



ESCUELA DE NEGOCIOS

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS

**MODELO DE MACHINE LEARNING PARA LA PREDICCIÓN DE
EXPERIENCIAS NEGATIVAS EN EL SERVICIO POSTVENTA: UNA
APLICACIÓN EN LA INDUSTRIA AUTOMOTRIZ**

PROFESOR

Ing. Mario Salvador Gonzáles. PhD

AUTOR

Econ. Renato Maximiliano Rivera Abad

2024

RESUMEN

El presente trabajo tiene como objetivo desarrollar e implementar un modelo de machine learning basado en XGBoost para predecir y mitigar las experiencias negativas de los clientes en el servicio postventa de una organización automotriz. Utilizando grandes volúmenes de datos de órdenes de reparación y características vehiculares, se identificaron patrones asociados con situaciones de recompra (buyback) y negociación asistida (trade assist). El modelo predictivo fue evaluado mediante métricas clave como precisión, recall y F1-Score, mostrando un desempeño superior en la predicción de ambas clases después de aplicar técnicas de balanceo de datos, como el submuestreo.

Los resultados evidencian que el modelo permite una mejor toma de decisiones, optimizando la asignación de recursos y mejorando la atención al cliente mediante la creación de un sistema de alerta temprana. Además, se propone la integración del modelo en el sistema CRM de la organización para automatizar las alertas en tiempo real y personalizar las soluciones ofrecidas a los clientes. Las conclusiones destacan la efectividad del modelo en la identificación de experiencias negativas y su potencial para mejorar la satisfacción del cliente y la eficiencia operativa. La implementación de este sistema representa una oportunidad para la empresa de innovar y mejorar su competitividad a largo plazo, posicionándose como líder en la experiencia del cliente en el sector automotriz.

Palabras Clave: Machine learning, predicción, postventa automotriz, XGBoost, experiencia del cliente, satisfacción del cliente, análisis de datos, modelo predictivo, trade assist, buyback.

ABSTRACT

This work aims to develop and implement a machine learning model based on XGBoost to predict and mitigate negative customer experiences in the after-sales service of an automotive organization. By analyzing large volumes of repair order data and vehicle characteristics, patterns related to buyback and trade assist situations were identified. The predictive model was evaluated using key metrics such as precision, recall, and F1-Score, demonstrating superior performance in predicting both classes after applying data balancing techniques like undersampling.

The results show that the model enables better decision-making by optimizing resource allocation and improving customer service through the creation of an early warning system. Additionally, it is proposed to integrate the model into the organization's CRM system to automate real-time alerts and customize solutions offered to customers. The conclusions highlight the model's effectiveness in identifying negative experiences and its potential to improve customer satisfaction and operational efficiency. The implementation of this system represents an opportunity for the company to innovate and enhance its long-term competitiveness, positioning itself as a leader in customer experience within the automotive sector.

Keywords: Machine learning, prediction, automotive after-sales, XGBoost, customer experience, customer satisfaction, data analysis, predictive model, trade assist, buyback.

ÍNDICE DEL CONTENIDO

INTRODUCCIÓN	1
REVISIÓN DE LITERATURA.....	3
IDENTIFICACIÓN DEL OBJETO DE ESTUDIO.....	15
PLANTEAMIENTO DEL PROBLEMA	17
OBJETIVO GENERAL	18
OBJETIVOS ESPECÍFICOS	18
JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA.....	19
RESULTADOS	29
DISCUSIÓN DE LOS RESULTADOS Y PROPUESTA DE SOLUCIÓN	36
CONCLUSIONES Y RECOMENDACIONES	44
REFERENCIAS	47

ÍNDICE DE TABLAS

TABLA 1 MATRIZ COMPARATIVA DE ESTUDIOS PREVIOS	10
TABLA 2 REPORTE DE CALIDAD DE DATOS.....	20
TABLA 3 RESUMEN ESTADÍSTICO, CÁLCULO DEL VALOR P Y COHEN'S D	22

ÍNDICE DE FIGURAS

FIGURA 1 DISTRIBUCIÓN DE LA VARIABLE CATEGÓRICA 'SOURCE' ...	21
FIGURA 2 ANÁLISIS DE CORRELACIÓN DEL SET DE DATOS	25
FIGURA 3 MATRIZ DE CONFUSIÓN LOGISTIC REGRESSION UTILIZANDO EL DATASET COMPLETO	29
FIGURA 4 REPORTE DE MÉTRICAS LOGISTIC REGRESSION CON TODO EL CONJUNTO DE DATOS.....	30
FIGURA 5 MATRIZ DE CONFUSIÓN LOGISTIC REGRESSION APLICANDO SUBMUESTREO	31
FIGURA 6 REPORTE DE MÉTRICAS LOGISTIC REGRESSION APLICANDO SUBMUESTREO	32
FIGURA 7 REPORTE DE MÉTRICAS XGBOOST APLICANDO SUBMUESTREO	34

