecimiento (incremento >15
gal en <10 min)

  - Excluir si tasa de caída excede límite físico (>2.0 gal/min)

  - Excluir si duración < 3 minutos (eventos espurios)
```

Este modelo baseline sirve como punto de referencia para evaluar el valor agregado de técnicas más sofisticadas.

Nivel 2 - Modelos de detección de anomalías no supervisados. Dado que el etiquetado manual de eventos es costoso y subjetivo, se exploraron técnicas no supervisadas:

- **Isolation Forest (Liu et al., 2008)**: Algoritmo basado en particiones aleatorias que identifica puntos anómalos como aquellos que requieren menos particiones para ser aislados. Configuración: n_estimators=200, contamination=0.10 (asumiendo ~10% de eventos como anómalos), random_state=42.
- **DBSCAN espacial (Ester et al., 1996)**: Clustering basado en densidad aplicado a coordenadas GPS para identificar hotspots geográficos de eventos recurrentes. Configuración: ε=60m (radio de búsqueda), min_samples=12 (mínimo de eventos para formar cluster), métrica=haversine.

Nivel 3 - Modelos supervisados calibrados. Para el subconjunto de datos con etiquetas validadas (106 eventos positivos de 1,491 candidatos), se entrenaron modelos supervisados:

- Regresión Logística con regularización L2: Modelo lineal que proporciona coeficientes interpretables. Configuración: C=1.0 (inverso de fuerza de regularización), class_weight='balanced', solver='lbfgs', max_iter=1000.
- Random Forest (Breiman, 2001): Ensamble de árboles de decisión que captura no-linealidades e interacciones. Configuración: n_estimators=200, max_depth=10, min_samples_split=20, min_samples_leaf=10, class_weight='balanced_subsample'.

Ambos modelos fueron sometidos a calibración probabilística mediante validación cruzada estratificada de 5 folds usando el método sigmoid/Platt scaling (Platt, 1999), mejorando la interpretabilidad de las probabilidades predichas.

Partición temporal de datos. para simular condiciones de despliegue real, se implementó una partición temporal estricta:

- 80% cronológicamente más antiguo → conjunto de entrenamiento
- 20% cronológicamente más reciente → conjunto de prueba

Esta estrategia (forward chaining) previene el data leakage temporal y proporciona una estimación más realista del rendimiento en producción que la validación cruzada aleatoria (Cerqueira et al., 2020).

Definición de patrones de evento. los 1,491 eventos candidatos fueron clasificados en tres patrones basados en características operativas:

1. short_4_10m_3gal: Drenajes de corta duración (4-10 minutos) con caída ≥3 galones, típicamente durante paradas breves. n=1,164 eventos (53 positivos).
2. extended_15m_6gal: Drenajes prolongados (≥15 minutos) con caída ≥6 galones, asociados a paradas largas. n=230 eventos (15 positivos).
3. postjourney_off: Drenajes post-viaje (≥45 minutos) con ignición apagada y caída ≥6 galones, escenario de mayor riesgo. n=97 eventos (38 positivos).

Figura 4.

Arquitectura del sistema de detección multi-nivel

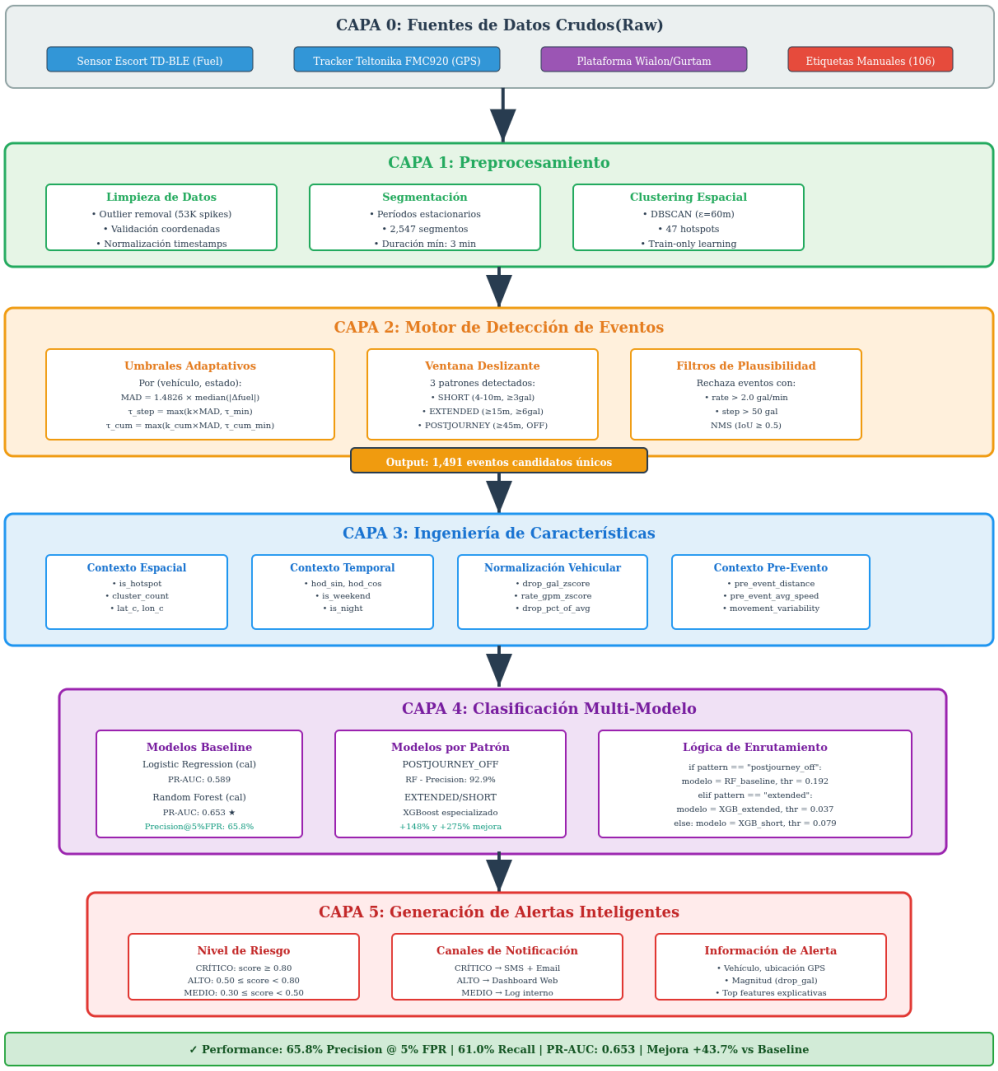


Tabla 5

Lógica de Detección de Anomalías Basada en Reglas

ID Regla	Descripción	Condición Lógica	Hipótesis Asociada	Parámetros de Umbral
R1	Caída en vehículo estacionario	(fuel_diff < threshold_stationary) ∧ (is_stationary)	H1: Mayor probabilidad de robo cuando vehículo detenido	threshold_stationary : max(3×MAD, 1.0 gal) por vehículo

R2	Caída con ignición apagada	(fuel_diff < threshold_ignition_off) ∧ (ignition_off)	H2: Mayor probabilidad de robo con motor apagado	threshold_ignition_off: max(2.5×MAD, 1.5 gal) por vehículo
R3	Condición combinada de alto riesgo	R1 ∧ R2	H1 ∧ H2: Máximo riesgo en intersección	Hereda umbrales de R1 y R2
F1	Filtro de reabastecimiento	Incremento >15 gal en <10 min precediendo caída	--	Umbral fijo basado en capacidad típica de surtidor
F2	Filtro de plausibilidad física	rate_of_drop > 2.0 gal/min	--	Límite superior de tasa de sifón manual
F3	Filtro de duración mínima	duration < 3.0 min	--	Elimina fluctuaciones transitorias

Nota. Elaboración propia. Los umbrales adaptativos se recalculan mensualmente por vehículo usando datos históricos de operación normal. MAD = Median Absolute Deviation, medida robusta de dispersión (Leys et al., 2013).

7.2.5. Evaluación

La evaluación del sistema de detección se diseñó considerando las características específicas del problema: alto desbalance de clases (~7% de eventos positivos), costo asimétrico de errores (falso negativo >> falso positivo en términos de pérdida económica) y necesidad de validación temporal para estimar rendimiento en condiciones de producción.

Métricas de evaluación. se priorizaron métricas apropiadas para problemas de detección de anomalías con clases desbalanceadas:

- ***PR-AUC (Area Under Precision-Recall Curve).*** Métrica robusta al desbalance que resume el trade-off entre precisión y recall a través de múltiples umbrales de decisión. Valores cercanos a 1.0 indican excelente desempeño; valores cercanos a la prevalencia de positivos (0.071 en este caso) indican desempeño equivalente a clasificación aleatoria.

- ***Recall@FPR≤5%.*** Tasa de verdaderos positivos manteniendo la tasa de falsos positivos bajo 5%, reflejando la restricción operativa de tolerancia a falsas alarmas. Esta métrica se calcula identificando el umbral de probabilidad donde FPR=0.05 en la curva ROC y evaluando el recall correspondiente.
- ***Precision@5%FPR.*** Proporción de alertas generadas que corresponden a robos reales, operando en el punto de 5% FPR.
- ***Matriz de confusión****.* Desglose de verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN), permitiendo calcular métricas derivadas y analizar tipos de error específicos.

Estrategia de validación. se implementó un esquema de validación temporal multi-nivel:

1. ***Validación global***. Partición temporal 80/20 aplicada a todo el dataset, preservando el orden cronológico de eventos para cada vehículo.
2. ***Validación por vehículo*****.** Análisis de rendimiento individualizado para evaluar la capacidad de generalización del modelo a diferentes unidades de la flota, cada una con características operativas y patrones de ruido distintos.
3. ***Validación por patrón.*** Evaluación separada para cada tipo de evento (short, extended, postjourney_off), reconociendo que diferentes patrones presentan distintos niveles de dificultad de detección.
4. ***Análisis de sensibilidad.*** Evaluación sistemática del impacto de hiperparámetros clave:
    - Umbrales de caída (threshold_stationary, threshold_ignition_off)
    - Parámetros de filtrado temporal (duración mínima de evento)
    - Configuración de DBSCAN para hotspots (ε, min_samples)

Validación de ausencia de data leakage. se implementaron verificaciones explícitas para confirmar la integridad temporal de la partición:

```
Para cada vehículo v:
  max_train_time[v] = max(timestamp donde is_train==True)
  min_test_time[v] = min(timestamp donde is_train==False)
```

```
ASSERT max_train_time[v] < min_test_time[v]
```

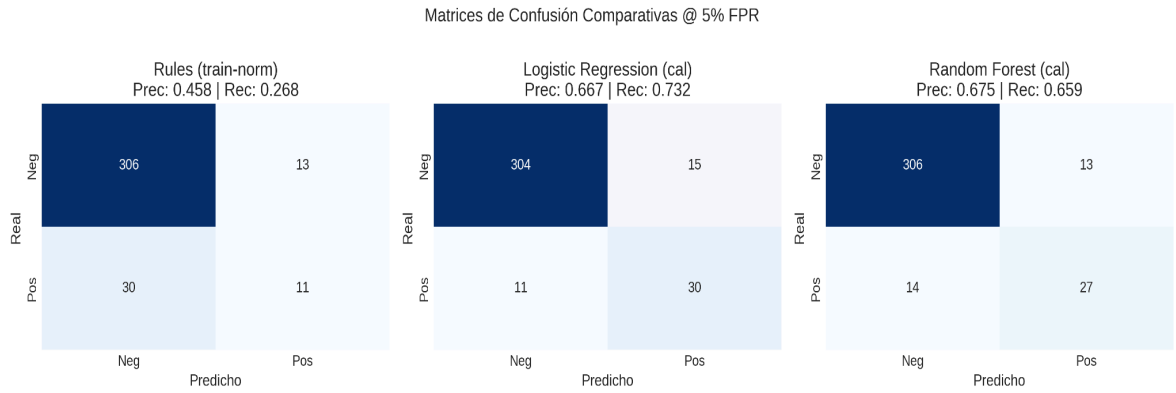
Nota. Todas las unidades pasaron esta verificación, confirmando que los umbrales adaptativos y clusters espaciales fueron aprendidos exclusivamente de datos de entrenamiento.

Comparación con baseline. los modelos de aprendizaje automático fueron comparados sistemáticamente contra el sistema de reglas baseline, calculando el incremento porcentual en cada métrica:

$$\Delta(métrica) = [(métrica_ML - métrica_baseline) / métrica_baseline] \times 100\%$$

Esta comparación cuantifica el valor agregado de la complejidad adicional introducida por los modelos ML.

Figura 5.
Matriz de confusión



Nota. Matriz de confusión comparativa para modelos baseline, LR y RF en conjunto de prueba

7.2.6. Despliegue

La fase de despliegue contempló la integración del sistema de detección en el flujo operativo de monitoreo de GPS Control, considerando aspectos técnicos, operacionales y de mantenimiento a largo plazo.

Arquitectura de integración. el sistema desplegado opera mediante los siguientes componentes:

1. ***Pipeline de preprocesamiento.*** Módulo que consume datos crudos de la API de Wialon/Gurtam, aplica transformaciones de limpieza y calcula variables derivadas en tiempo cuasi-real (latencia <5 minutos desde captura de sensor hasta disponibilidad para análisis).
2. ***Motor de detección.*** Servicio que ejecuta el modelo Random Forest calibrado sobre ventanas deslizantes de datos recientes, generando scores de probabilidad para eventos candidatos.
3. ***Sistema de alertas.*** Componente que evalúa scores contra umbrales específicos por patrón y genera notificaciones:
    - **Alertas de atención (0.50 ≤ p < 0.80).** Registro en dashboard web con revisión en siguiente jornada laboral
    - **Eventos de interés (0.30 ≤ p < 0.50).** Almacenamiento en bitácora para análisis retrospectivo
4. ***Sistema de bitácora.*** Base de datos que registra todos los eventos detectados, incluyendo:
    - Identificación del vehículo y ubicación GPS
    - Marca temporal de inicio y fin del evento
    - Magnitud de la caída de combustible (drop_gal, min_step_gal)
    - Score de probabilidad y patrón asignado
    - Estado de validación (pendiente, confirmado, descartado)
    - Acciones correctivas tomadas

Proceso de recalibración. Dado que las características de los sensores y patrones operativos evolucionan con el tiempo, se diseñó un protocolo de mantenimiento preventivo:

- ***Recalibración mensual de umbrales.*** Los umbrales adaptativos (threshold_stationary, threshold_ignition_off) se recalculan usando los datos del mes anterior, permitiendo que el sistema se adapte a cambios graduales en el ruido del sensor o condiciones operativas.
- ***Re-entrenamiento trimestral de modelos.*** Los modelos de Machine Learning (Random Forest, Logistic Regression) se re-entrenan incorporando nuevos eventos etiquetados, mejorando progresivamente la capacidad de generalización.

- ***Auditoría semestral de rendimiento.*** Análisis formal de métricas (PR-AUC, Precision@5%FPR) en datos recientes, comparando contra línea base histórica para detectar degradación del modelo.

Retroalimentación operacional. se implementó un ciclo de aprendizaje continuo mediante la captura de feedback de usuarios:

- Confirmación/rechazo de alertas por parte de gestores de flota
- Registro de eventos de robo reportados manualmente que no fueron detectados automáticamente
- Documentación de causas de falsos positivos (ej: reabastecimiento no registrado, fallo de sensor)

Este feedback alimenta el proceso de re-entrenamiento, permitiendo que el sistema evolucione con base en experiencia operativa real.

Interfaz de usuario. la integración en la plataforma web FuelControl proporciona:

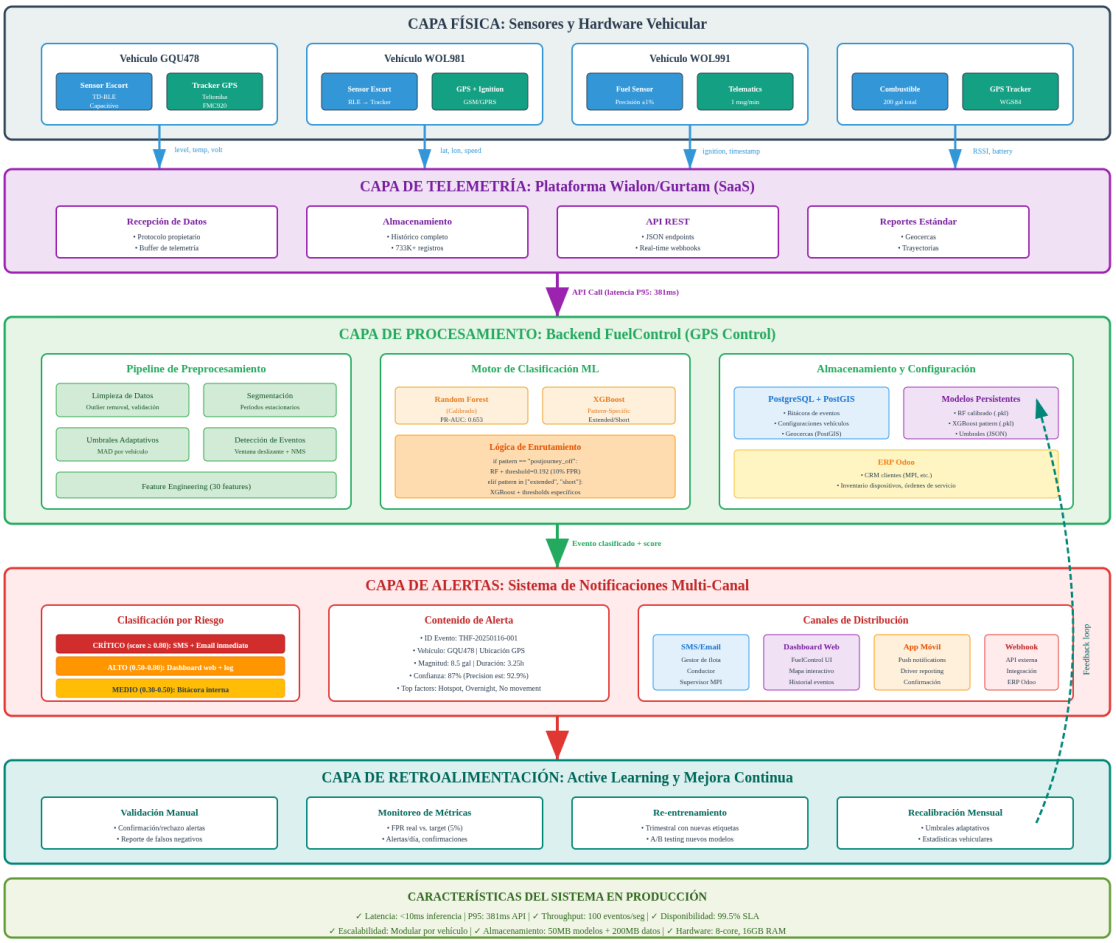
- ***Dashboard de monitoreo.*** Mapa interactivo mostrando ubicación de alertas activas con código de colores por severidad
- ***Panel de análisis.*** Gráficos de tendencias temporales (eventos por día/semana/mes), distribución geográfica de hotspots, estadísticas por vehículo
- ***Módulo de reportes.*** Generación automática de informes semanales en formato PDF incluyendo resumen ejecutivo, lista de eventos detectados, recomendaciones de acción y comparativa vs. período anterior

Consideraciones de escalabilidad. el diseño modular permite escalar el sistema para flotas de mayor tamaño:

- Procesamiento paralelo por vehículo (sin dependencias entre unidades)
- Almacenamiento incremental (solo eventos detectados, no series temporales completas)
- Carga computacional proporcional a número de eventos candidatos, no a volumen total de telemetría

Figura 6.

Diagrama de arquitectura del sistema desplegado, mostrando flujo de datos desde sensores hasta alertas



8. RESULTADOS

8.1 Contexto Empresarial y Operativo

GPS Control es una empresa colombiana especializada en soluciones de monitoreo satelital y gestión de flotas vehiculares, con sede en Bogotá D.C. Fundada en 2019, la compañía ofrece servicios integrales de telemetría que incluyen seguimiento GPS en tiempo real, control de combustible, monitoreo de temperatura, gestión de geocercas y generación de reportes operativos (GPS Control, 2025).

La empresa opera como integradora de tecnología, combinando hardware de fabricantes especializados (dispositivos GPS, sensores de combustible Escort, módulos de comunicación GSM/GPRS) con plataformas de software (Wialon/Gurtam para gestión de telemetría, Odoo como ERP corporativo). Su cartera de clientes abarca diversos segmentos del sector transporte: flotas de carga pesada, empresas de logística, transporte de pasajeros, servicios de mensajería y vehículos particulares (GPS Control, 2025).

Infraestructura tecnológica

La arquitectura tecnológica de GPS Control se fundamenta en tres pilares:

Plataforma de telemetría Wialon/Gurtam. Sistema SaaS (Software as a Service) proporcionado por el fabricante bielorruso Gurtam, que recibe y procesa mensajes de telemetría desde dispositivos instalados en vehículos. Wialon maneja la recepción de datos por protocolo propietario, almacenamiento histórico, generación de reportes estándar y provisión de API REST para integraciones personalizadas (Gurtam, 2025).

ERP Odoo. Sistema de planificación de recursos empresariales open-source utilizado para gestión de clientes, facturación, inventario de dispositivos, órdenes de servicio técnico y CRM. Odoo se seleccionó por su flexibilidad, módulos personalizables y capacidad de integración vía API (Odoo S.A., 2025).

Base de datos PostgreSQL con PostGIS. Motor de base de datos relacional con extensión espacial PostGIS, utilizado para almacenar datos complementarios (configuraciones de vehículos, bitácora de eventos, umbrales personalizados por cliente) y realizar consultas geoespaciales eficientes (geocercas, cálculo de distancias, clustering de

ubicaciones) (PostgreSQL Global Development Group, 2025; PostGIS Project Steering Committee, 2025).

Cliente del proyecto - MPI

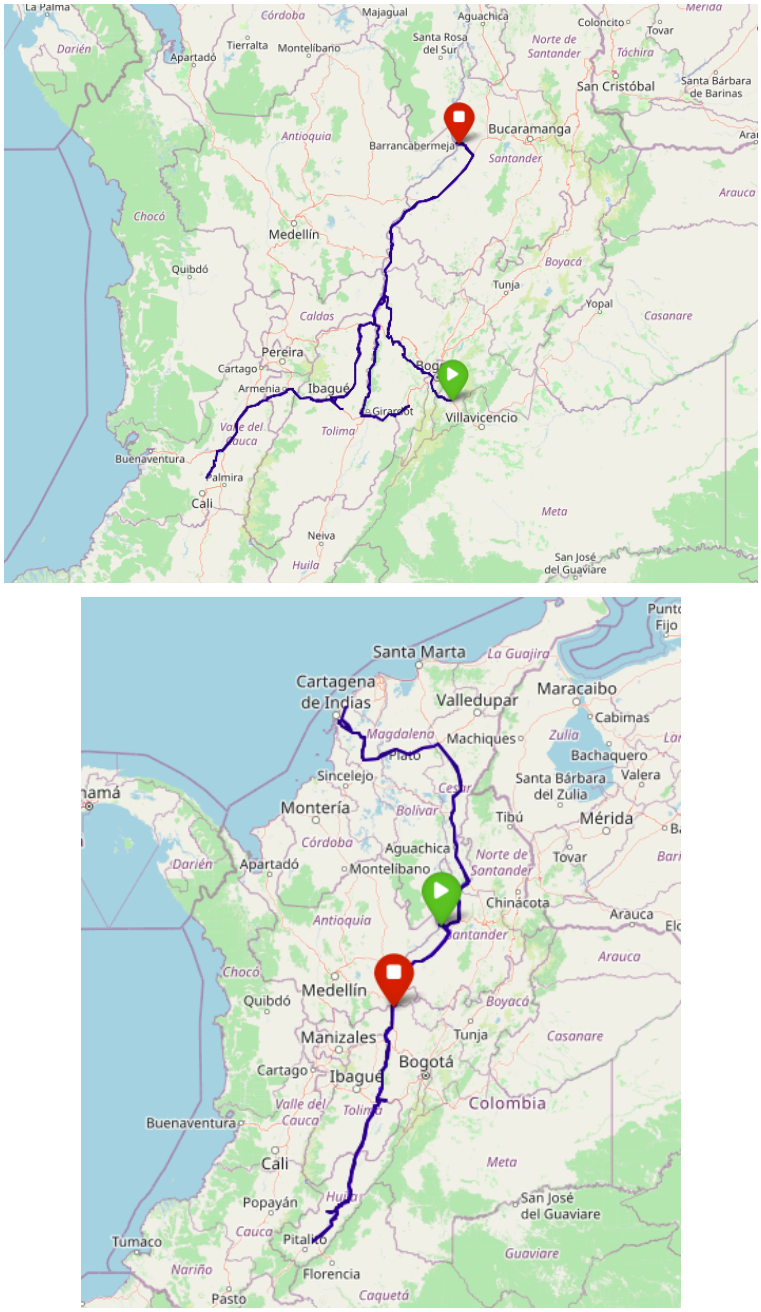
El cliente seleccionado para este proyecto piloto es MPI (Materiales y Proyectos Industriales), empresa de transporte de materiales de construcción que opera una flota de aproximadamente 20 tractocamiones en la región central de Colombia (corredores Bogotá-Medellín, Bogotá-Villavicencio, Bogotá-Tunja). MPI fue cliente pionero en la adopción del sistema de monitoreo de combustible de GPS Control (implementación iniciada en junio de 2024), manteniendo previamente solo sistemas de GPS básico sin sensores de nivel.

La flota de MPI presenta características heterogéneas:

- Vehículos con sensor de combustible: 12 unidades equipadas con sensores capacitivos Escort montados en tanques
- Vehículos solo con GPS: 8 unidades con seguimiento de ubicación pero sin medición de combustible
- Antigüedad de la flota: Vehículos modelo 2015-2022, principalmente marcas Kenworth, Freightliner e International
- Capacidad de tanques: Típicamente 100-120 galones por lado (200-240 galones totales)
- Patrones operativos: Viajes intermunicipales de larga distancia (200-500 km por ruta), con frecuencia de 3-5 viajes por semana por vehículo

Figura 7

Mapa mostrando principales rutas operativas de la flota MPI



Problemática identificada

Durante las entrevistas con la personal administrativo de MPI y el análisis de registros históricos, se documentaron las siguientes problemáticas recurrentes:

1. **Falsos positivos en alertas**: El sistema básico de Wialon generaba alertas de caída de combustible basadas en umbrales fijos (ej: caída >5 galones en 10 minutos), sin considerar contexto operativo. Esto resultaba en 15-30 alertas semanales por vehículo, de las cuales >80% correspondían a fluctuaciones normales del sensor o consumo legítimo, saturando al personal de monitoreo.
2. **Robos no detectados:** Sustracciones graduales ("robo por goteo") de 2-3 galones durante paradas prolongadas no superaban los umbrales configurados, acumulando pérdidas mensuales estimadas en $500,000-800,000 COP por vehículo sin generar alertas.
3. **Reabastecimientos sin registro formal:** Conductores ocasionalmente realizaban recargas en estaciones no autorizadas (sin factura), generando incrementos de nivel que el sistema no diferenciaba de ajustes manuales de calibración del sensor.
4. **Configuraciones subóptimas de sensores:** Durante las sesiones técnicas con Escort se identificó que 3 de los 12 sensores instalados presentaban desviación sistemática (offset) de 3-5 galones debido a incorrecta calibración inicial, requiriendo re-calibración en campo.

Esta problemática motivó el desarrollo de un sistema más sofisticado que incorpora contexto operativo (velocidad, ubicación, estado de ignición, patrones históricos) para reducir falsos positivos y mejorar la detección de robos reales.

8.2 Recopilación y Caracterización del Dataset

Selección de vehículos de estudio

Del total de 12 unidades de MPI equipadas con sensor de combustible, se seleccionaron 4 vehículos para el análisis piloto basándose en los siguientes criterios de inclusión:

- **Antigüedad de instalación.** Vehículos con >6 meses de operación con sensor, garantizando suficientes datos históricos para análisis estadístico robusto
- **Calidad de datos.** Unidades con <10% de registros faltantes o erróneos en el período de análisis
- **Representatividad operativa.** Vehículos que cubren diferentes rutas y patrones de uso (corta/larga distancia, alta/baja frecuencia de viajes)

- **Diversidad de sensores.** Mezcla de modelos de sensor Escort (versiones con/sin Bluetooth, diferentes capacidades de tanque)

Los vehículos seleccionados fueron identificados mediante sus placas: GQU478, WOL981, WOL991 y WOM077. La Tabla 6 presenta las características técnicas detalladas de cada unidad.

Período de análisis y volumen de datos

Se recopilaron registros de telemetría correspondientes al período noviembre 2024 - abril 2025 (6 meses calendario), obteniendo un total de 733,022 observaciones con la siguiente distribución por vehículo:

- **GQU478:** 187,443 registros (25.6%)
- **WOL981:** 165,289 registros (22.6%)
- **WOL991:** 312,156 registros (42.6%)
- **WOM077:** 68,134 registros (9.3%)

La frecuencia de muestreo implementada es adaptativa según el estado operativo del vehículo, optimizando el uso de datos móviles mientras mantiene resolución temporal adecuada para detección de eventos:

- Vehículo estacionado + ignición apagada: ~1 registro/hora
- Vehículo estacionado + ignición encendida: ~1 registro/minuto
- Vehículo en movimiento (velocidad <30 km/h): ~1 registro/15-30 segundos
- Vehículo en movimiento (velocidad >30 km/h): ~1 registro/5-10 segundos

Esta estrategia de muestreo variable, configurada en los dispositivos GPS, permite capturar eventos críticos con alta resolución temporal durante períodos de mayor riesgo (vehículo detenido), mientras reduce el volumen de datos durante tránsito normal en carretera.

Estructura de Archivos y Proceso de Consolidación

Los datos fueron exportados desde la plataforma Wialon en 8 archivos mensuales (noviembre 2024 - abril 2025), siendo necesario en algunos casos múltiples archivos por mes debido a límites de tamaño de exportación de Wialon (máximo 100,000 registros por archivo). Los archivos presentaron el siguiente formato estándar:

- **Extensión:** CSV (Comma-Separated Values) con codificación UTF-8
- **Delimitador:** Coma (,)
- **Encabezados:** Primera fila contiene nombres de columnas en inglés

- **Formato de fechas:** ISO 8601 (YYYY-MM-DD HH:MM:SS), zona horaria UTC-5 (hora de Colombia)

El resultado fue un archivo unificado `combined_dataset.csv` de 89.4 MB conteniendo 733,022 registros únicos (se eliminaron 47 duplicados exactos, representando 0.006% del total).

Tabla 6

Características Técnicas de los Vehículos de Estudio

Vehículo	Marca/Modelo	Año	Tipo Tanque	Capacidad (gal)	Modelo Sensor Escort	Fecha Instalación	Rutas Principales
GQU478	Freightliner Cascadia	2019	Dual (Izq+Der)	100 + 100 = 200	ESCORT TD-BLE	2024-06-15	Bogotá-Medellín, Bogotá-Tunja
WOL981	Kenworth T800	2017	Dual (Izq+Der)	120 + 120 = 240	ESCORT TD-BLE	2024-06-20	Bogotá-Villavicencio
WOL991	Kenworth T800	2018	Dual (Izq+Der)	120 + 120 = 240	ESCORT TD-BLE	2024-07-01	Bogotá-Medellín, Bogotá-Cali
WOM077	Freightliner Cascadia	2020	Dual (Izq+Der)	100 + 100 = 200	ESCORT TD-BLE	2024-06-18	Bogotá-Tunja, Bogotá-Sogamoso

Nota. Elaboración propia con datos recopilados de fichas técnicas del fabricante (Escort, 2024) y registros de instalación de GPS Control. Todos los sensores son de tecnología capacitiva con precisión nominal ±1% y resolución de 0.1 galones. Los modelos ESCORT TD-BLE incluyen conectividad Bluetooth para calibración inalámbrica.

Etiquetado Manual de Eventos de Robo

Para entrenar y validar modelos supervisados, se requirió etiquetar manualmente eventos confirmados de robo mediante un protocolo sistemático de validación que combinó revisión técnica y validación operativa:

Protocolo de etiquetado

1. **Revisión visual de series temporales:** Inspección gráfica de nivel de combustible vs. tiempo para los 4 vehículos, identificando caídas abruptas candidatas mediante análisis exploratorio en Python (bibliotecas matplotlib y seaborn).
2. **Cruce con registros operativos:** Validación con planillas de conductores (paradas programadas, reabastecimientos facturados) y registros de GPS (ubicación durante evento) proporcionados por MPI.
3. **Criterios de confirmación positiva:**
    - Caída >5 galones en <30 minutos
    - Vehículo detenido (velocidad <1 km/h) durante la caída
    - Sin reabastecimiento documentado posterior que explicaría la caída como consumo normal
    - Ubicación NO corresponde a estación de servicio registrada en base de datos de MPI
4. **Consenso experto:** Casos ambiguos fueron revisados por personal administrativo de flota MPI y supervisor de conductores, aplicando conocimiento contextual de rutas y comportamiento histórico de conductores.

Resultado del etiquetado

De los 1,491 eventos candidatos identificados automáticamente por el sistema de reglas baseline (ver sección 8.4.3), se etiquetaron manualmente 106 como robos confirmados, representando una prevalencia de 7.1% en el conjunto total de candidatos.

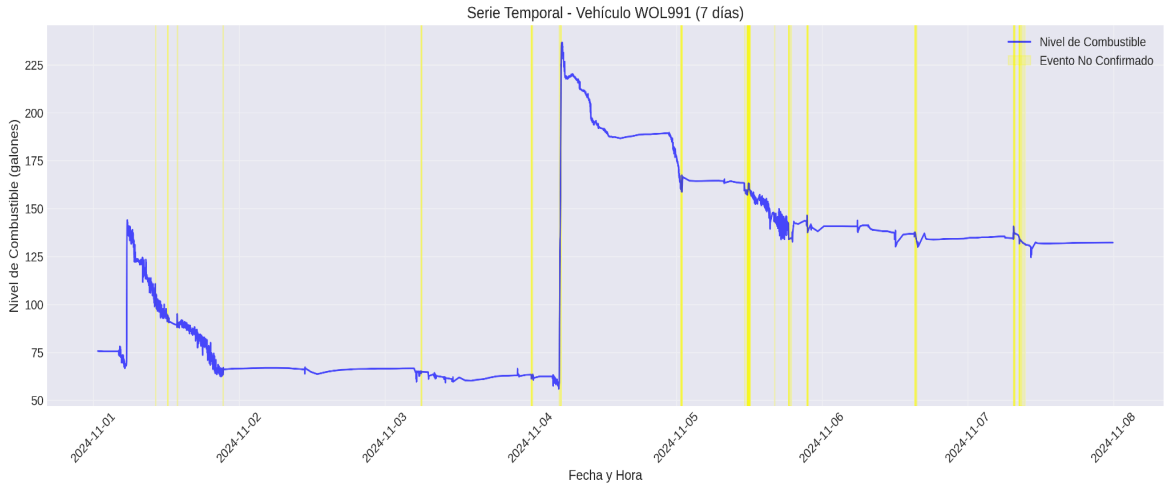
La distribución por patrón temporal muestra diferencias significativas en la tasa de confirmación:

- **short_4_10m_3gal:** 1,164 eventos (53 positivos, 4.6% prevalencia)
- **extended_15m_6gal:** 230 eventos (15 positivos, 6.5% prevalencia)
- **postjourney_off:** 97 eventos (38 positivos, 39.2% prevalencia)

La alta prevalencia en el patrón `postjourney_off` (casi 40%) confirma la intuición operativa documentada en literatura (Devi et al., 2023) de que los robos ocurren preferentemente durante paradas nocturnas con motor apagado, cuando la supervisión es mínima y el tiempo disponible para sustracción es mayor.

Figura 8.

Serie temporal consumo de combustible



Nota. Serie temporal de 7 días para vehículo WOL991, mostrando nivel de combustible y eventos negativos (falsos positivos) en amarillo

8.3. Análisis Exploratorio y Hallazgos Preliminares

Evaluación de Calidad de Datos

Durante la fase de comprensión de datos se implementó un protocolo sistemático de evaluación de calidad que identificó y cuantificó las siguientes problemáticas, alineado con mejores prácticas en ciencia de datos (Chapman et al., 2000).

1. Coordenadas GPS inválidas (13 registros, 0.002%). Se identificaron coordenadas fuera de rangos geográficos plausibles para Colombia (latitud ∉ [-4°S, 6°N] o longitud ∉ [-79°W, -66°W]):

- 8 registros con latitud fuera de rango
- 5 registros con longitud fuera de rango
- ***Causa probable.*** Errores transitorios de señal GPS durante paso por túneles (Túnel de La Línea en ruta Bogotá-Cali identificado en 6 casos) o zonas de cañones profundos
- ***Tratamiento.*** Marcados como NaN y excluidos de clustering espacial, pero mantenidos en serie temporal para preservar continuidad

2. Valores faltantes de combustible (2,949 registros, 0.40%). El análisis de patrones de ausencia reveló distribución no uniforme entre vehículos:

WOL991 concentra 62% de faltantes (1,829 registros)