INTRODUCCIÓN

En la industria automotriz, especialmente en el segmento de autos, la experiencia del cliente es uno de los factores clave que determina la fidelización y la percepción de la marca. Los clientes de este segmento tienen expectativas particularmente elevadas no solo en cuanto a la calidad del producto, sino también respecto al servicio postventa. Una mala experiencia en esta área puede tener consecuencias significativas, como la pérdida de lealtad y una reducción en la tasa de recompra, lo que puede dañar la reputación de la marca en un mercado altamente competitivo (Smith et al., 2020). A medida que las empresas automotrices enfrentan estos desafíos, surge la necesidad de integrar tecnologías avanzadas, como el machine learning (ML), para mejorar la calidad del servicio postventa y anticipar posibles problemas que podrían generar insatisfacción en el cliente.

En este contexto, el uso de órdenes de reparación (Repair Orders, RO), que generan grandes volúmenes de datos valiosos, ofrece una oportunidad significativa para analizar patrones recurrentes que pueden estar asociados a experiencias negativas del cliente. Sin embargo, muchas organizaciones aún carecen de las herramientas analíticas necesarias para aprovechar de manera efectiva estos datos. Como resultado, la respuesta a las quejas de los clientes sigue siendo mayormente reactiva en lugar de proactiva, lo que limita la capacidad de la empresa para prevenir incidentes de insatisfacción y optimizar sus procesos internos (Kim et al., 2021). El uso de modelos predictivos basados en machine learning se presenta como una solución innovadora para abordar este problema y transformar la gestión del servicio postventa en las empresas automotrices.

Este proyecto se enfoca en diseñar e implementar un sistema de machine learning que permita predecir y mitigar experiencias negativas en el servicio postventa de una empresa automotriz de auto, mediante el análisis de datos históricos de órdenes de reparación. La capacidad de anticiparse a los

problemas no solo mejorará la satisfacción del cliente, sino que también optimizará la utilización de recursos en los centros de servicio, aumentando la eficiencia operativa y reduciendo los costos asociados a la resolución de problemas. Además, la creación de un sistema que pueda retroalimentarse y aprender de nuevos datos permitirá que el modelo evolucione continuamente, mejorando su precisión y adaptabilidad a las condiciones cambiantes del mercado (Singh et al., 2020).

REVISIÓN DE LITERATURA

1. Antecedentes y Relevancia del Estudio

1.1. Análisis de la Industria Automotriz

La industria automotriz ha sido, históricamente, una de las más influyentes y determinantes en el desarrollo de tecnologías avanzadas. Desde el siglo XX, los vehículos han evolucionado, integrando nuevos sistemas tecnológicos que no solo mejoran la experiencia de conducción, sino también los servicios que rodean a la vida útil del vehículo, como es el caso del servicio postventa. Los estudios recientes señalan que la satisfacción del cliente en el sector automotriz depende, en gran medida, del servicio postventa, el cual incluye actividades que van desde la reparación y el mantenimiento del vehículo hasta la atención de quejas y garantías (Smith et al., 2020). Dado este contexto, es crucial que las compañías automotrices implementen sistemas avanzados para prever y mitigar experiencias negativas, ya que estas pueden afectar la percepción general de la marca.

En el contexto global, la industria automotriz enfrenta desafíos significativos, como el cambio hacia vehículos eléctricos, la automatización, y las expectativas crecientes de los consumidores en cuanto a servicios. Estas tendencias tecnológicas, sumadas a la competencia creciente, obligan a las compañías automotrices a innovar constantemente para mantenerse competitivas. Según Jones y Li (2021), el uso de datos masivos y tecnologías de inteligencia artificial, como el machine learning, ha permitido a estas compañías no solo optimizar la producción, sino también personalizar el servicio postventa de acuerdo con las necesidades y expectativas del cliente. Estos avances han generado un enfoque más centrado en el cliente, aumentando la necesidad de predecir y corregir experiencias insatisfactorias antes de que ocurran.