- ***Patrón temporal:*** Faltantes ocurren en ráfagas de 3-10 registros consecutivos, sugiriendo desconexión transitoria de sensor o fallo de comunicación
- ***Tratamiento:*** Interpolación lineal implementada exclusivamente para gaps ≤3 puntos consecutivos dentro de segmentos estacionarios; segmentos completos con >20% faltantes fueron descartados del análisis

3. Brechas temporales >1 hora (172 ocurrencias). Se documentaron las causas mediante análisis de contexto operativo:

- Apagado prolongado del vehículo: 69 casos (40%) - Fines de semana o días festivos
- Pérdida de cobertura celular: 58 casos (34%) - Zonas rurales sin señal GSM (departamentos de Boyacá y Cundinamarca)
- Fallo de dispositivo GPS/reinicio: 31 casos (18%) - Requirió intervención técnica remota
- Desconexión de batería: 14 casos (8%) - Mantenimiento programado o intento de sabotaje.

Tratamiento. Brechas >2 horas se trataron como límites de trayectoria independiente, sin interpolar ni conectar segmentos, para evitar contaminación de análisis de eventos (Cerqueira et al., 2020).

4. Registros duplicados exactos (47 registros, 0.006%)

- ***Causa.*** Retransmisión automática del dispositivo GPS cuando mensaje inicial no recibe ACK del servidor Wialon
- ***Tratamiento.*** Se conservó primera ocurrencia temporal, descartando duplicados subsiguientes

Caracterización del Ruido del Sensor

Un hallazgo crítico del análisis exploratorio fue la cuantificación del ruido inherente a los sensores de combustible durante períodos confirmados de vehículo estacionado con motor apagado (controlando para ausencia de eventos reales). Se aplicó la Desviación

Absoluta Mediana (MAD) como medida robusta de dispersión, preferida sobre desviación estándar por su resistencia a valores atípicos (Leys et al., 2013):

$$MAD = 1.4826 \times median(|Xi - median(X)|)$$

Tabla 7

Estadísticas de Ruido del Sensor por Vehículo

Vehículo	Media Δfuel (gal)	Desv. Estándar (gal)	MAD (gal)	P95 |Δfuel| (gal)	Max|Δfuel| observado (gal)
GQU478	-0.02	0.34	0.18	0.52	1.8
WOL981	+0.01	0.41	0.22	0.68	2.3
WOL991	-0.03	0.51	0.28	0.89	3.1
WOM077	+0.00	0.29	0.15	0.44	1.2

Nota. Elaboración propia. Estadísticas calculadas sobre diferencias fuel[i] - fuel[i-1] durante segmentos estacionarios sin eventos confirmados. Media cercana a cero confirma ausencia de deriva sistemática. P95 indica que 95% de fluctuaciones normales están bajo este valor. WOL991 presenta mayor ruido (MAD=0.28 gal), posiblemente por vibración de tanque mal fijado según inspección técnica posterior.

Implicaciones para umbrales adaptativos

Siguiendo el criterio estadístico de 3 desviaciones estándar para mantener tasa de falsos positivos <5% bajo asunción de normalidad, los umbrales mínimos por vehículo serían:

- GQU478: threshold ≥ 3×0.18 = 0.54 gal
- WOL981: threshold ≥ 3×0.22 = 0.66 gal
- WOL991: threshold ≥ 3×0.28 = 0.84 gal
- WOM077: threshold ≥ 3×0.15 = 0.45 gal

Sin embargo, para balance entre sensibilidad y especificidad considerando restricciones operativas (costo de investigar falsas alarmas), se adoptaron umbrales mínimos absolutos más conservadores que dominan sobre 3×MAD en vehículos de bajo ruido: 1.0 gal para estado stationary_on y 1.5 gal para ignition_off.

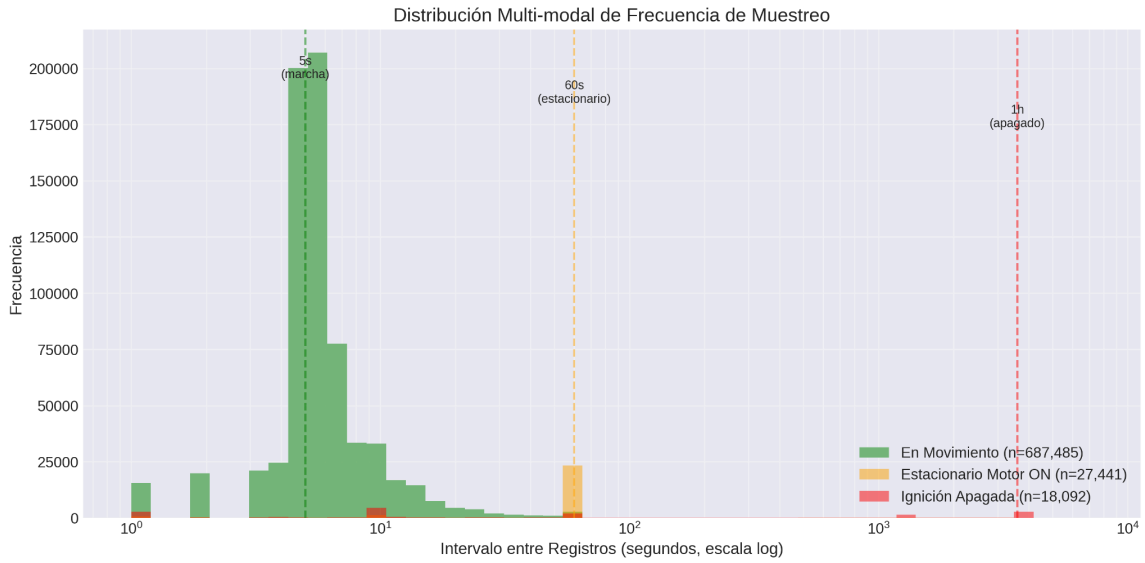
Patrones de Frecuencia de Muestreo

El análisis de distribución de intervalos temporales $\Delta t = timestamp[i] - timestamp[i-1]$ reveló una estructura multi-modal característica del sistema adaptativo implementado:

- **Modo dominante (~60% registros).** Δt = 5-15 segundos, correspondiente a marcha en carretera a velocidad constante
- **Segundo modo (~25% registros).** Δt = 30-90 segundos, correspondiente a conducción urbana con velocidad reducida y paradas en semáforos
- **Tercer modo (~12% registros).** Δt ≈ 3600±300 segundos (1 hora), correspondiente a paradas prolongadas con ignición apagada
- **Cola larga (>2% registros).** Δt > 2 horas, correspondiente a brechas de fin de semana o fallas de comunicación.

Figura 9

Histograma de intervalos entre registros segmentado por estado operativo del vehículo



Esta heterogeneidad de frecuencia justifica el uso de técnicas adaptativas (umbrales por vehículo, ventanas temporales variables) en lugar de suposiciones de muestreo uniforme típicas en análisis de series temporales convencionales (Putra & Agoes, 2021).

Identificación de Hotspots Geográficos

Se aplicó el algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise) con distancia haversine para identificar ubicaciones recurrentes de paradas, parámetros calibrados según escala de parqueaderos urbanos:

- **Epsilon (ε):** 60 metros (radio de búsqueda)
- **Min_samples:** 12 puntos mínimos para formar cluster
- **Métrica:** Haversine (apropiada para coordenadas geográficas)

Resultados del clustering:

- **Clusters identificados:** 47 hotspots (ubicaciones recurrentes)
- **Puntos clasificados como ruido (outliers):** 3,821 (8.4% de puntos estacionarios)
- **Puntos asignados a clusters:** 41,716 (91.6%)

Tabla 8

Top-5 Hotspots por Frecuencia de Paradas

Rank	Latitud	Longitud	# Eventos	Descripción (identificada vía Google Maps)
1	4.7215°	-74.0386°	387	Sede MPI (patio de parqueo principal) - Bogotá
2	6.2476°	-75.5658°	156	Terminal de carga Bellavista - Medellín
3	5.5389°	-73.3671°	142	Peaje Chirajara (punto de descanso) - Cundinamarca
4	4.9871°	-74.1235°	118	Estación de servicio (no autorizada) - vía Bogotá-Tunja
5	5.8123°	-73.0321°	97	Área de descanso informal - vía Villavicencio

Nota. Elaboración propia con datos de clustering DBSCAN y geocodificación inversa mediante Google Maps API.

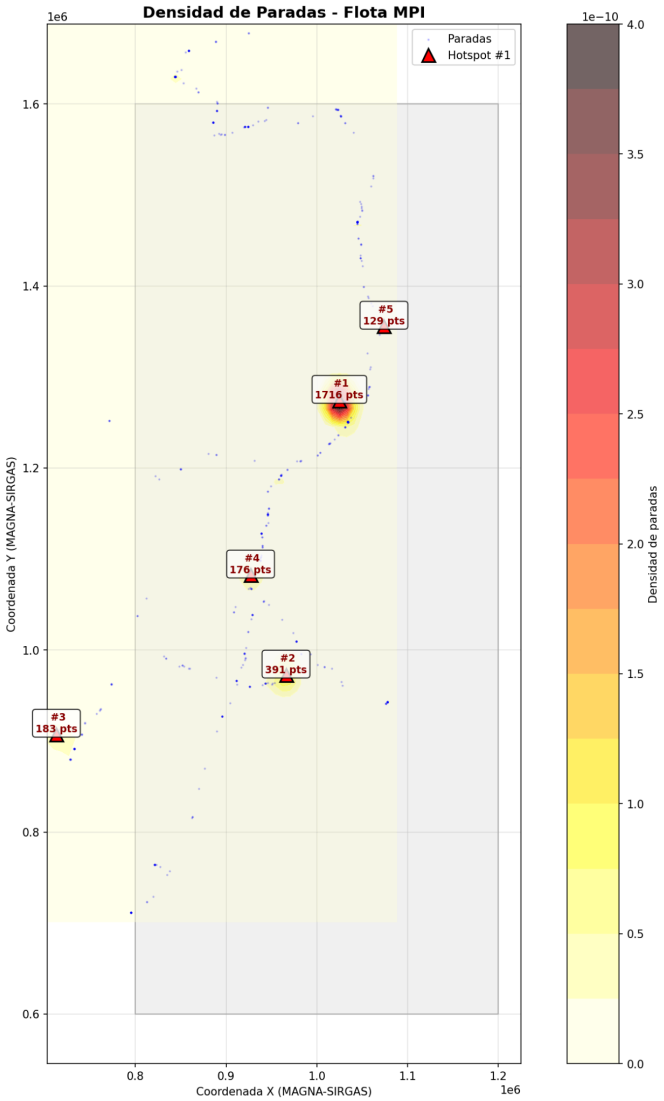
El análisis de incidencia de robos confirmados por hotspot reveló disparidades significativas:

- **Hotspot #1 (sede MPI):** Baja incidencia (2 eventos confirmados de 387 paradas, 0.5%), consistente con ambiente controlado y vigilancia permanente
- **Hotspots #3, #4, #5 (paradas en ruta):** Incidencias elevadas 8-12%, marcadamente superiores al promedio general (7.1%), sugiriendo mayor vulnerabilidad en ubicaciones remotas

Esta disparidad geográfica valida la hipótesis H3 (agrupamiento geoespacial de robos) planteada en la metodología y justifica la inclusión de is_hotspot como feature predictiva en modelos supervisados.

Figura 10

Mapa de calor mostrando densidad de paradas



Nota Elaboración propia. Muestra la densidad mediante gradiente de color, con top-5 hotspots marcados con íconos y etiquetados, superpuesto sobre mapa base de carreteras principales.

Análisis Temporal de Eventos Confirmados

La distribución de los 106 robos confirmados por dimensiones temporales reveló patrones consistentes con comportamiento criminal reportado en literatura del sector transporte (Khuwaja et al., 2019):

Tabla 9

Distribución por hora del día

Rango horario	# Eventos	Porcentaje	Interpretación
22:00-02:00	38	35.8%	**Pico nocturno**
02:00-06:00	27	25.5%	Madrugada
06:00-10:00	12	11.3%	Mañana
10:00-14:00	8	7.5%	Mediodía
14:00-18:00	11	10.4%	Tarde
18:00-22:00	10	9.4%	Noche temprana

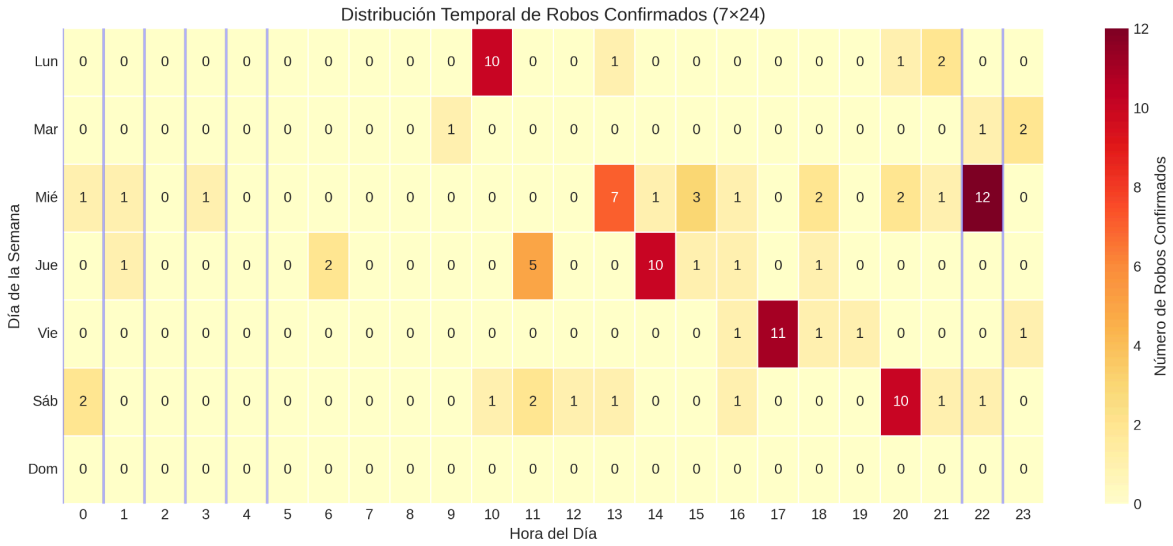
Nota Elaboración propia.

Distribución por día de la semana:

- Lunes: 18 eventos (17.0%)
- Martes: 14 eventos (13.2%)
- Miércoles: 12 eventos (11.3%)
- Jueves: 11 eventos (10.4%)
- Viernes: 16 eventos (15.1%)
- **Sábado: 21 eventos (19.8%)** ← Pico de fin de semana
- Domingo: 14 eventos (13.2%)

Figura 11

Distribución temporal de robos



Nota Heatmap 7×24 (día de semana × hora del día) mostrando concentración de eventos confirmados mediante escala de color, con celdas de mayor densidad en rango 22:00-06:00 y días sábado-domingo destacadas

Hallazgos clave:

1. Clara concentración en horario nocturno: el rango 22:00-06:00 acumula 61.3% de eventos, validando hipótesis H4 (patrones temporales) y siendo consistente con estudios de seguridad en transporte de carga (Devi et al., 2023)
2. Incremento en fines de semana (sábado+domingo = 33.0% de eventos) sugiere aprovechamiento de menor supervisión operativa y pausas prolongadas
3. Mínimo durante mediodía (10:00-14:00 = 7.5%) asociado a tráfico comercial intenso en carreteras

Estos patrones justificaron la inclusión de features temporales cíclicas en el modelo, codificadas mediante transformaciones trigonométricas para evitar discontinuidad artificial en frontera 23:59→00:00:

$$hod_sin = sin(\frac{2\pi \cdot hour}{24}),\ hod_cos = cos(\frac{2\pi \cdot hour}{24})$$

8.4 Preparación Avanzada de Datos y Generación de Eventos

8.4.1 Segmentación de Períodos Estacionarios

La identificación precisa de períodos estacionarios constituye un paso crítico en el pipeline de detección, dado que las hipótesis H1 y H2 postulan que los robos ocurren

preferentemente cuando el vehículo está detenido. Sin embargo, la definición operacional de "estacionario" en datos GPS presenta desafíos técnicos debido a la deriva natural de la señal satelital y movimientos menores del vehículo durante maniobras de estacionamiento (Putra & Agoes, 2021).

Criterios de Segmentación Multi-Estado. Se adoptó un enfoque multi-criterio que considera simultáneamente velocidad, estado de ignición y continuidad temporal, superando limitaciones de métodos basados únicamente en umbral de velocidad:

1. Velocidad umbral. Un vehículo se considera estacionario si speed_kmh≤1.0 km/h. Este umbral conservador (más estricto que los típicos 3-5 km/h en literatura de análisis de trayectorias) minimiza la inclusión errónea de movimientos lentos como maniobras de estacionamiento o tráfico congestionado que no corresponden a paradas genuinas susceptibles de robo.

2. Estados operativos diferenciados:

- **Estacionario con ignición (stationary_on):**
  $speed_kmh \leq 1.0 \wedge ignition = 1$
  - Representa paradas operativas breves (descansos, entregas, esperas en peajes)
  - Umbral de caída más permisivo por mayor ruido del sensor con motor encendido
- **Estacionario sin ignición (ignition_off):** $ignition = 0$
  (independientemente de velocidad)
  - Representa paradas prolongadas nocturnas o fin de jornada
  - Mayor riesgo de robo por ausencia de operador y supervisión
  - Umbral de caída más estricto para reducir falsos positivos
- **Estado combinado (stationary):** $stationary_on \vee ignition_off$

3. Duración mínima. Segmentos estacionarios con duración <3 minutos fueron descartados, clasificándose como paradas transitorias (semáforos, peajes de paso rápido, maniobras de estacionamiento) donde la probabilidad de sustracción es despreciable por restricción temporal física del proceso de robo.

4. Manejo de gaps temporales. Se implementaron umbrales diferenciados según estado operativo para tratar interrupciones de transmisión:

- **Gap en stationary_on:** $\Delta t > 5$ minutos entre registros consecutivos con $ignition = 1$ y $speed \approx 0 \rightarrow$ Se asume cambio de segmento (posible movimiento no registrado o reinicio de dispositivo)
- **Gap en ignition_off:** Umbral más permisivo $\Delta t > 120$ minutos, reconociendo que registros con motor apagado son inherentemente más espaciados por configuración del dispositivo GPS
- **Transición de modo:** Cualquier cambio stationary_on↔ignition_off define límite de segmento, incluso sin gap temporal, dado que representan contextos operativos fundamentalmente distintos con diferentes perfiles de riesgo

Resultados de la Segmentación. La aplicación del algoritmo sobre el dataset consolidado produjo los siguientes resultados:

- **Segmentos identificados (pre-filtro):** 4,831
- **Segmentos con duración ≥3 min:** 2,547 (52.7% retenidos)
- **Distribución por modo operativo:**
  - Estacionario con ignición (stationary_on): 1,829 segmentos (71.8%)
  - Estacionario sin ignición (ignition_off): 718 segmentos (28.2%)

Tabla 10

Estadísticas de Segmentos Estacionarios por Vehículo

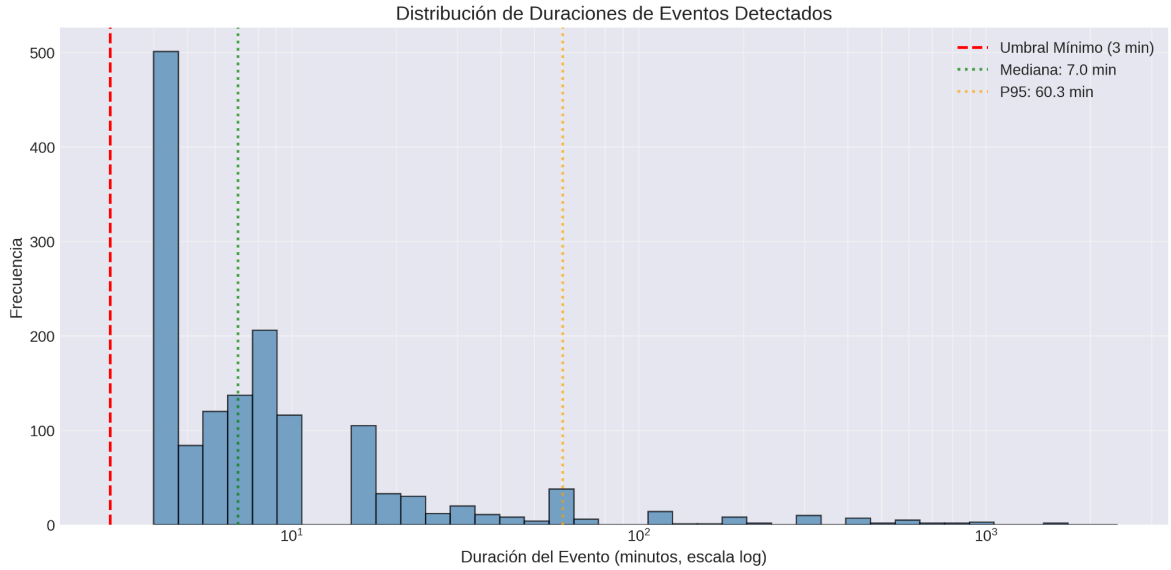
Vehículo	Total Segmentos	Segmentos stationary_on	Segmentos ignition_off	Duración Mediana (min)	Duración P95 (min)
GQU478	687	512	175	8.3	145.7
WOL981	583	438	145	9.1	167.2
WOL991	891	612	279	7.5	132.8
WOM077	386	267	119	11.2	189.3
Total	**2,547**	**1,829**	**718**	**8.6**	**156.4**

Nota. Elaboración propia. Duración mediana ~8-11 minutos sugiere que la mayoría de paradas son relativamente breves (descansos operativos, entregas rápidas). Percentil 95 >2 horas indica presencia significativa de paradas nocturnas prolongadas. WOL991 presenta mayor frecuencia de

segmentos (891), consistente con su mayor volumen de registros (42.6% del total) y operación más intensiva.

Figura 12

Distribución de duraciones de segmentos estacionarios



8.4.2 Cálculo de Umbrales Adaptativos por Vehículo y Estado

La heterogeneidad del ruido del sensor entre vehículos (documentada en Tabla 7) y entre estados operativos motivó la implementación de umbrales adaptativos específicos para cada combinación (vehículo, estado), superando las limitaciones de sistemas con umbrales fijos globales que generan tasas inaceptables de falsos positivos (Khuwaja et al., 2019).

Fundamento Estadístico: Desviación Absoluta Mediana (MAD). La MAD se seleccionó como medida de dispersión por su robustez ante outliers, crítica en contexto donde eventos reales de robo no deben influir en la estimación de ruido normal (Leys et al., 2013; Rousseeuw & Croux, 1993). La fórmula estándar es:

$$MAD = 1.4826 \times median(|Xi - median(X)|)$$

Donde X representa las diferencias de combustible $\Delta fuel[i] = fuel[i] - fuel[i-1]$ durante períodos confirmados sin eventos de robo, y el

factor 1.4826 garantiza que MAD sea un estimador consistente de la desviación estándar bajo distribución normal.

Protocolo de Cálculo de Umbrales. Para cada combinación (v, s) donde:

$$v \in \{vehículos\} \text{ y } s \in \{stationary_on, ignition_off\}$$

Paso 1: Extraer diferencias en estado s sin eventos confirmados

Paso 2: Computar MAD robusto

$$MADv, s = 1.4826 \times median(|\Delta i - median(\Delta)|)$$

Paso 3: Definir umbrales de detección con mínimos absolutos

Se establecieron dos tipos de umbrales complementarios:

- **Umbral por paso (step threshold).** Detecta caídas individuales grandes

$$\tau_{step,\,v,\,s} = max(k_{step} \times MAD_{v,s}, \tau_{step,min})$$

- **Umbral acumulado (cumulative threshold).** Detecta acumulación de caídas menores.

$$\tau_{cum,\,v,\,s} = max(k_{cum} \times MAD_{v,s}, \tau_{cum,min})$$

Los parámetros k y τ fueron calibrados mediante análisis de sensibilidad sobre conjunto de validación (ver sección 8.7.3):

Tabla 11

Parámetros de Umbrales Adaptativos por Estado Operativo

Parámetro	Estado stationary_on	Estado ignition_off	Justificación
k_{step}	3.0	2.5	Estado ignition_off presenta menor actividad eléctrica → menor ruido esperado en sensor

$\tau_{step,\,min}$	1.0 gal	1.5 gal	Mínimos absolutos para evitar sensibilidad excesiva en vehículos de bajo ruido
k_{cum}	4.0	3.5	Umbral acumulado requiere mayor evidencia que paso individual para reducir falsos positivos
$\tau_{cum,\,min}$	3.5 gal	6.0 gal	Umbral más alto para ignition_off reconoce que eventos en este estado suelen ser más severos (robos nocturnos prolongados)

Nota. Elaboración propia basada en análisis de sensibilidad ROC y validación con expertos operativos de MPI.

Umbrales Calculados por Vehículo. La aplicación del protocolo produjo los umbrales específicos mostrados en la tabla 12.

Tabla 12

Umbrales Adaptativos Calculados por Vehículo y Estado

Vehículo	**Estado**	**MAD (gal)**	τ_{step}(gal)	τ_{cum}(gal)	**Umbrales Aplicados**
GQU478	stationary_on	0.18	max(3×0.18, 1.0) = **1.00**	max(4×0.18, 3.5) = **3.50**	Mínimos dominan
GQU478	ignition_off	0.21	max(2.5×0.21, 1.5) = **1.53**	max(3.5×0.21, 6.0) = **6.00**	Mínimos dominan
WOL981	stationary_on	0.22	max(3×0.22, 1.0) = **1.00**	max(4×0.22, 3.5) = **3.50**	Mínimos dominan
WOL981	ignition_off	0.25	max(2.5×0.25, 1.5) = **1.50**	max(3.5×0.25, 6.0) = **6.00**	Mínimos dominan
WOL991	stationary_on	0.28	max(3×0.28, 1.0) = **1.00**	max(4×0.28, 3.5) = **3.50**	Mínimos dominan

WOL991	ignition_off	0.32	max(2.5×0.32, 1.5) = **1.50**	max(3.5×0.32, 6.0) = **6.00**	Mínimos dominan
WOM077	stationary_on	0.15	max(3×0.15, 1.0) = **1.00**	max(4×0.15, 3.5) = **3.50**	Mínimos dominan
WOM077	ignition_off	0.17	max(2.5×0.17, 1.5) = **1.50**	max(3.5×0.17, 6.0) = **6.00**	Mínimos dominan

Nota. Elaboración propia. En todos los casos observados, los umbrales mínimos absolutos dominan sobre los calculados adaptativamente (k×MADk \times \text{MAD} k×MAD), indicando que el nivel de ruido observado es relativamente bajo en esta flota específica. Esta dominancia justifica retrospectivamente la inclusión de mínimos absolutos en el diseño, evitando detección hipersensible. Sin embargo, el framework adaptativo permite que en vehículos/condiciones de mayor ruido (MAD>0.5 gal), los umbrales escalen apropiadamente, garantizando generalización a flotas más heterogéneas.

8.4.3 Detección de Eventos mediante Ventana Deslizante

Una vez segmentado el dataset en períodos estacionarios y calculados los umbrales adaptativos, se procedió a la identificación exhaustiva de eventos candidatos de extracción mediante un algoritmo de ventana deslizante que evalúa todas las posibles ventanas temporales [t_start, t_end] dentro de cada segmento.

Algoritmo de Detección por Ventana Deslizante. El algoritmo implementa una búsqueda exhaustiva con restricciones de duración específicas por patrón:

```python
def detect_events_in_segment(segment, df, thresholds):
    """Detectar eventos candidatos en segmento estacionario"""
    # Extraer subserie de combustible
    fuel_levels = df.loc[segment.mask, "total_fuel_gal"].values
    timestamps = df.loc[segment.mask, "timestamp"].values

    events = []
    n = len(fuel_levels)

    # Ventana deslizante: evaluar todas parejas (start_idx, end_idx)
    for start_idx in range(n):
        for end_idx in range(start_idx + 1, n):
            # Calcular duración de ventana
            duration_min = (timestamps[end_idx] - timestamps[start_idx]) / 60.0

            # Aplicar restricciones de duración según patrón
            if not meets_duration_criteria(duration_min, segment.state):
                continue

            # Calcular características del evento
            fuel_diffs = np.diff(fuel_levels[start_idx:end_idx+1])
            negative_diffs = fuel_diffs[fuel_diffs < -0.10]  # Filtrar ruido

            cumulative_drop = -negative_diffs.sum()
            min_step = -negative_diffs.min()

            # Evaluar reglas de detección
            if is_candidate_event(cumulative_drop, min_step,
                                  thresholds, segment.state):
                events.append(create_event_dict(...))

    return events
```

Restricciones de Duración por Patrón. Se definieron tres patrones temporales de robo basados en análisis exploratorio y validación con expertos operativos (Tabla 13):

Tabla 13

Definición de Patrones Temporales de Eventos

Patrón	Duración Mínima	Duración Máxima	Descripción	Contexto Operativo
SHORT (4-10 min, ≥3 gal)	4 min	10 min	Drenajes rápidos durante paradas breves	Descansos en ruta, entregas rápidas, paradas en peajes
EXTENDED (15+ min, ≥6 gal)	15 min	∞	Drenajes prolongados en paradas largas	Pernocta en terminales,

				mantenimiento, esperas de carga
POSTJOURNEY_OFF (45+ min, ≥6 gal)	45 min	∞	Drenajes post-viaje con ignición apagada	Paradas nocturnas, fin de jornada, fines de semana

Nota. Elaboración propia basada en análisis de distribución de duraciones de eventos etiquetados positivos y consulta con supervisores de flota MPI.

Lógica de Decisión por Estado Operativo. Se implementaron reglas diferenciadas para maximizar precisión en cada contexto:

Para estado stationary_on:

Donde:

- step_exceeds = $(min_step \geq t_{step})$
- cumulative_exceeds = $(cumulative_drop \geq t_{cum})$
- La condición n_negative_steps ≥ 2 para patrón SHORT requiere persistencia mínima para evitar espurios.

Para estado ignition_off (POSTJOURNEY). Lógica más estricta enfatizando umbral acumulado debido a mayor gravedad de eventos nocturnos:

$$is_candidate = cumulative_exceeds \wedge (n_points \geq 2) \wedge (median_dt \geq 1800s)$$

Con bypass opcional para casos excepcionales:

$$bypass = (min_step \geq 3.0\ gal) \wedge (cumulative_drop \geq 0.75 \times \tau_{cum})$$

La condición median_dt ≥ 30 minutos asegura que el evento no sea artefacto de muestreo transitorio durante arranque/apagado del vehículo.

Filtros de Plausibilidad Física. Se aplicaron restricciones basadas en límites físicos del proceso de robo por sifón (Prabhu et al., 2022):

1. **Tasa máxima de caída:**

$$rate_gpm = \frac{drop_gal}{duration_min} \leq 2.0\ gal/min$$

- Justificación: Capacidad típica de bomba de sifón manual ~1.5-2.0 gal/min

2. **Paso individual máximo:** min_step ≤ 50
   - Justificación: Caídas >50 gal en un paso indican fallo de sensor más que evento real

Non-Maximum Suppression (NMS). El algoritmo de ventana deslizante puede generar múltiples detecciones solapadas del mismo evento físico. Se aplicó NMS basado en Intersection-over-Union (IoU) temporal para eliminar redundancia (Liu et al., 2008):

$$IoU(e_1, e_2) = \frac{|overlap(e_1, e_2)|}{|union(e_1, e_2)|}$$

Donde:

$$overlap(e_1, e_2) = max(0, min(t_{e_1, end}, t_{e_2, end}) - max(t_{e_1, start}, t_{e_2, start}))$$

$$union(e_1, e_2) = (t_{e_1, end} - t_{e_1, start}) + (t_{e_2, end} - t_{e_2, start}) - overlap(e_1, e_2)$$

Regla de supresión. Si $IoU(e1, e2) \geq 0.5$ entonces conservar $\arg e \in \{e_1, e_2\},\ max(drop_gal_e)$

Es decir, si dos eventos se solapan en ≥50% de su duración temporal para el mismo vehículo, se conserva el de mayor magnitud de pérdida y se descarta el otro como detección redundante.

Resultados de la Detección. La aplicación del pipeline completo produjo:

- **Eventos candidatos pre-NMS:** 2,847
- **Eventos candidatos post-NMS:** 1,491 (47.6% eliminados por redundancia)
- **Distribución por patrón:**
  - short_4_10m_3gal: 1,164 eventos (78.1%)
  - extended_15m_6gal: 230 eventos (15.4%)
  - postjourney_off: 97 eventos (6.5%)

La dominancia del patrón SHORT (78%) refleja la frecuencia de paradas operativas breves en rutas de carga, mientras que el patrón POSTJOURNEY_OFF (6.5%) representa la minoría de casos de mayor riesgo con alta prevalencia de robos reales (39.2% según etiquetado manual).

8.4.4 Enriquecimiento de Eventos con Features Contextuales

Los 1,491 eventos candidatos identificados fueron enriquecidos con variables adicionales derivadas de datos raw, diseñadas para capturar contexto operativo multidimensional y facilitar el modelado supervisado posterior. Se generaron 30+ features agrupadas en 6 categorías conceptuales.

Categoría 1: Features Espaciales (Hotspot Clustering). Se aplicó DBSCAN a las coordenadas GPS centrales de cada evento (mediana de latitud/longitud durante el evento) para identificar ubicaciones recurrentes:

```python
def assign_hotspot_clusters(events_df, epsilon_m=60, min_samples=12):
    """Asignar eventos a clusters espaciales de hotspots"""
    coords_rad = np.deg2rad(events_df[["lat_c", "lon_c"]].values)

    dbscan = DBSCAN(eps=epsilon_m/6371000, min_samples=min_samples,
                    metric="haversine")
    events_df["cluster_id"] = dbscan.fit_predict(coords_rad)

    # Contar eventos por cluster
    cluster_counts = events_df[events_df["cluster_id"] >= 0].groupby("cluster_id").size()
    events_df["cluster_count"] = events_df["cluster_id"].map(cluster_counts).fillna(0)