1.2. Servicio Postventa en la Industria Automotriz

El servicio postventa juega un papel esencial en la fidelización de clientes dentro del sector automotriz. La eficiencia de este servicio no solo impacta en las

decisiones de recompra, sino que también incide en la imagen global de la marca (Doe et al., 2019). Según un análisis realizado por Lopez y Rodríguez (2020), los consumidores insatisfechos suelen difundir sus malas experiencias a través de redes sociales y otros canales digitales, lo que puede perjudicar gravemente la reputación de la empresa. En este contexto, la predicción de posibles experiencias negativas mediante el uso de machine learning surge como una estrategia efectiva para reducir estos riesgos y aumentar la satisfacción del cliente.

2. Importancia de la predicción de las potenciales experiencias negativas

Prever las posibles experiencias negativas en el servicio postventa es fundamental para mejorar los niveles de calidad del servicio, disminuir las quejas de los clientes y fortalecer la lealtad a la marca. De acuerdo con un estudio de Singh et al. (2020), los modelos predictivos basados en machine learning permiten identificar patrones de insatisfacción que no serían evidentes a través de análisis tradicionales. Estos modelos pueden utilizar datos históricos para prever problemas comunes, como retrasos en las reparaciones, costos inesperados o falta de disponibilidad de piezas. Además, la capacidad de tomar medidas preventivas basadas en predicciones permite a las empresas no solo corregir fallos, sino también mejorar continuamente la experiencia del cliente.

3. Machine Learning para Predicciones

3.1. Definiciones de Machine Learning

El uso de técnicas de aprendizaje automático en la industria automotriz ha crecido exponencialmente, gracias a la ventaja de esta tecnología para analizar grandes volúmenes de datos, aprender de ellos y generar predicciones precisas. Un estudio de Zhao et al. (2021) destacó cómo los algoritmos de aprendizaje de máquina son especialmente útiles en la predicción de fallos en componentes automotrices y en la mejora de los servicios postventa. La integración de estos modelos ha permitido optimizar procesos que, en muchos casos, dependen de datos que carecen de una estructura adecuada, como los comentarios de clientes o los registros de reparación.

El aprendizaje automático es una rama de la inteligencia artificial que se enfoca en crear algoritmos capaces de permitir que los sistemas adquieran conocimientos a partir de los datos y optimicen sus predicciones o decisiones sin la necesidad de programar cada tarea de manera específica. Goodfellow et al. (2019) afirman que los modelos de aprendizaje automático aprovechan patrones obtenidos de grandes volúmenes de datos para generar predicciones exactas y decisiones bien fundamentadas. Estos modelos pueden clasificarse en supervisados, no supervisados o de aprendizaje por refuerzo, en función del tipo de datos y del propósito del análisis.

3.2. Tipos de algoritmos de Machine Learning

3.2.1. Aprendizaje Supervisado

El aprendizaje supervisado es una de las técnicas más utilizadas en el ámbito del machine learning. Este enfoque se basa en trabajar con conjuntos de datos previamente etiquetados, lo que permite que el modelo identifique las relaciones entre las variables de entrada y los resultados esperados. Entre los algoritmos supervisados más comunes se encuentran la regresión lineal, las máquinas de soporte vectorial (SVM) y las redes neuronales. Según Arora et al. (2021), los modelos supervisados, especialmente las redes neuronales profundas, han demostrado un excelente desempeño en tareas complejas de clasificación, como la identificación de defectos en el sector automotriz. Asimismo, algoritmos como los árboles de decisión, incluidos los bosques aleatorios, son muy utilizados por su capacidad para gestionar grandes volúmenes de datos y su facilidad de interpretación en problemas de clasificación (Zhou et al., 2020).

Algunos de los algoritmos más utilizados en aprendizaje supervisado incluyen:

- **Regresión lineal:** Empleado principalmente en problemas de regresión, donde la variable a predecir es continua. Este modelo busca ajustar una línea que minimice el error entre las predicciones y los valores reales.

- **Máquinas de soporte vectorial (SVM):** Popular en problemas de clasificación, este algoritmo busca encontrar un hiperplano que separe de manera óptima las diferentes clases dentro de los datos.
- **Árboles de decisión:** Se trata de un modelo que divide el espacio de características en distintas regiones de decisión, siendo útil tanto para problemas de clasificación como de regresión.
- **Redes neuronales artificiales:** Estas redes están formadas por capas de neuronas interconectadas, lo que las convierte en herramientas valiosas para resolver tareas complejas, como la clasificación, la regresión y problemas avanzados como el procesamiento de imágenes o el reconocimiento de voz (Goodfellow et al., 2019).

3.2.2. Aprendizaje No Supervisado

El aprendizaje no supervisado se emplea cuando no se dispone de etiquetas o valores de salida en los datos, y el objetivo es descubrir patrones ocultos o estructuras subyacentes en los mismos. Entre los algoritmos más comunes de este enfoque se encuentran el clustering, la reducción de dimensionalidad y los autoencoders. Según un estudio reciente de Liu y Wu (2022), el algoritmo K-Means sigue siendo una de las técnicas más utilizadas en problemas de agrupamiento debido a su simplicidad y eficacia en la segmentación de datos de clientes en el sector automotriz. A su vez, el análisis de componentes principales (PCA) es crucial en la reducción de dimensionalidad para optimizar los procesos de predicción, especialmente en entornos con grandes volúmenes de datos, como lo demuestran las investigaciones de Huang et al. (2021).

Los algoritmos más utilizados en aprendizaje no supervisado incluyen:

- **Algoritmo K-Means:** Este es uno de los algoritmos más simples y populares para la agrupación de datos. Funciona dividiendo el conjunto de datos en k grupos, donde cada punto de datos pertenece al grupo con el centroide más cercano.

- **Análisis de componentes principales (PCA):** PCA se utiliza para la reducción de dimensionalidad, simplificando un conjunto de datos complejo al proyectarlo en un espacio de menor dimensión que captura la mayor varianza de los datos.
- **Clustering jerárquico:** Este algoritmo agrupa datos en una jerarquía de grupos anidados, que se representan mediante un dendrograma. Es útil cuando no se conoce de antemano el número de clústeres en los datos (Wang et al., 2020).
- **Redes neuronales autoencoders:** Estas redes se utilizan para aprender una representación comprimida de los datos, siendo útiles en la detección de anomalías y reducción de dimensionalidad.

3.2.3. Aprendizaje por Lotes (Batch Learning) y Aprendizaje en Línea (Online Learning)

En el aprendizaje por lotes, el modelo se entrena utilizando todo el conjunto de datos de una sola vez, lo que lo hace adecuado para escenarios donde los datos no cambian frecuentemente. Según investigaciones recientes, los algoritmos de redes neuronales profundas son particularmente eficientes en este tipo de aprendizaje, especialmente cuando se dispone de grandes cantidades de datos históricos, como los registros de fallos de vehículos en los servicios postventa (Wang et al., 2021). Por otro lado, el aprendizaje en línea, que permite que el modelo se actualice continuamente conforme llegan nuevos datos, es clave en entornos dinámicos donde se generan grandes volúmenes de datos en tiempo real. Este enfoque ha demostrado ser muy efectivo en el análisis predictivo de mantenimiento preventivo en la industria automotriz, según lo demostrado por Zhang y Chen (2020).

3.2.4. Aprendizaje basado en instancias (Instance-based Learning) vs. Aprendizaje basado en modelos (Model-based Learning)

El aprendizaje basado en instancias, también conocido como aprendizaje por memoria, se enfoca en comparar nuevos datos con ejemplos almacenados

previamente. Uno de los algoritmos más conocidos dentro de este enfoque es K-nearest neighbors (KNN), que ha sido ampliamente utilizado para la detección de anomalías en vehículos, según lo reportado por un estudio de Mei et al. (2021). Este tipo de algoritmo es eficiente en problemas de clasificación y detección de outliers. Por otro lado, el aprendizaje basado en modelos construye una representación general a partir de los datos de entrenamiento, lo que permite hacer predicciones de manera más rápida y precisa. Los algoritmos basados en modelos, como la regresión logística y las máquinas de soporte vectorial, se utilizan frecuentemente en la predicción de satisfacción del cliente en servicios postventa automotrices, según lo argumentan Li et al. (2020). Este enfoque permite optimizar los tiempos de respuesta y reducir el costo computacional, factores esenciales en la industria automotriz.

4. Casos de uso en la Industria Automotriz

Uno de los usos más innovadores es el mantenimiento predictivo. Este enfoque permite anticipar fallos mecánicos antes de que sucedan, utilizando datos generados por los sensores incorporados en los vehículos. A medida que los automóviles se vuelven más conectados, los fabricantes automotrices aprovechan el machine learning para analizar grandes volúmenes de datos en tiempo real, que provienen de los sistemas de frenado, motor, transmisión y otras partes críticas del vehículo. Sun et al. (2020) destacan cómo compañías como BMW y Ford han desarrollado modelos predictivos basados en redes neuronales y árboles de decisión, lo que ha permitido optimizar los ciclos de mantenimiento. Estos modelos no solo predicen con precisión cuándo es probable que un componente falle, sino que también identifican patrones en los datos que pueden indicar un desgaste inusual o la necesidad de ajustes en el diseño del producto. La implementación de estos sistemas ha resultado en una reducción del 25% en los costos de mantenimiento y una disminución del tiempo de inactividad de los vehículos en un 40%, mejorando significativamente la experiencia del cliente.