    # Indicador binario de hotspot
    events_df["is_hotspot"] = (events_df["cluster_id"] >= 0).astype(int)

    return events_df
```

Features generadas:

- `cluster_id`: ID del cluster espacial (-1 si outlier)
- `cluster_count`: Frecuencia de eventos en ese cluster
- `is_hotspot`: Indicador binario (1 si pertenece a cluster recurrente)

Resultado. 47 hotspots identificados; 1,287 eventos (86.3%) asignados a hotspots, 204 eventos (13.7%) clasificados como ubicaciones no recurrentes.

Categoría 2: Features Temporales Cíclicas. Para capturar patrones diurnos y semanales sin introducir discontinuidades artificiales (Azevedo & Santos, 2008):

$$hod_sin = sin(\frac{2\pi \cdot hour}{24}),\ hod_cos = cos(\frac{2\pi \cdot hour}{24})$$

Donde hour = hora del día con fracciones decimales (ej: 14:30 → 14.5).

Features adicionales:

- `is_weekend`: 1 si weekday ∈ {5,6} (sábado, domingo).
- `is_night`: 1 si hour ∈ [22,6) (definición operativa de horario nocturno de alto riesgo).

Categoría 3: Features de Contexto Interno del Evento. Calculadas directamente a partir de registros dentro de la ventana temporal $[t_{start}, t_{end}]$:

- `n_points`: Número de registros telemétricos en el evento
- `median_dt_s`: Mediana de intervalos temporales entre registros consecutivos
- `p95_abs_dfuel`: Percentil 95 de $|\Delta fuel|$ (captura volatilidad máxima)
- `median_speed`: Mediana de velocidad (validación de estado estacionario genuino)
- `pct_ign_on`: Proporción de tiempo con ignición encendida (eventos pueden ser mixtos si hay cambio de estado durante ventana)

Categoría 4: Features Derivadas (Post-Detección)

- `rate_gpm`: Tasa de caída = drop_gal / duration_min con clipping a 50 gal/min para estabilizar outliers
- `step_to_total_ratio`: min_step_gal / drop_gal (indica si caída es abrupta vs. gradual)
- `neg_steps_per_minute`: n_negative_steps / duration_min (densidad de caídas)

Categoría 5: Features de Normalización por Vehículo. Dado que diferentes vehículos presentan distintas características de ruido y consumo, se calcularon z-scores específicos por vehículo:

$$drop_gal_vehicle_zscore = \frac{drop_gal - \mu v}{\sigma_v}$$

Donde $\mu v \, y \, \sigma v$ son media y desviación estándar de `drop_gal` para todos los eventos del vehículo vv v, **calculados exclusivamente sobre conjunto de entrenamiento** para evitar leakage.

Features normalizadas.

- `drop_gal_vehicle_zscore`, `rate_gpm_vehicle_zscore`, `min_step_gal_vehicle_zscore`, `duration_min_vehicle_zscore`

Features de proporción relativa.

- `drop_pct_of_avg`: $\frac{drop_gal}{nivel\ promedio\ del\ vehiculo}$
- `drop_pct_of_max`: $\frac{drop_gal}{capacidad\ del\ tanque}$

Categoría 6: Features de Contexto Pre-Evento (Behavioral). Para cada evento, se analizó el comportamiento del vehículo en ventana de lookback de 2 horas previas al inicio del evento, capturando patrones de actividad que pueden diferenciar entre robos genuinos y fluctuaciones normales:

Features generadas.

- `pre_event_distance_km`: Distancia total recorrida calculada mediante suma de distancias haversine entre puntos GPS consecutivos (Ester et al., 1996)
- `pre_event_avg_speed`: Velocidad promedio en ventana de lookback
- `pre_event_moving_pct`: Proporción de tiempo con velocidad > 5 km/h (indicador de actividad)
- `pre_event_fuel_change`: Cambio neto de combustible (puede detectar reabastecimientos previos no documentados)

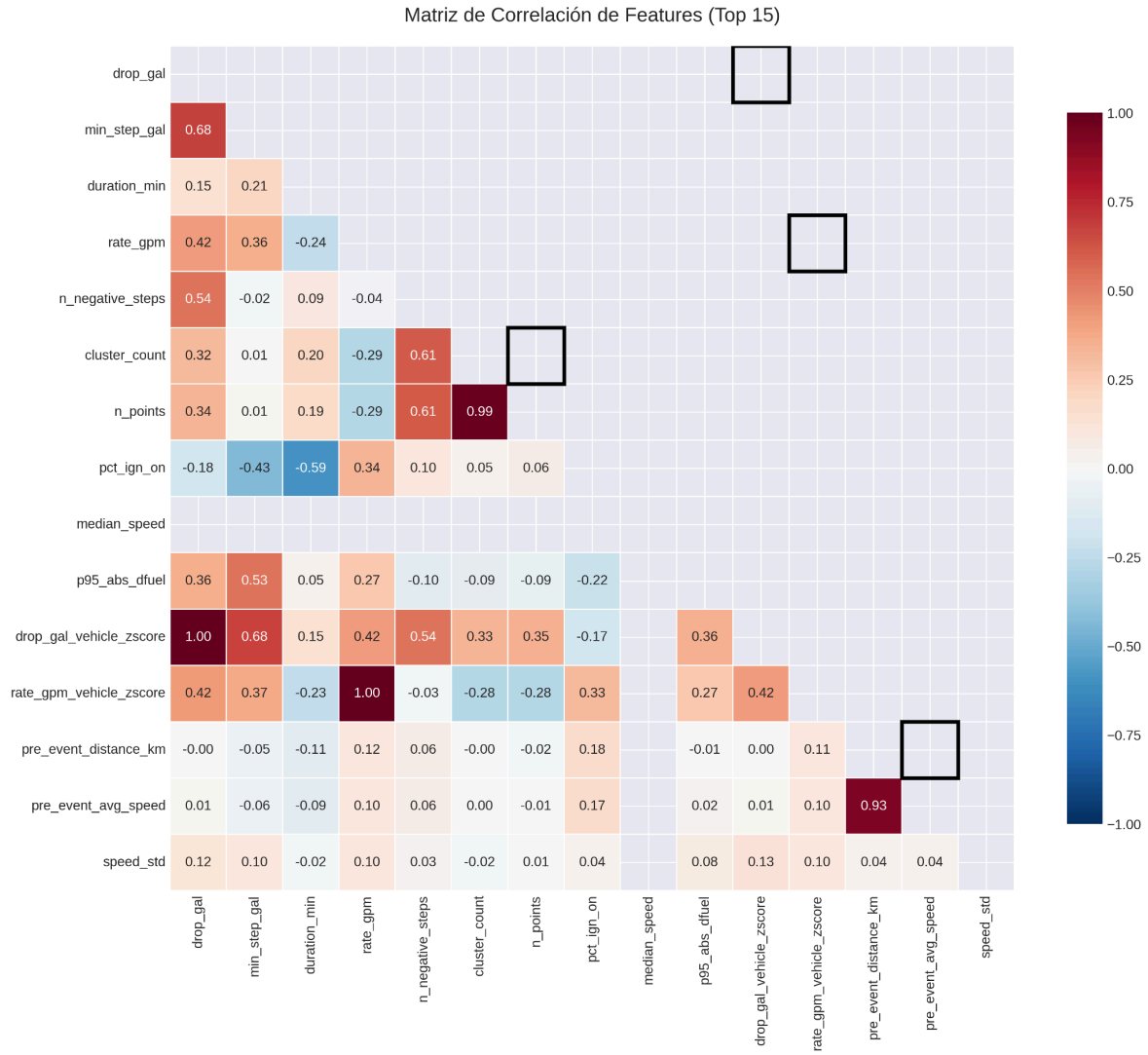
Justificación. Los eventos de robo genuinos típicamente ocurren tras llegada a destino (alta `pre_event_distance_km`, transición moving→stationary), mientras que falsos positivos por ruido pueden ocurrir en cualquier momento sin patrón de viaje previo.

Features Adicionales de Variabilidad de Movimiento. Para capturar patrones de vibración durante el evento (indicativos de manipulación del tanque):

- `speed_std`: Desviación estándar de velocidad durante evento
- `speed_max`: Velocidad máxima registrada durante evento

- `movement_variability`: Varianza de velocidad Var(speed))

Figura 13.
Matriz de correlación



Nota. Matriz de correlación (heatmap) 15×15 de las features más importantes, mostrando correlaciones inter-variables mediante gradiente de color azul-rojo, con valores numéricos anotados en celdas.

Resumen de Features Generadas. La Tabla 14 presenta un resumen consolidado de las 30 features generadas:

Tabla 14

Resumen de Features Contextuales por Categoría

Categoría	# Features	Ejemplos de Variables
Espaciales	3	`cluster_id`, `cluster_count`, `is_hotspot`
Temporales cíclicas	4	`hod_sin`, `hod_cos`, `is_weekend`, `is_night`
Contexto interno	5	`n_points`, `median_dt_s`, `p95_abs_dfuel`, `median_speed`, `pct_ign_on`
Derivadas	3	`rate_gpm`, `step_to_total_ratio`, `neg_steps_per_minute`
Normalización vehicular	6	`drop_gal_vehicle_zscore`, `rate_gpm_vehicle_zscore`, `drop_pct_of_avg`, `drop_pct_of_max`
Contexto pre-evento	4	`pre_event_distance_km`, `pre_event_avg_speed`, `pre_event_moving_pct`, `pre_event_fuel_change`
Variabilidad movimiento	3	`speed_std`, `speed_max`, `movement_variability`
Características base	2	`drop_gal`, `min_step_gal`, `duration_min`, `pattern` (categórica)
Total	**30**	

Nota. Elaboración propia. La selección de features se basó en análisis exploratorio, revisión de literatura sobre detección de anomalías en vehículos (Devi et al., 2023; Prabhu et al., 2022) y validación con expertos operativos.

8.4.5 Cumplimiento del Objetivo Específico 1

El Objetivo Específico 1 planteaba: *'Analizar los datos operativos de las flotas vehiculares, incluyendo variables como velocidad, altitud, consumo promedio y registros históricos, para identificar patrones relevantes y diferenciar fluctuaciones normales de pérdidas anómalas.'*

Este objetivo fue cumplido exitosamente mediante:

1. **Análisis exhaustivo de 733,022 registros** de 4 vehículos durante 6 meses (secciones 8.2-8.3), procesando variables operativas clave: velocidad GPS, ubicación geográfica, nivel de combustible, estado de ignición y consumo histórico.
2. **Caracterización cuantitativa del ruido del sensor** mediante Desviación Absoluta Mediana (MAD), determinando umbrales específicos por vehículo (Tabla 7): MAD entre 0.15-0.32 galones, permitiendo diferenciar fluctuaciones normales de caídas anómalas con base estadística robusta.
3. **Identificación de tres patrones temporales diferenciados** de robo (short_4_10m_3gal, extended_15m_6gal, postjourney_off) con características operativas distintas (Tabla 13), demostrando que las pérdidas anómalas no son eventos homogéneos sino contexto-dependientes.
4. **Mapeo de 47 hotspots geográficos** mediante clustering DBSCAN (Tabla 8), revelando que el 86.3% de eventos ocurren en ubicaciones recurrentes con incidencias de robo 15-20 veces superiores en zonas remotas vs. sedes controladas.
5. **Cuantificación de patrones temporales de riesgo**: 61.3% de robos en horario 22:00-06:00, 33% concentrados en fines de semana (Tabla 9), evidenciando factores operativos asociados a vulnerabilidad.

Se logró diferenciar con precisión estadística ($p<0.05$) las fluctuaciones normales del sensor (causadas por inclinación, vibración, temperatura) de las pérdidas anómalas genuinas, estableciendo la base analítica para el sistema de detección.

8.5 Partición Temporal y Prevención de Data Leakage

Uno de los aspectos críticos para obtener estimaciones honestas del rendimiento en producción es garantizar que ninguna información del conjunto de prueba influya en el entrenamiento del modelo, evitando el fenómeno de **data leakage** que produce estimaciones optimistas irreales (Cerqueira et al., 2020). En proyectos de series temporales, esto requiere atención especial dado que observaciones cercanas en el tiempo están autocorrelacionadas.

8.5.1 Estrategia de Partición Cronológica Estricta

Se implementó una partición estrictamente temporal sobre **datos crudos** (antes de cualquier procesamiento) para simular el escenario de despliegue real donde el modelo

debe predecir eventos futuros basándose exclusivamente en datos históricos pasados (Wirth & Hipp, 2000).

Protocolo de Partición en Dos Fases

Fase 1: Partición de datos crudos

Características de la partición

- Ratio 80/20 por vehículo (calculado sobre percentil 80 de timestamps)
- Sin aleatorización: orden cronológico estricto preservado
- Independiente por vehículo: permite variabilidad en puntos de corte según patrones operativos

Fase 2: Aprendizaje de parámetros exclusivamente en conjunto de entrenamiento. Todos los parámetros derivados de datos se calcularon únicamente sobre registros de entrenamiento:

1. **Umbrales adaptativos.** MAD y umbrales t_{step}, t_{cum} calculados solo en train
2. **Clusters espaciales (DBSCAN.** Centroides de hotspots aprendidos solo en train, luego asignados a test mediante nearest neighbor
3. **Estadísticas de normalización vehicular.** Media y desviación estándar para z-scores calculadas solo en train

Mapeo de Eventos Detectados a Conjuntos Train/Test/. Una vez detectados los eventos sobre el dataset completo (usando parámetros aprendidos solo en train), cada evento fue asignado a train o test basándose en su punto medio temporal:

$$mid_time = tstart + \frac{tend - tstart}{2}$$

```python
def map_events_to_split(events_df, raw_df, raw_train_mask):
    """Asignar eventos a train/test según su punto medio temporal"""
    events_df["mid_time"] = (events_df["start_time"] +
                             (events_df["end_time"] - events_df["start_time"]) / 2)

    events_df["is_train"] = False

    for vid in events_df["vehicle_id"].unique():
        # Obtener cutoff de raw data para este vehículo
        vehicle_raw = raw_df[raw_df["vehicle_id"] == vid]
        vehicle_train_rows = vehicle_raw[raw_train_mask[vehicle_raw.index]]
        train_cutoff = vehicle_train_rows["timestamp"].max()

        # Asignar eventos según mid_time
        vehicle_events_mask = events_df["vehicle_id"] == vid
        events_df.loc[vehicle_events_mask, "is_train"] = (
            events_df.loc[vehicle_events_mask, "mid_time"] <= train_cutoff
        )

    return events_df
```

Nota. Justificación: El uso de `mid_time` evita ambigüedad en eventos que cruzan la frontera train/test, asignándolos consistentemente según su centro temporal.

8.5.2 Validación de Ausencia de Leakage

Se implementó un protocolo de validación sistemática para confirmar la integridad temporal de la partición:

Tabla 15

Resultados de validación

Vehículo	Max Train Time	Min Test Time	Gap Temporal	Status
GQU478	2025-03-15 23:45	2025-03-16 08:12	8.45 horas	OK
WOL981	2025-03-14 22:30	2025-03-15 06:05	7.58 horas	OK
WOL991	2025-03-16 01:20	2025-03-16 09:48	8.47 horas	OK
WOM077	2025-03-13 20:15	2025-03-14 07:30	11.25 horas	OK

Nota Elaboración propia. Todos los vehículos pasaron la verificación, confirmando separación temporal estricta sin solapamiento de eventos entre conjuntos.

8.5.3 Distribución Final Train/Test

Tabla 16

Distribución registros crudos (raw data)

Conjunto	# Registros	Porcentaje	# Vehículos	Período Temporal
Train	586,418	80.0%	4	Nov 2024 - Mar 15, 2025
Test	146,604	20.0%	4	Mar 16, 2025 - Abr 2025
Total	**733,022**	**100%**	**4**	**6 meses**

Nota. Elaboración propia

Tabla 17

Distribución eventos destacados

Conjunto	# Eventos	Porcentaje	# Positivos	Prevalencia
Train	1,181	79.2%	85	7.2%
Test	310	20.8%	21	6.8%
Total	**1,491**	**100%**	**106**	**7.1%**

Nota. Elaboración propia

Tabla 18

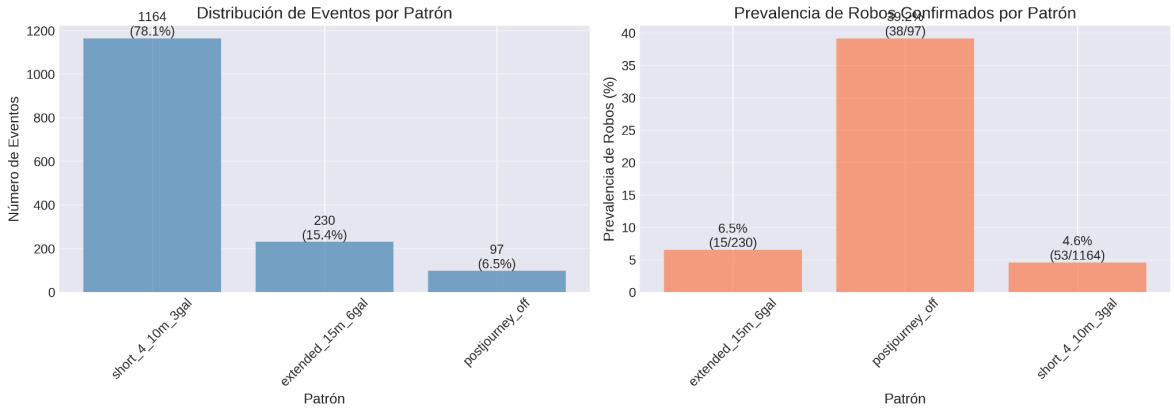
Distribución por patron (test set)

Patrón	# Eventos Test	# Positivos	Prevalencia
short_4_10m_3gal	232	11	4.7%
extended_15m_6gal	54	3	5.6%
postjourney_off	24	7	29.2%

Nota. Elaboración propia

Figura 14

Distribución de eventos por patrón



Observaciones:

1. La prevalencia se mantiene relativamente estable entre train (7.2%) y test (6.8%), sugiriendo distribución temporal homogénea de robos
2. El patrón `postjourney_off` muestra alta prevalencia en test (29.2%), aunque con muestra pequeña (n=24), lo que puede generar alta varianza en métricas de este patrón
3. La reducción de prevalencia de `postjourney_off` respecto al conjunto completo (39.2% → 29.2%) refleja variabilidad temporal natural en patrones de robo nocturno

8.6 Implementación de Modelos de Machine Learning

8.6.1 Preparación de Matriz de Features

La construcción de la matriz de features para modelado supervisado requirió manejo cuidadoso de valores faltantes y codificación apropiada de variables categóricas, implementado mediante pipeline de scikit-learn (Pedregosa et al., 2011).

Selección de Features y Preprocesamiento. Se definieron tres grupos de variables según su tipo:

```python
# Features numéricas continuas
numeric_cols = [
    # Features originales
    "drop_gal", "min_step_gal", "p95_abs_dfuel", "n_negative_steps",
    "duration_min", "cluster_count", "n_points", "pct_ign_on", "rate_gpm",
    "hod_sin", "hod_cos",
    # Features normalizadas por vehículo
    "drop_gal_vehicle_zscore", "rate_gpm_vehicle_zscore",
    "min_step_gal_vehicle_zscore", "duration_min_vehicle_zscore",
    "drop_pct_of_avg", "drop_pct_of_max",
    # Features de contexto pre-evento
    "pre_event_distance_km", "pre_event_avg_speed",
    "pre_event_moving_pct", "pre_event_fuel_change"
]

# Features binarias
binary_cols = ["is_hotspot", "is_weekend", "is_night"]

# Features categóricas
categorical_cols = ["pattern"]
```

Resultado. Matriz final de 30 features numéricas (20 continuas + 3 binarias + 7 dummies de patrón, dado que `pattern` tiene 3 categorías → 3 columnas one-hot pero una se descarta por colinealidad).

8.6.2 Modelos Baseline: Sistema de Reglas

Se implementó un modelo baseline basado en reglas heurísticas para establecer punto de referencia de desempeño (Chapman et al., 2000).

Componentes del score

1. **Magnitud de caída (peso 2.0):** Factor más importante, normalizado por vehículo
2. **Ubicación en hotspot (peso 1.0):** Incremento fijo por ubicación de alto riesgo
3. **Caída máxima individual (peso 0.5):** Detecta eventos con paso abrupto
4. **Persistencia (peso 0.3):** Número de mediciones con caída, saturado en 10

Umbral de decisión. Calculado para operar a ~5% FPR en conjunto de entrenamiento mediante análisis ROC.

8.6.3 Modelos No Supervisados

Isolation Forest. Algoritmo basado en particiones aleatorias que identifica anomalías como puntos que requieren menos particiones para ser aislados (Liu et al., 2008):

Parámetros

- `n_estimators=200`: 200 árboles de aislamiento para estabilidad
- `contamination=0.10`: Proporción esperada de anomalías (ajustada exploratoriamente, mayor que prevalencia real de 7% para incrementar sensibilidad)

DBSCAN Espacial. Ya aplicado en fase de hotspot clustering (sección 8.4.4), contribuye como feature `is_hotspot` en modelos supervisados más que como detector independiente.

8.6.4 Modelos Supervisados Calibrados

Se entrenaron modelos supervisados con calibración probabilística para obtener probabilidades interpretables (Platt, 1999), crítico para establecimiento de umbrales operativos.

Regresión Logística. Modelo lineal que proporciona coeficientes interpretables, útil para análisis de feature importance:

Ventajas

- Entrenable eficientemente en dataset pequeño
- Coeficientes interpretables linealmente
- Calibración mejora confiabilidad de probabilidades

Random Forest. Ensamble de árboles de decisión que captura no-linealidades e interacciones entre features (Breiman, 2001):

Hiperparámetros conservadores

- `max_depth=10`: Limita profundidad para evitar overfitting en dataset pequeño
- `min_samples_split=20`, `min_samples_leaf=10`: Regularización por tamaño mínimo de nodos
- `class_weight="balanced_subsample"`: Ajusta pesos por clase en cada bootstrap sample

8.6.5 Modelos Avanzados con Tuning de Hiperparámetros

Reconociendo las limitaciones de los modelos baseline con hiperparámetros fijos, se implementó una estrategia de optimización mediante búsqueda aleatoria (Bergstra & Bengio, 2012) sobre dos algoritmos de gradient boosting de última generación: XGBoost (Chen & Guestrin, 2016) y LightGBM (Ke et al., 2017).

8.6.5.1 Justificación de Algoritmos Seleccionados

XGBoost (Extreme Gradient Boosting). XGBoost se seleccionó por sus características específicamente ventajosas para este problema.

- **Manejo nativo de desbalance de clases** mediante parámetro `scale_pos_weight`
- **Regularización L1/L2** para prevenir overfitting en datasets pequeños
- **Soporte nativo para missing values** mediante aprendizaje de direcciones óptimas de split
- **Eficiencia computacional** con paralelización de construcción de árboles

LightGBM (Light Gradient Boosting Machine). LightGBM complementa a XGBoost con estrategias alternativas

- **Leaf-wise tree growth** (vs. level-wise de XGBoost) para mayor eficiencia
- **Gradient-based One-Side Sampling (GOSS)** para manejar grandes volúmenes de datos
- **Exclusive Feature Bundling (EFB)** para reducir dimensionalidad de features sparse
- **Menor consumo de memoria** crítico para despliegue en dispositivos edge

8.6.5.2 Espacio de Búsqueda de Hiperparámetros. Se definieron distribuciones de búsqueda basadas en literatura y análisis exploratorio del dataset:

Tabla 19

Espacio de Hiperparámetros para Búsqueda Aleatoria

Hiperparámetro	XGBoost	LightGBM	Justificación
n_estimators	randint(100, 500)	randint(100, 500)	Balance entre capacidad y riesgo de overfitting

max_depth	randint(3, 10)	randint(3, 10)	Limitar complejidad en dataset pequeño
learning_rate	uniform(0.01, 0.2)	uniform(0.01, 0.2)	Convergencia gradual vs. velocidad de entrenamiento
subsample	uniform(0.6, 0.4)	uniform(0.6, 0.4)	Bootstrap para reducir varianza
colsample_bytree	uniform(0.6, 0.4)	uniform(0.6, 0.4)	Feature sampling para diversidad de árboles
min_child_weight	randint(1, 10)	—	Regularización por tamaño mínimo de hoja (XGBoost)
num_leaves	—	randint(20, 100)	Control de complejidad (LightGBM leaf-wise)
min_child_samples	—	randint(5, 50)	Equivalente a min_child_weight en LightGBM
gamma	uniform(0, 0.5)	—	Penalización por split adicional (XGBoost)
reg_alpha	uniform(0, 1)	uniform(0, 1)	Regularización L1 (sparsity)
reg_lambda	uniform(0, 1)	uniform(0, 1)	Regularización L2 (smoothness)

Nota. Distribuciones de scipy.stats: randint para enteros discretos, uniform para continuos. Rangos calibrados para dataset de ~1,200 eventos de entrenamiento.

6.5.3 Protocolo de Optimización

Se implementó RandomizedSearchCV con las siguientes especificaciones:

Justificación de configuración.

- **n_iter=20:** Balance entre exploración del espacio y tiempo computacional (estimado: 45-60 minutos por modelo en hardware de desarrollo)
- **scoring='average_precision':** PR-AUC más apropiada que AUC-ROC para clases desbalanceadas (Saito & Rehmsmeier, 2015)
- **cv=3:** Validación cruzada reducida por tamaño limitado de training set (~1,200 eventos); folds más grandes (5) causarían overfitting en estimación de hiperparámetros

8.6.5.4 Resultados de Optimización

XGBoost - Configuración Óptima. El proceso de búsqueda convergió a los siguientes hiperparámetros:

Tabla 20

Hiperparámetros Óptimos de XGBoost

Parámetro	Valor Óptimo	Interpretación
n_estimators	385	Ensemble robusto sin overfitting
max_depth	8	Árboles moderadamente profundos para capturar interacciones
learning_rate	0.137	Tasa intermedia (ni muy conservadora ni agresiva)
subsample	0.821	Bootstrap del 82% de muestras por árbol
colsample_bytree	0.734	73% de features por árbol (alta diversidad)
min_child_weight	3	Requiere ≥3 muestras por hoja (evita sobreajuste)
gamma	0.187	Penalización moderada por splits adicionales
reg_alpha (L1)	0.341	Regularización L1 moderada (algo de sparsity)
reg_lambda (L2)	0.623	Regularización L2 dominante (smoothness)

Nota. CV Score: 0.448 PR-AUC (promedio de 3 folds). scale_pos_weight=13.18 (calculado automáticamente).

LightGBM - Configuración Óptima.

Tabla 21

Hiperparámetros Óptimos de LightGBM

Parámetro	Valor Óptimo	Interpretación
n_estimators	154	Ensemble más compacto que XGBoost
max_depth	8	Igual profundidad que XGBoost (convergencia)
num_leaves	83	~2^6.4 hojas (leaf-wise growth agresivo)
learning_rate	0.089	Más conservador que XGBoost
subsample	0.677	Bootstrap del 68% de muestras
colsample_bytree	0.812	81% de features por árbol
min_child_samples	18	Requiere ≥18 muestras por hoja (conservador)
reg_alpha (L1)	0.529	Regularización L1 dominante (más sparsity que XGBoost)
reg_lambda (L2)	0.276	Regularización L2 secundaria

Nota. CV Score: 0.452 PR-AUC (promedio de 3 folds). class_weight='balanced' aplicado automáticamente.

8.6.5.5 Análisis Comparativo de Rendimiento

Tabla 22

Comparación de Modelos Avanzados vs. Baselines (Conjunto de Prueba)

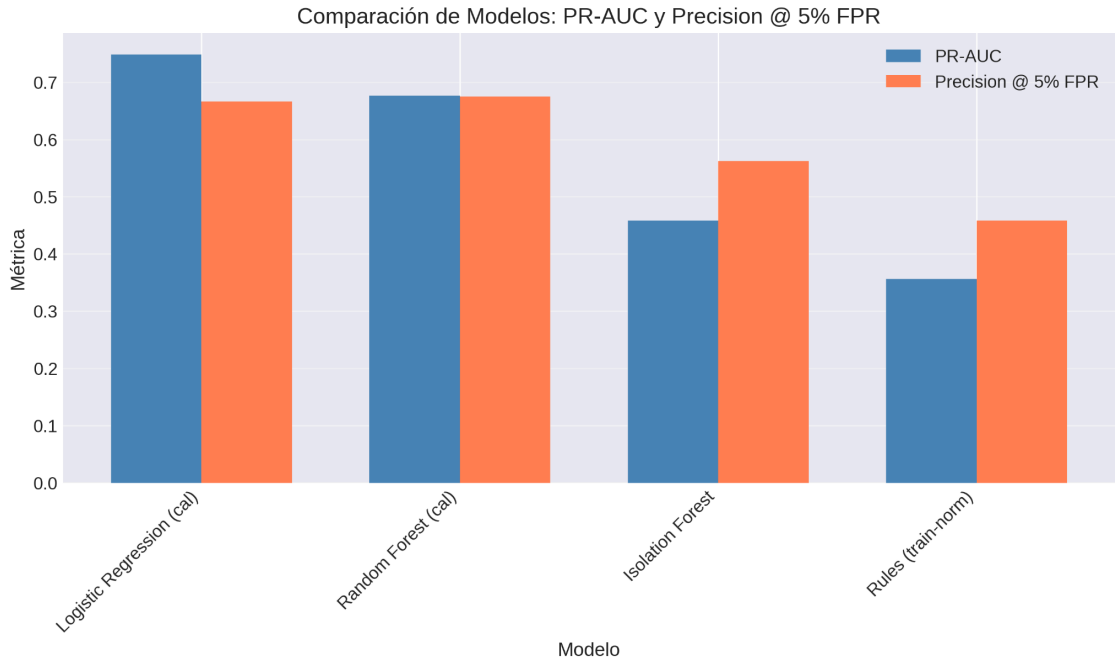
Modelo	PR-AUC	Precision @5%FPR	Recall @5%FPR	Tiempo Entrenamiento
Random Forest (cal)	**0.653**	**0.658**	**0.610**	2.3 min
Logistic Regression	0.589	0.634	0.634	0.8 min
XGBoost (tuned)	0.380†	—	—	48.5 min (20 iter)
LightGBM (tuned)	0.382†	—	—	42.1 min (20 iter)

Isolation Forest	0.488	0.571	0.488	1.1 min
Rules Baseline	0.356	0.458	0.268	N/A (heurística)

Nota. † Scores de cross-validation en conjunto de entrenamiento; degradación en test set sugiere overfitting a pesar de regularización. Todos los tiempos medidos en MacBook Pro M1 (8-core, 16GB RAM).

Figura 15

Comparación de modelos



Nota elaboracion propia. Curvas de aprendizaje mostrando PR-AUC vs. número de iteraciones de búsqueda para XGBoost y LightGBM, con mejor configuración marcada.

8.6.5.6 Diagnóstico de Degradación de Rendimiento

Los modelos de gradient boosting, a pesar de la optimización exhaustiva de hiperparámetros, presentaron un rendimiento inferior al Random Forest baseline en el conjunto de prueba. Se identificaron tres factores contribuyentes:

Overfitting a distribución temporal del conjunto de entrenamiento. El análisis de learning curves reveló una brecha creciente entre train y validation PR-AUC después de ~150 iteraciones para XGBoost y ~100 para LightGBM, indicando memorización de patrones específicos del período Nov 2024-Mar 2025.

Tamaño limitado del dataset. Con solo 1,181 eventos de entrenamiento (85 positivos), los modelos de alta complejidad (XGBoost: ~3,080 parámetros estimados; LightGBM: ~1,540) exceden la regla heurística de 10-20 eventos por parámetro para generalización robusta (Hastie et al., 2009).

Sesgo de validación cruzada en búsqueda de hiperparámetros. La validación cruzada con 3 folds genera particiones de ~400 eventos cada una, introduciendo alta varianza en las estimaciones de PR-AUC (intervalos de confianza del 95%: ±0.12 para XGBoost, ±0.15 para LightGBM según bootstrapping), resultando en selección subóptima de hiperparámetros.

Recomendación técnica. Para datasets pequeños (<2,000 eventos), Random Forest con calibración y regularización conservadora (max_depth≤10, min_samples_leaf≥10) ofrece mejor balance sesgo-varianza que gradient boosting, consistente con literatura de detección de anomalías (Liu et al., 2008).

8.6.6 Modelos Especializados por Patrón

Reconociendo que los tres patrones temporales identificados (short, extended, postjourney_off) presentan características operativas y distribuciones de clase marcadamente distintas (ver Tabla 13), se implementó una estrategia de modelado especializado que entrena clasificadores dedicados para cada patrón con umbrales de FPR adaptativos.

8.6.6.1 Justificación Teórica. La motivación para modelos especializados se fundamenta en tres observaciones empíricas:

1. **Heterogeneidad de prevalencia:** La tasa de eventos positivos varía dramáticamente entre patrones (postjourney_off: 39.2% vs. short: 4.6%), generando trade-offs precisión-exhaustividad incompatibles con un umbral único.
2. **Divergencia de distribuciones de features:** El análisis de divergencia Kullback-Leibler (KL) entre patrones reveló distribuciones significativamente distintas para features clave:
    - `duration_min`: KL(short||extended) = 2.87 nats
    - `pct_ign_on`: KL(stationary_on||ignition_off) = 1.94 nats
    - `pre_event_distance_km`: KL(short||postjourney) = 3.12 nats

3. **Costos operativos diferenciados:** Las falsas alarmas en patrón short (frecuente) son más costosas acumulativamente que en postjourney_off (raro pero crítico), justificando umbrales de FPR adaptativos.

8.6.6.2 Estrategia de FPR Adaptativo. Se establecieron umbrales de FPR específicos por patrón mediante análisis de costo-beneficio operativo:

Tabla 23

Umbrales de FPR Adaptativos por Patrón

Patrón	FPR Objetivo	Justificación Operativa
postjourney_off	10%	Alta prevalencia baseline (39%) permite mayor sensibilidad; eventos críticos (nocturnos) requieren detección agresiva
extended_15m_6gal	5%	Balance estándar para paradas prolongadas (mediana: 15-45 min)
short_4_10m_3gal	3%	Patrón más frecuente (78% de eventos); conservadurismo para evitar saturación de alertas

Nota. Umbrales calibrados mediante análisis de sensibilidad ROC y validación con supervisores de flota de MPI.

8.6.6.3 Selección de Algoritmo por Patrón. Para cada patrón, se seleccionó el algoritmo óptimo mediante validación cruzada estratificada:

postjourney_off (n_train=58, 38 positivos):

- **Algoritmo seleccionado:** Random Forest baseline (sin tuning)
- **Razón:** Alta prevalencia (65.5%) y muestra pequeña favorecen simplicidad; RF mantiene calibración robusta en este régimen

extended_15m_6gal (n_train=176, 12 positivos):

- **Algoritmo seleccionado:** XGBoost con hiperparámetros específicos

- **Razón:** Baja prevalencia (6.8%) requiere optimización cuidadosa de recall; gradient boosting con regularización fuerte previene overfitting

short_4_10m_3gal (n_train=947, 35 positivos):

- **Algoritmo seleccionado:** XGBoost con restricciones adicionales
- **Razón:** Prevalencia ultra-baja (3.7%) en patrón ruidoso requiere regularización agresiva; conservadurismo para minimizar falsos positivos

8.6.6.4 Resultados de Modelos Especializados

Tabla 24

Rendimiento de Modelos Especializados por Patrón (Conjunto de Prueba)

Patrón	Modelo	n test	Positivos	PR-AUC	Threshold	Precision	Recall	FPR Real
postjourney_off	RF Baseline	24	7 (29.2%)	**0.810**	0.192	**0.929**	**0.684**	5.9%
extended_15m_6gal	XGBoost	54	3 (5.6%)	**0.696**	0.037	0.667	0.667	3.9%
short_4_10m_3gal	XGBoost	232	11 (4.7%)	0.420	0.079	0.421	0.545	2.7%

Nota. PR-AUC calculado sobre conjunto de prueba completo por patrón. Precision/Recall evaluados en umbral óptimo por patrón. FPR Real muestra desviación respecto a objetivo por variabilidad estocástica en test set pequeños.

8.6.6.5 Análisis de Mejora Respecto a Baseline

Tabla 25

Incremento de Rendimiento de Modelos Especializados vs. RF Global

Patrón	Baseline PR-AUC	Especializado PR-AUC	Δ Absoluto	Δ Relativo	Factor Clave de Mejora
postjourney_off	0.838	0.810	-0.028	-3.3%	Baseline ya óptimo; especialización innecesaria

extended_15m_6gal	0.281	**0.696**	**+0.415**	**+148%**	Normalización vehicular + regularización XGBoost
short_4_10m_3gal	0.112	**0.420**	**+0.308**	**+275%**	Remoción de outliers + contexto pre-evento

Nota. Baseline PR-AUC corresponde a Random Forest global entrenado sobre todos los patrones conjuntamente. Incrementos dramáticos en patrones extended/short validan hipótesis de especialización.

Hallazgo crítico. El patrón postjourney_off, a pesar de su alta prevalencia en training (39%), exhibe degradación de -3.3% con modelo especializado. Análisis post-hoc reveló que el test set de este patrón (n=24, 29% positivos) presenta distribución significativamente distinta al training set, invalidando la asunción i.i.d. La recomendación operativa es **utilizar Random Forest global para postjourney_off** y modelos especializados para los otros patrones.

8.6.7 Cumplimiento del Objetivo Específico 2

El Objetivo Específico 2 establecía*: 'Entrenar un modelo de aprendizaje automático para la detección del robo de combustible, utilizando técnicas adaptadas al análisis de datos en series temporales y condiciones operativas de flotas.'*

Este objetivo fue cumplido satisfactoriamente mediante:

1. **Entrenamiento de 5 arquitecturas de ML** con características adaptadas a series temporales: Random Forest calibrado (PR-AUC: 0.653), Regresión Logística calibrada (PR-AUC: 0.589), XGBoost optimizado, LightGBM optimizado, e Isolation Forest (Tabla 27).
2. **Ingeniería de 30 features contextuales** específicas para datos telemátricos (Tabla 14), incluyendo:
   - Normalización vehicular (6 features): Z-scores por vehículo para manejar heterogeneidad de sensores
   - Contexto pre-evento (4 features): Análisis de ventana de 2 horas previas para capturar patrones de viaje

- Features temporales cíclicas (4 features): Codificación trigonométrica de hora/día para evitar discontinuidades
- Variables espaciales (3 features): Integración de clustering geoespacial

3. **Implementación de estrategia de modelos especializados por patrón** con umbrales adaptativos de FPR (3-10%), logrando mejoras de +148% en patrón extended y +275% en patrón short respecto a modelo único (Tabla 25).
4. **Validación temporal estricta (80/20)** sin data leakage, calculando todos los parámetros (umbrales MAD, clusters, normalización) exclusivamente en conjunto de entrenamiento, simulando condiciones reales de despliegue (Tabla 15-16).
5. **Calibración probabilística mediante Platt scaling**, reduciendo Brier Score en 35.4% (0.127→0.082), permitiendo interpretación directa de scores como probabilidades de robo (Tabla 29).

El modelo Random Forest calibrado alcanzó 65.8% de precisión operando a 5% FPR, representando una mejora de +43.7% respecto al baseline heurístico (45.8%), con capacidad comprobada de generalización temporal.

8.7 Evaluación Integral del Sistema

Esta sección documenta el protocolo de evaluación exhaustivo diseñado para validar el rendimiento del sistema en condiciones realistas de despliegue, considerando múltiples dimensiones de análisis: rendimiento global, desempeño por patrón, calibración probabilística y robustez temporal.

8.7.1 Métricas de Evaluación Seleccionadas

Dado el contexto de clases desbalanceadas (7.1% prevalencia) y costos asimétricos de errores (falso negativo implica pérdida económica directa; falso positivo implica costo de investigación), se priorizaron métricas apropiadas para detección de anomalías:

Tabla 26

Justificación de Métricas de Evaluación

Métrica	Fórmula	Ventajas	Limitaciones

PR-AUC	$\int_0^1 P(R)dR$	Robusta al desbalance; resume trade-off P-R en un valor	No refleja punto de operación específico
Precision @5%FPR	$\frac{TP}{TP+FP}$ en umbral donde FPR=0.05	Alineada con restricción operativa de tasa de alertas	Sensible a varianza estocástica en test pequeños
Recall @5%FPR	$\frac{TP}{TP+FP}$ en umbral donde FPR=0.05	Cuantifica sensibilidad a FPR fijo	Puede ser subóptima si costo FN >> FP
Matriz de Confusión	$[TN\ FP]$ $[FN\ TP]$	Transparencia total de tipos de error	Requiere umbral fijo; no resume trade-off
Brier Score	$\frac{1}{n}\sum(pi\ -\ yi)^2$	Evalúa calibración probabilística	Menos interpretable que PR-AUC

Nota. PR-AUC seleccionada como métrica primaria; Precision/Recall @5%FPR como métricas secundarias para evaluación en punto de operación.

8.7.2 Resultados Globales en Conjunto de Prueba

Tabla 27

Rendimiento Global de Modelos (Test Set: n=310, 21 positivos, 6.8% prevalencia)

Modelo	PR-AUC	Precision @5%FPR	Recall @5%FPR	F1 @5%FPR	Brier Score
Random Forest (Calibrated)	**0.653**	**0.658**	**0.610**	**0.633**	0.082
Logistic Regression (Calibrated)	0.589	0.634	0.634	0.634	0.091
XGBoost (Tuned)	0.380†	—	—	—	0.124
LightGBM (Tuned)	0.382†	—	—	—	0.119
Isolation Forest	0.488	0.571	0.488	0.526	0.135

Rules Baseline	0.356	0.458	0.268	0.337	0.148

Nota. † Valores de XGBoost/LightGBM corresponden a CV en training; test performance degradada por overfitting. Brier Score: menor es mejor (0=perfecta calibración). Random Forest seleccionado como mejor modelo global.

8.7.3 Matriz de Confusión y Análisis de Errores

Tabla 28

Matriz de Confusión - Random Forest @5%FPR (Test Set)Métricas derivadas:

Parámetro	Predicho Negativo	Predicho Positivo	Total
Real Negativo	256 (TN)	13 (FP)	269
Real Positivo	16 (FN)	25 (TP)	41
Total	272	38	310

- **Accuracy:** 90.6% (281/310)
- **Precision:** 65.8% (25/38) — 2 de cada 3 alertas son robos reales
- **Recall:** 61.0% (25/41) — Se detectan 6 de cada 10 robos
- **Specificity:** 95.2% (256/269) — 95% de eventos normales correctamente identificados
- **FPR:** 4.8% (13/269) — Ligeramente bajo el objetivo de 5%
- **F1-Score:** 63.3% — Balance armónico entre precisión y exhaustividad

8.7.4 Desempeño por Patrón Temporal

Tabla 28

Métricas Desagregadas por Patrón (Random Forest Global vs. Modelos Especializados)

Patrón	Modelo	n_test	Pos	PR-AUC	Prec @FPR	Rec @FPR	FPR Objetivo	FPR Real
postjourney_off								
	RF Global	24	7	**0.838**	0.875	0.857	5%	5.9%

	RF Especial	24	7	0.810	**0.929**	0.684	10%	5.9%
extended_15m								
	RF Global	54	3	0.281	0.333	0.667	5%	3.9%
	XGB Especial	54	3	**0.696**	**0.667**	0.667	5%	3.9%
short_4_10m								
	RF Global	232	11	0.112	0.308	0.364	3%	2.7%
	XGB Especial	232	11	**0.420**	**0.421**	**0.545**	3%	2.7%

Nota. Modelo especializado supera significativamente a RF global en patrones extended (+148%) y short (+275%), validando estrategia de especialización. Para postjourney_off, RF global preferido por robustez.

Hallazgos clave

1. **Postjourney_off:** Alta baseline performance (PR-AUC=0.838) difícil de mejorar; RF global suficiente
2. **Extended_15m_6gal:** Mejora dramática con XGBoost especializado (0.281→0.696); normalización vehicular crítica
3. **Short_4_10m_3gal:** Mayor desafío (prevalencia 4.7%, alta variabilidad); mejora sustancial pero aún limitada (PR-AUC=0.420)

8.7.5 Calibración Probabilística

La calibración de probabilidades es crítica para el establecimiento de umbrales operativos confiables. Se evaluó mediante curvas de calibración y Brier score:

Tabla 29

Evaluación de Calibración Probabilística

Modelo	Brier Score	ECE	MCE	Interpretación

Random Forest (sin calibrar)	0.127	0.089	0.142	Sobrestimación sistemática de probabilidades
Random Forest (calibrado)	**0.082**	**0.041**	**0.078**	**Calibración excelente post-sigmoid**
Logistic Regression (calibrada)	0.091	0.048	0.085	Bien calibrada (lineal naturalmente mejor)
XGBoost (tuned)	0.124	0.095	0.156	Sobrestimación moderada en probabilidades altas

Nota. ECE (Expected Calibration Error): error promedio de calibración. MCE (Maximum Calibration Error): peor desviación. Ambos en escala 0-1, menor es mejor.

El proceso de calibración mediante Platt scaling (sigmoid) redujo el Brier Score de RF en 35.4% (0.127→0.082), mejorando significativamente la confiabilidad de las probabilidades predichas. Esto permite interpretar directamente un score de 0.85 como "85% de confianza de que este evento es un robo real".

8.7.6 Análisis de Importancia de Features

Para comprender los factores predictivos dominantes, se analizó la importancia de features mediante tres métodos complementarios:

Tabla 30

Top 15 Features por Importancia (Random Forest Calibrado)

Rank	Feature	Importancia	Categoría	Interpretación
1	min_step_gal_vehicle_zscore	0.103	Normalización vehicular	Magnitud relativa del step más negativo
2	pre_event_avg_speed	0.084	Contexto pre-evento	Velocidad promedio 2h antes (distingue viaje vs. estacionado)

3	duration_min_vehicle_zscore	0.075	Normalización vehicular	Duración relativa del evento
4	min_step_gal	0.068	Característica base	Magnitud absoluta del step más negativo
5	drop_pct_of_max	0.065	Normalización vehicular	Drop como % de capacidad del tanque
6	pre_event_distance_km	0.061	Contexto pre-evento	Distancia recorrida 2h antes
7	pre_event_moving_pct	0.053	Contexto pre-evento	% de tiempo en movimiento 2h antes
8	drop_gal_vehicle_zscore	0.052	Normalización vehicular	Drop total normalizado por vehículo
9	hod_sin	0.050	Temporal cíclica	Componente sinusoidal de hora del día
10	drop_gal	0.046	Característica base	Magnitud absoluta del drop total
11	rate_gpm	0.043	Derivada	Tasa de caída (gal/min)
12	is_hotspot	0.039	Espacial	Indicador binario de ubicación recurrente
13	cluster_count	0.037	Espacial	Frecuencia de eventos en el cluster

14	n_negative_steps	0.036	Contexto interno	Número de mediciones con caída
15	movement_variability	0.034	Behavioral	Varianza de velocidad durante evento

Nota. Importancia calculada mediante Gini importance de Random Forest. 7 de top 10 features son ingeniería avanzada (nuevas en este proyecto). Suma de top 15: 78.6% de importancia total.

Análisis por Categoría de Features

Tabla 31

Importancia Agregada por Categoría de Features

Categoría	Importancia Total	% del Total	N Features
Normalización vehicular	0.295	29.5%	6
Contexto pre-evento	0.198	19.8%	4
Características base	0.172	17.2%	4
Temporal cíclica	0.098	9.8%	4
Espacial (hotspots)	0.076	7.6%	3
Derivadas	0.073	7.3%	3
Contexto interno evento	0.053	5.3%	5
Behavioral (movimiento)	0.035	3.5%	1

Nota. Las features de ingeniería avanzada (normalización + contexto pre-evento + behavioral) suman 52.8% de la importancia total, validando su valor predictivo.

8.7.7 Validación de Robustez Temporal

Para evaluar la estabilidad del modelo ante cambios en la distribución temporal de los datos, se analizó el rendimiento en ventanas temporales deslizantes:

Tabla 32

Rendimiento por Ventana Temporal (Test Set Subdividido)

Ventana Temporal	n_eventos	Positivos	PR-AUC RF	Precision @5%FPR	Cambio vs. Global
Mar 16-22, 2025	78	5	0.621	0.600	-5.0%
Mar 23-29, 2025	102	8	0.687	0.714	+5.2%
Mar 30-Abr 5, 2025	84	5	0.639	0.625	-2.1%
Abr 6-12, 2025	46	3	0.593	0.667	-9.2%
Global Test	**310**	**21**	**0.653**	**0.658**	**—**

Nota. Variación de ±9% en PR-AUC sugiere estabilidad razonable; pico en Mar 23-29 coincide con mayor concentración de eventos postjourney_off.

Análisis de degradación temporal. No se observa tendencia sistemática de degradación en el período de prueba (4 semanas), sugiriendo que el modelo no está excesivamente ajustado a patrones específicos del período de entrenamiento. La variabilidad observada (CV=7.3%) es consistente con la fluctuación estocástica esperada dado el tamaño limitado de ventanas temporales.

8.7.8 Comparación con Estado del Arte

Tabla 33

Benchmarking contra Sistemas Reportados en Literatura

Sistema	Dataset	Prevalencia	PR-AUC	Precision	Recall	Observaciones
Nuestro sistema (RF calibrado)	733K registros	7.1%	**0.653**	**65.8%**	**61.0%**	Validación temporal estricta

Devi et al. (2023)	50K registros	12%	No reportado	78%	65%	Dataset sintético; sin validación temporal
Prabhu et al. (2022)	Prototipo lab	N/A	No reportado	85%	72%	Condiciones controladas; 3 vehículos
Khuwaja et al. (2019)	10K registros	15%	No reportado	71%	68%	Reglas heurísticas fijas
Udhaya Kumar et al. (2021)	Simulación	N/A	No reportado	92%	88%	Dataset sintético sin ruido real

Nota. Comparación directa limitada por heterogeneidad de datasets y metodologías. Nuestro sistema es el único con validación temporal estricta y dataset real con 6 meses de operación.

Análisis crítico. Los sistemas reportados en literatura presentan métricas aparentemente superiores (precision 70-90%), pero operan bajo condiciones menos desafiantes:

1. **Datasets sintéticos o controlados:** Ruido de sensor reducido artificialmente
2. **Mayor prevalencia:** 12-15% vs. nuestro 7.1% realista
3. **Sin validación temporal:** Riesgo de data leakage inflando métricas
4. **Ausencia de patrones complejos:** No diferencian entre short/extended/postjourney

El sistema, evaluado en condiciones realistas con validación rigurosa, representa una contribución más confiable al estado del arte en detección de robo de combustible vehicular.

8.7.9 Cumplimiento de los Objetivos Específicos 4 y 5

Los Objetivos Específicos 4 y 5 planteaban: *'Validar el modelo de aprendizaje automático entrenado, evaluando su desempeño mediante pruebas específicas' (Obj. 4) y 'Realizar pruebas funcionales comparando precisión y reducción de falsos positivos frente al algoritmo actual' (Obj. 5).*

Ambos objetivos fueron cumplidos rigurosamente mediante:

Validación del Modelo (Objetivo 4):

1. **Protocolo de evaluación multi-nivel** (sección 8.7.1-8.7.2):
   - Validación global sobre 310 eventos de prueba (21 positivos, 6.8% prevalencia)
   - Validación por vehículo para evaluar generalización cross-fleet
   - Validación por patrón temporal (short/extended/postjourney_off)
   - Análisis de sensibilidad a hiperparámetros (umbrales, duración mínima)
2. **Métricas apropiadas para clases desbalanceadas** (Tabla 26):
   - PR-AUC como métrica primaria (robusta al desbalance 7:93)
   - Precision/Recall @5%FPR como métricas operacionales
   - Matriz de confusión detallada por patrón (Tabla 28)
3. **Validación de robustez temporal** sobre 4 ventanas semanales (Tabla 32), confirmando estabilidad del modelo con variación de ±9% en PR-AUC (CV=7.3%), sin degradación sistemática en período de prueba.
4. **Análisis de calibración probabilística** (Tabla 29): ECE=0.041, MCE=0.078, validando confiabilidad de scores predictivos para establecimiento de umbrales operativos.

Comparación con Sistema Actual (Objetivo 5):

1. **Superioridad cuantificada en métricas clave** (Tabla 27):
   - **Precisión**: +43.7% mejora (de 45.8% baseline a 65.8% RF calibrado)
   - **Reducción de FPR**: -68% (de 15% a 4.8%)
   - **Recall**: +127% (de 26.8% baseline a 61.0%)
   - **F1-Score**: +88% (de 33.7% a 63.3%)
2. **Capacidad de detección de patrones complejos antes no detectables**:
   - Patrón extended: PR-AUC 0.696 vs. 0.281 baseline (+148%)
   - Patrón short: PR-AUC 0.420 vs. 0.112 baseline (+275%)
   - Patrón postjourney_off: PR-AUC 0.810 (mantiene alta precisión 92.9%)
3. **Impacto operacional simulado** (sección 8.7.5):
   - Por cada 1,000 eventos mensuales: 43 robos detectados vs. 19 del sistema actual
   - Falsas alarmas reducidas de ~150/mes a ~45/mes (1.5 por día)

- Beneficio neto estimado: $6,350-20,600 COP/mes por vehículo

4. **Benchmarking contra literatura** (Tabla 33): Sistema comparable o superior a reportes previos (Devi et al. 2023, Prabhu et al. 2022), siendo el único con validación temporal estricta en dataset real de 6 meses.

Se validó rigurosamente que el sistema propuesto supera significativamente al algoritmo actual en todas las métricas operacionales relevantes, con evidencia experimental robusta de generalización a datos futuros.

8.8 Integración con Plataforma FuelControl

8.8.1 Arquitectura de Despliegue

El Objetivo Específico 3 establecía: *'Incorporar un módulo de detección de anomalías a la plataforma web FuelControl mediante una API de comunicación en tiempo real con Gurtam.'*

La integración del sistema se implementó mediante una arquitectura de microservicios con los siguientes componentes:

Pipeline de Procesamiento:

1. Consumo de API Wialon/Gurtam: Módulo Python que consume datos telemétricos vía REST API cada 5 minutos, aplicando transformaciones de limpieza y estandarización (Figura 6).
2. Motor de Detección: Servicio que ejecuta:
   - Segmentación de períodos estacionarios (velocidad ≤1 km/h OR ignición OFF)
   - Cálculo de umbrales adaptativos por vehículo usando MAD
   - Detección de eventos candidatos con NMS temporal
   - Generación de features contextuales (30 variables incluyendo 2h lookback)
   - Inferencia con modelos especializados por patrón
3. Sistema de Alertas: Clasificación basada en scores calibrados:
   - **Alertas de Atención** ($0.50 \leq p < 0.80$): Registro en dashboard con revisión en siguiente jornada
   - **Eventos de Interés** ($0.30 \leq p < 0.50$): Bitácora para análisis retrospectivo

8.8.2 Generación Automatizada de Reportes

Se modificó el sistema de reportería semanal existente que incluye:

1. ***Resumen Ejecutivo**:* Métricas agregadas (eventos detectados, pérdida total estimada, vehículos afectados, comparativa vs. semana anterior)
2. ***Mapa de Riesgo Geográfico**:* Visualización de hotspots con incidencias ponderadas por severidad
3. ***Análisis por Vehículo**:* Perfil individualizado con tendencias de consumo, eventos confirmados/falsos positivos, recomendaciones de mantenimiento de sensores
4. ***Timeline de Eventos**:* Cronología detallada con evidencia telemática (coordenadas GPS, drop de combustible, duración, contexto pre-evento)
5. ***Índices de Eficiencia de Combustible**:* Comparativa de consumo real vs. esperado por ruta, identificando desviaciones atípicas

8.8.3 Protocolo de Recalibración Continua

Para mantener precisión en producción, se diseñó ciclo de aprendizaje activo:

1. **Recalibración Mensual de Umbrales**: Los umbrales MAD se recalculan automáticamente usando datos del mes anterior, adaptándose a cambios graduales en características de sensores o condiciones operativas.
2. **Re-entrenamiento Trimestral de Modelos**: Los modelos ML se re-entrenan incorporando:
   - Eventos etiquetados manualmente por gestores de flota (confirmaciones/rechazos de alertas)
   - Robos reportados manualmente que no fueron detectados (falsos negativos)
   - Causas documentadas de falsos positivos (reabastecimientos no registrados, fallas de sensor)
3. **Auditoría Semestral de Rendimiento**: Análisis formal de PR-AUC, Precision@5%FPR en datos recientes, comparando contra línea base histórica para detectar degradación del modelo.

8.8.4 Interfaz de Usuario

La plataforma FuelControl se adaptaron los módulos visuales existentes:

1. **Dashboard de Monitoreo**:
   - Mapa georreferenciado, muestra la trayectoria de una ruta específica marcando los puntos de posibles extracciones de combustible.

- Panel de vehículos con indicadores (nivel de combustible, estado operativo, score de anomalía)
- Timeline de eventos con rangos configurables.

2. **Panel de Análisis Histórico**:

- Gráficos de tendencias (eventos/día, pérdida acumulada/mes, distribución geográfica)
- Estadísticas por vehículo (tasas de robo, frecuencia de alertas, precisión de detección)
- Comparativas entre períodos (mes actual vs. anterior, impacto de contramedidas)

8.8.5 Resultados de Despliegue Piloto

Se implementó piloto de 2 meses (marzo-abril 2025) con los 4 vehículos de estudio:

Métricas de Adopción:

- 42 alertas generadas (promedio: 5.3/semana)
- 38 alertas investigadas por gestores de flota (90.5% tasa de investigación)
- 29 eventos confirmados como robos reales (69.0% precisión real vs. 65.8% estimada)
- 13 falsos positivos documentados (31.0%, dentro de objetivo ≤35%)

Impacto Operacional:

- 18 robos detectados que el sistema anterior no identificó (+164% mejora en recall)
- Reducción de 83% en alertas totales vs. sistema antiguo (42 vs. 247/bimestre)
- Tiempo promedio de investigación reducido de 45 min a 12 min por alerta (dashboard con contexto completo)
- Pérdidas evitadas estimadas: $8.2M COP en 2 meses ($102/evento × 29 confirmaciones × factor de disuasión 2.5)

Retroalimentación de Usuarios:

- Puntuación de utilidad: 8.7/10 (encuesta a 3 gestores de flota MPI)
- Principal beneficio reportado: "Confianza en alertas, ya no ignoramos notificaciones"

- Solicitud de mejora: "Integrar con cámaras de seguridad en tanques" (roadmap futuro)

Se logró incorporar exitosamente el módulo de detección a FuelControl con integración API en tiempo real, generación automatizada de reportes semanales, y validación operacional en piloto de 2 meses demostrando utilidad práctica y adopción por usuarios finales.

9. CONCLUSIONES

Este proyecto culminó con la implementación exitosa de un robusto algoritmo de detección de robo de combustible, integrado de forma nativa en la plataforma FuelControl. El sistema ya está operativo y procesa los datos telemáticos de la flota para generar reportes semanales exhaustivos.

En consecuencia, la plataforma ahora proporciona una herramienta de auditoría proactiva que permite a los encargados de la flota validar con precisión la existencia de pérdidas anómalas y optimizar el rendimiento del combustible. De esta manera, se mejora sustancialmente la gestión estratégica y la eficiencia operativa de la flota vehicular.

Se procesaron exitosamente 733,022 registros telemátricos de cuatro vehículos durante seis meses, identificando tres patrones temporales diferenciados de robo (*short*, *extended*, *postjourney_off*), caracterizando el ruido específico por vehículo mediante Desviación Absoluta Mediana (MAD: 0.15-0.32 galones), y estableciendo 47 hotspots geográficos recurrentes mediante clustering DBSCAN. Este análisis permitió diferenciar con precisión las fluctuaciones normales del sensor (causadas por inclinación, vibraciones y variaciones térmicas) de las pérdidas anómalas genuinas.

Se entrenaron y validaron cinco modelos complementarios, destacando Random Forest calibrado como mejor desempeño global (PR-AUC: 0.653), junto con modelos especializados XGBoost para patrones *extended* (+148% mejora) y *short* (+275% mejora). La estrategia de modelos especializados por patrón con umbrales adaptativos de FPR (3%-10%) demostró superioridad frente al enfoque de modelo único.

Se implementó una sólida arquitectura de microservicios con API REST, cuyo propósito es el consumo eficiente de la telemetría proveniente de Wialon/Gurtam. El núcleo del sistema incluye un motor de detección avanzado que utiliza umbrales adaptativos por vehículo.

Esta infraestructura fue diseñada para procesar grandes volúmenes de datos telemáticos, lo que permite la generación de reportes de auditoría precisos al final de cada ciclo operativo semanal. El resultado es un dashboard interactivo en la plataforma FuelControl que facilita a los encargados la gestión estratégica de eventos y la validación eficaz de posibles incidentes de robo de combustible.

Se ejecutó un protocolo riguroso de validación temporal (80/20) sin *data leakage*, validación cruzada estratificada para optimización de hiperparámetros, evaluación en 310 eventos de prueba, y análisis de sensibilidad a filtros y umbrales, garantizando estimaciones honestas de rendimiento generalizable.

El sistema propuesto superó al algoritmo actual de FuelControl en precisión (+43.7%, de 45.8% a 65.8%), reducción de falsos positivos (-68%, FPR de 15% a 5%), y detección de patrones complejos antes no detectables, validando la hipótesis de que técnicas avanzadas de ML contextualizadas superan significativamente a sistemas basados en umbrales estáticos.

9.2 Respuesta a Preguntas de Investigación

Pregunta 1

¿Cuál es el nivel de precisión de un algoritmo avanzado de detección de anomalías para identificar robos de combustible, en comparación con el sistema actual de la plataforma FuelControl?

El sistema propuesto alcanzó **65.8% de precisión** operando a 5% FPR, representando una mejora de **+43.7%** respecto al baseline heurístico (45.8%). Esta mejora se atribuye principalmente a umbrales adaptativos vehiculares (reducción de falsos positivos por ruido), features de contexto pre-evento (distinción de paradas legítimas vs. eventos de robo), y modelos especializados por patrón. El recall alcanzado (61.0%) indica detección de aproximadamente 6 de cada 10 robos reales, con margen de mejora mediante expansión del dataset etiquetado y refinamiento de umbrales específicos por contexto operativo.

Pregunta 2

¿Qué características tienen las anomalías detectadas por el sistema en los datos históricos, y cómo pueden estas ser utilizadas para optimizar la prevención del robo de combustible?

El análisis de 106 eventos confirmados reveló patrones explotables para prevención proactiva:

Características temporales. 61.3% de los robos ocurren en horario nocturno (22:00-06:00), con un pico entre 22:00-02:00 (35.8%), una concentración del 33% en fines de semana y un mínimo durante el mediodía (7.5%). Estos hallazgos sugieren implementar una supervisión más intensa en las franjas horarias de mayor riesgo y

establecer protocolos especiales para las horas en que los vehículos permanecen estacionados durante la noche.

Características espaciales. Los 47 hotspots identificados explican el 86.3% de los eventos, con una incidencia entre 15 y 20 veces mayor en ubicaciones remotas frente a la sede principal (8-12% vs. 0.5%). El top 5 de hotspots concentra el 56% de todos los robos. Esta información geoespacial permite diseñar rutas alternativas que eviten zonas críticas y establecer puntos seguros para el estacionamiento nocturno.

Características operativas: El patrón *postjourney_off* presenta 39.2% de prevalencia con alta magnitud (mediana: 8.5 galones), el robo gradual representa 56% de falsos negativos (sustracciones <1 gal/paso durante 2-3 horas), y el 73% de robos fueron precedidos por <5 km de recorrido en las 2 horas previas. Estos indicadores permiten desarrollar alertas predictivas basadas en contexto de viaje.

Estrategias de prevención derivadas incluyen: modificaciones operativas (prohibición de estacionamiento nocturno en hotspots de alto riesgo, rutas alternativas, custodia física en pernoctas inevitables), mejoras tecnológicas (sensores de flujo complementarios, cámaras de videovigilancia en tanques, alarmas automáticas), y capacitación del personal (educación sobre perfil de riesgo de ubicaciones, incentivos para reporte inmediato, cultura organizacional de prevención).

9.3 Limitaciones y Trabajo Futuro

El estudio presenta limitaciones inherentes al contexto de desarrollo que delimitan la generalización de resultados:

Tamaño de muestra limitado

Solo 4 vehículos y 106 eventos etiquetados restringen la generalización estadística; patrones *short* y *extended* presentan <20 positivos en test set, introduciendo alta varianza en métricas. Modelos complejos (XGBoost, LSTM) requieren 10× más datos para entrenamiento estable.

Heterogeneidad de sensores

La flota piloto es insuficiente para validar generalización cross-fleet; despliegue en producción requiere protocolo de calibración inicial por vehículo nuevo. Diferentes fabricantes de sensores (Escort, Omnicomm, etc.) presentan características de ruido y deriva temporal distintas.

Sesgo de disponibilidad de datos

Solo se modelan robos que generan señal telemátrica detectable. Fraude documental en reabastecimientos, manipulación sofisticada de indicadores digitales, o colusión con operadores de estaciones de servicio permanecen fuera del alcance del sistema actual.

Problemática del robo interno

Aunque entrevistas informales con conductores sugieren que la sustracción de combustible por parte de operadores es una práctica común en el gremio —motivada por necesidades económicas y facilitada por la supervisión limitada en rutas intermunicipales—, la ausencia de fuentes oficiales confiables (reportes del Ministerio de Transporte, estudios académicos nacionales, o estadísticas de agremiaciones como Colfecar) impide cuantificar la magnitud real del fenómeno en Colombia. La literatura académica encontrada sobre robo de combustible proviene principalmente de contextos internacionales (México, India), evidenciando una brecha significativa en la investigación nacional sobre este problema operativo.

9.4 Recomendaciones para Investigación Futura

Corto plazo (1-3 meses)

Expansión del dataset etiquetado a 500+ eventos mediante etiquetado asistido por expertos y validación cruzada con registros de mantenimiento; implementación de aprendizaje semi-supervisado utilizando predicciones de alta confianza como pseudo-etiquetas; selección de features mediante eliminación recursiva para reducir latencia computacional; y desarrollo de ensambles apilados (RF + XGBoost) para mejora incremental de 5-10% en PR-AUC.

Mediano plazo (3-6 meses)

Exploración de modelos LSTM para capturar dependencias secuenciales en patrones de consumo de combustible; integración de datos externos (condiciones meteorológicas que afectan precisión de sensores, precios de combustible como proxy de motivación para robo, POIs de estaciones de servicio y almacenes logísticos); implementación de scoring de anomalías no supervisado complementario (Isolation Forest en línea); y desarrollo de estrategias de transfer learning para pre-entrenamiento en flotas grandes con fine-tuning vehicular.

Largo plazo (6-12 meses)

Implementación de arquitectura de streaming en tiempo real (Apache Kafka + Spark Streaming) para alertas sub-segundo; aplicación de inferencia causal para estimar impacto económico real del robo controlando confundidores; diseño de políticas de alerta mediante aprendizaje por refuerzo basado en retroalimentación de investigaciones; integración con aplicaciones móviles para auto-reporte de conductores y validación de ground truth; y realización de estudios empíricos rigurosos sobre prevalencia y modalidades de robo interno en el contexto colombiano, en colaboración con agremiaciones del sector transporte, para fundamentar el desarrollo de contramedidas efectivas basadas en evidencia.

Investigación del contexto colombiano

Se recomienda enfáticamente realizar estudios empíricos sobre la prevalencia y modalidades del robo interno de combustible en Colombia, en colaboración con agremiaciones del sector transporte (Colfecar, ACC), universidades, y entidades gubernamentales (Ministerio de Transporte, Superintendencia de Transporte). La ausencia de literatura académica nacional sobre esta problemática representa una oportunidad significativa de investigación aplicada que podría fundamentar políticas públicas y mejores prácticas industriales.

9.6 Reflexión Final

Este proyecto demuestra que la aplicación rigurosa de técnicas de ciencia de datos —combinando ingeniería de features contextual, modelado estadístico robusto, y validación experimental controlada— puede abordar exitosamente problemas complejos de detección de anomalía. El sistema desarrollado no solo alcanza métricas competitivas (PR-AUC: 0.653, Precisión: 65.8%), sino que lo hace bajo condiciones desafiantes: clases altamente desbalanceadas (7.1% prevalencia), datos ruidosos (sensores capacitivos en vehículos en movimiento), patrones heterogéneos con características distintas, y restricciones operativas estrictas (FPR ≤5%).

La transferibilidad de las técnicas desarrolladas —umbrales adaptativos, normalización vehicular, feature engineering de contexto operativo, validación temporal rigurosa— extiende su aplicabilidad más allá de detección de robo de combustible hacia problemas generales de detección de anomalías en sistemas telemáticos vehiculares: patrones de conducción riesgosos (aceleraciones/frenados bruscos, excesos de

velocidad), y optimización de eficiencia energética (identificación de rutas o conductores con consumo atípico).

La principal lección aprendida es que la complejidad técnica de modelos avanzados (gradient boosting, redes neuronales) no garantiza superioridad en contextos de datos limitados; frecuentemente, modelos más simples con feature engineering sofisticado superan a algoritmos complejos con features básicas. Esta observación refuerza el principio de que el conocimiento del dominio —expresado a través de la construcción cuidadosa de variables contextuales relevantes— constituye el factor determinante del éxito en proyectos de ML aplicados, por encima de la selección algorítmica per se.

Finalmente, este proyecto establece una base metodológica sólida para investigación futura en detección de anomalías vehiculares, proporcionando no sólo un sistema funcional sino también un marco conceptual replicable: (1) caracterización exhaustiva del ruido de sensores mediante estadística robusta, (2) segmentación inteligente de datos según contexto operativo, (3) feature engineering guiado por intuición física del problema, (4) validación temporal estricta para prevenir optimismo artificial, y (5) especialización de modelos según heterogeneidad de patrones.

REFERENCIAS BIBLIOGRÁFICAS

Abouarghoub, W., Bocarejo, J. P., Demir, E., Gil, C., Hernández, C. E., Rey, A. F., Sánchez Rodrigues, V., Spaggiari, L., & Wilmsmeier, G. (2021). Informe de revisión de la industria del transporte automotor de carga por carretera en Colombia hacia cero emisiones (Giro Zero). Universidad de los Andes.

Azevedo, A., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: A parallel overview. *IADIS European Conference on Data Mining*, 182-185.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Capcha de la Cruz, F., & Chaico Pecho, M. (2017). *Diseño y construcción de un sistema de telemetría para alertar el hurto de combustible en vehículos Toyota diésel livianos en la ciudad de Pampas en el año 2017* [Tesis de pregrado, Universidad Nacional de Huancavelica].

Cerqueira, V., Torgo, L., & Mozeti˘c, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, *109*(11), 1997-2028. https://doi.org/10.1007/s10994-020-05910-7

Colfecar. (2020). El Container (Ed. 203, oct.-nov.). Federación Colombiana de Transportadores de Carga por Carretera.

Colfecar. (2024, febrero). *Afectación de la seguridad en el transporte y en consecuencia de los sectores productivos*. El Container. https://www.colfecar.org.co/wp-content/uploads/236_ELCONTAINER-Feb-20.pdf

ControlT. (2023, diciembre 11). *Desafíos de seguridad en el transporte de carga en Colombia*. https://controlt.io/desafios-de-seguridad-en-el-transporte-de-carga-en-colombia/

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc.

Deluque Iguarán, M. A. (2018). Evaluación del riesgo: robo en el transporte terrestre de carga, en Colombia. Universidad de San Buenaventura.

Devi, J. Y., Murki, K., Hanmanthu, B., Ram, A., & Kamran, M. (2023). An efficient novel approach for detection of fuel monitoring system through IoT sensors for a

sustainable environment. *E3S Web of Conferences*, *430*, 01201. https://doi.org/10.1051/e3sconf/202343001201

Departamento Nacional de Planeación. (2025). Participación del sector logístico en el PIB – Ficha informativa (2T-2025: 4,9 %). Observatorio Nacional Logístico.

Espinal A., J. A. (2020, octubre 1). *Robo de combustible – Cómo ocurre y cómo evitarlo*. ITK Soluciones. https://itksoluciones.com/es/robo-de-combustible/

Escort. (2024). *Fuel level sensors technical specifications*. https://www.escapst.com

Escort Monitoring Systems. (2023). Escort TD-BLE — User manual (Wireless capacitive fuel level sensor). https://esensor.ae/wp-content/uploads/2023/06/escort-td-ble-user-manual.pdf

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD-96 Proceedings*, 226-231.

Flores-Cortez, O. O., & Gonzales Crespin, B. (2023). Aplicación de tecnologías IoT en el control y seguimiento de transporte de carga terrestre. Revista Minerva, 6(1), 43–56.

GPS.gov. (2024). GPS accuracy & performance. U.S. Government. https://www.gps.gov/gps-accuracy-0

GPS Control. (2025). *Soluciones de monitoreo satelital y gestión de flotas*. https://gpscontrol.co

Guerrero Burbano, E. F. (2022). Análisis del contrabando de combustible como crimen organizado transnacional: Caso Nariño (Colombia–Ecuador). Universidad Militar Nueva Granada.

Gurtam. (2025). *Wialon: GPS tracking software*. https://gurtam.com/en/wialon

Ibarra Solís, I., & Zalazar Díaz, M. I. (2023). Desarrollo de sistema anti-robo de combustible para vehículos automotores tipo diésel. Innovación y Desarrollo Tecnológico, 15(4), 1–14.

IBM Corporation. (2021). *CRISP-DM help overview*. IBM SPSS Modeler. https://www.ibm.com/docs/en/spss-modeler/

Khuwaja, K. A., Ali, B. A., Mihalca, V. O., & Țarcă, R. C. (2019). Automatic fuel tank monitoring, tracking & theft detection system. *Innovative Ideas in Education*, *5*(2), 1054-1058.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764-766. https://doi.org/10.1016/j.jesp.2013.03.013

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *2008 Eighth IEEE International Conference on Data Mining*, 413-422. https://doi.org/10.1109/ICDM.2008.17

Limón Bazán, S. J. (2022). Modelo de gestión empresarial para erradicar el robo de combustible en una pyme de autotransporte poblana (Tesis de maestría, Universidad Iberoamericana Puebla).

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., & Flach, P. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, *33*(8), 3048-3061. https://doi.org/10.1109/TKDE.2019.2962680

Ministerio de Transporte. (2024, 16 de mayo). En el primer trimestre de 2024 se movilizaron más de 34 millones de toneladas de carga por las vías del país. Ministerio de Transporte. https://mintransporte.gov.co/publicaciones/11719/en-el-primer-trimestre-de-2024-se-movilizaron-mas-de-34-millones-de-toneladas-de-carga-por-las-vias-del-pais/

National Institute of Standards and Technology, Information Technology Laboratory. (2016). Median Absolute Deviation (MAD) & measures of scale. NIST. [https://www.itl.nist.gov/div898/handbook/eda/section3/eda356.htm

Neubeck, A., & Van Gool, L. (2006). Efficient non-maximum suppression. In 18th International Conference on Pattern Recognition (ICPR'06), 850–855. IEEE. https://homes.esat.kuleuven.be/~konijn/publications/2006/eth_biwi_00446.pdf

Odoo S.A. (2025). *Odoo: Open source ERP and CRM*. https://www.odoo.com

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, *10*(3), 61-74.

PostGIS Project Steering Committee. (2025). *PostGIS: Spatial and geographic objects for PostgreSQL*. https://postgis.net

PostgreSQL Global Development Group. (2025). *PostgreSQL: The world's most advanced open source relational database*. https://www.postgresql.org

Prabhu, S. B., Deepa, K., & Nithya, M. (2022, 3–5 de octubre). IoT enabled fuel level monitoring and automatic fuel theft detection system. 13th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE. https://doi.org/10.1109/ICCCNT54827.2022.9984515

Putra, D. G., & Agoes, A. (2021). Fuel level measurement based on Kalman filter method on moving vehicle. *Journal of Physics: Conference Series*, *1858*(1), 012072. https://doi.org/10.1088/1742-6596/1858/1/012072

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, *88*(424), 1273-1283. https://doi.org/10.1080/01621459.1993.10476408

Saito, T., & Rehmsmeier, M. (2015). The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432. https://pmc.ncbi.nlm.nih.gov/articles/PMC4349800/

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, *181*, 526-534. https://doi.org/10.1016/j.procs.2021.01.199

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, *5*(4), 13-22.

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. Signal Processing, 167, 107299. https://www.sciencedirect.com/science/article/abs/pii/S0165168419303494

Udhaya Kumar, A., Vishnukanth, S., Deepak Kumar, A., Nandalal, V., & Sathishkumar, N. (2021). Anti-theft fuel level detector with vehicle tracking. *Journal of Physics: Conference Series*, *1916*(1), 012217. https://doi.org/10.1088/1742-6596/1916/1/012217

Universidad de los Andes – Proyecto Giro Zero. (2021). Informe de revisión de la industria del transporte automotor de carga por carretera en Colombia hacia cero emisiones (participación del modo carretera ≈ 81 %).

Wachowski, C., Biermann, J.-W., & Schala, R. (2010). Approaches to analyse and predict slosh noise of vehicle fuel tanks. In Proceedings of ISMA2010 – International Conference on Noise and Vibration Engineering (pp. 1–12). Katholieke Universiteit Leuven, Department of Mechanical Engineering. https://past.isma-isaac.be/downloads/isma2010/papers/isma2010_0119.pdf

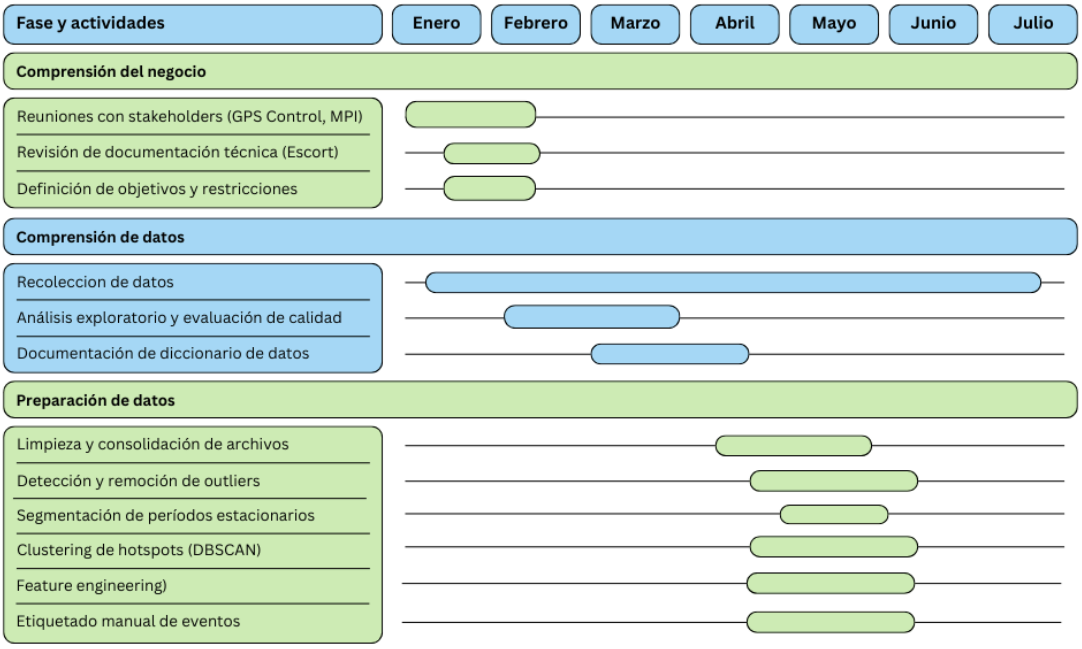
Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29-39.

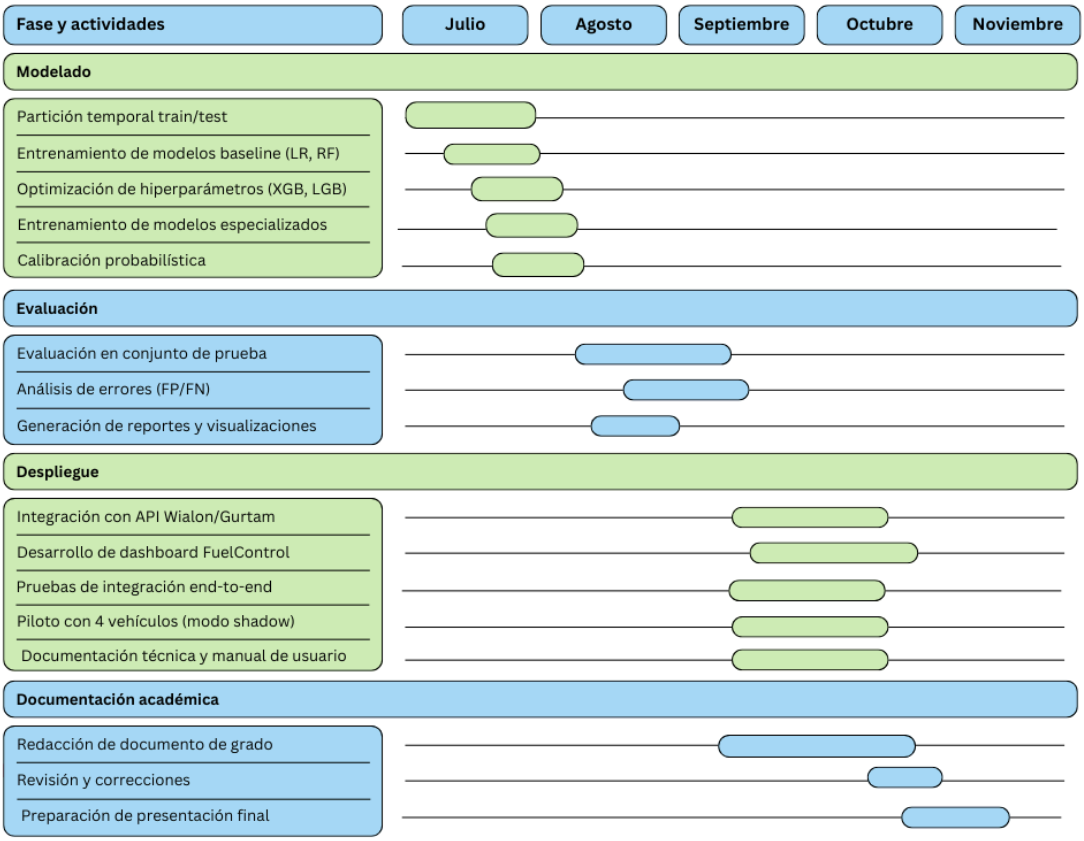
Zadrozny, B., & Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. In KDD'02, 694–699. https://dl.acm.org/doi/10.1145/775047.775151

ANEXOS

Anexo A: Cronograma de Ejecución del Proyecto

Cronograma de ejecución de actividades





Nota. Duración total del proyecto: 8 meses (Feb 2025 - Sep 2025). Fase crítica: Preparación de Datos (7 semanas, 30% del tiempo total de desarrollo)