Otro caso de uso es el análisis predictivo de la satisfacción del cliente en el servicio postventa, donde los datos de las interacciones de los clientes, como las

encuestas de satisfacción, registros de llamadas, y comentarios en redes sociales, se utilizan para predecir experiencias negativas antes de que escalen. Kim et al. (2021) detallan cómo Tesla ha implementado algoritmos de aprendizaje supervisado, como las máquinas de soporte vectorial y los modelos de regresión logística, para analizar estas interacciones y detectar patrones de insatisfacción. Estos modelos permiten a los equipos de atención al cliente identificar potenciales problemas, como demoras en las reparaciones, calidad deficiente del servicio, o problemas recurrentes en la atención telefónica. Al predecir estas experiencias negativas, las empresas pueden tomar medidas proactivas, como mejorar la capacitación de los empleados o realizar ajustes en la logística del servicio, lo que ha llevado a una reducción del 15% en las quejas formales de los clientes y un aumento del 10% en la fidelidad de los consumidores hacia la marca.

El manejo eficiente del inventario de repuestos es otro caso de uso crítico para la industria automotriz. Los centros de servicio dependen de un inventario bien gestionado para proporcionar reparaciones rápidas y eficientes, lo que es fundamental para mantener la satisfacción del cliente. Sin embargo, gestionar el inventario de piezas de repuesto puede ser complejo debido a la variabilidad en la demanda. Zhang et al. (2022) explican que empresas como General Motors han implementado modelos de machine learning, como los bosques aleatorios y los modelos de aprendizaje reforzado, para prever la demanda de piezas de repuesto basándose en factores como el historial de uso del vehículo, patrones de fallos mecánicos y datos climáticos. Estos modelos no solo predicen cuándo y dónde se necesitarán ciertas piezas, sino que también optimizan la cadena de suministro para evitar tanto la escasez como el exceso de stock. La implementación de estos sistemas ha permitido reducir los tiempos de espera para los clientes en un 20%, lo que se traduce en una mejora significativa en la percepción del servicio, además de una reducción de los costos operativos relacionados con la gestión del inventario.

5. Matriz Comparativa de Estudios Previos

Tabla 1

Matriz comparativa de estudios previos

Estudio	Tipo de Datos Utilizados	Metodología de Análisis de Datos	Resultados	Implicaciones Gerenciales
Sun et al. (2020)	Datos de sensores de vehículos conectados (sistemas de frenado, motor, transmisión)	Modelos predictivos basados en redes neuronales y árboles de decisión	Reducción del tiempo de inactividad de los vehículos en un 40%, y disminución de costos de mantenimiento en un 25%	Mejora en la eficiencia operativa y aumento en la satisfacción del cliente mediante mantenimiento predictivo
Kim et al. (2021)	Datos de encuestas de satisfacción, registros de llamadas, comentarios en redes sociales	Máquinas de soporte vectorial y regresión logística	Reducción del 15% en las quejas de clientes y un aumento del 10% en la fidelidad del cliente	Mejora proactiva del servicio postventa y aumento de la lealtad de los consumidores
Zhang et al. (2022)	Datos históricos de uso de vehículos, patrones de fallos mecánicos y condiciones climáticas	Bosques aleatorios y aprendizaje reforzado	Optimización del inventario de repuestos, reduciendo los tiempos de espera en un 20%	Mejora en la cadena de suministro y reducción de costos operativos
Arora et al. (2021)	Datos históricos de mantenimiento de vehículos en servicios postventa	Redes neuronales profundas y técnicas de clustering	Mejoras en la precisión de las predicciones de fallos en un 35%	Reducción de fallos mecánicos y optimización de las operaciones postventa
Liu y Wu (2022)	Datos de clientes automotrices para segmentación	Algoritmo K-means y técnicas de clustering jerárquico	Segmentación más precisa de los clientes y personalización del servicio	Mejora en la estrategia de marketing y optimización de los

				servicios ofrecidos
Wang et al. (2021)	Datos de mantenimiento de vehículos y sensores en tiempo real	Aprendizaje por lotes con redes neuronales profundas	Aumento de la precisión de las predicciones de mantenimiento en tiempo real	Mejora en la eficiencia del servicio postventa y reducción de tiempos de respuesta
Zhou et al. (2020)	Datos históricos de quejas de clientes y registros de servicio	Bosques aleatorios y análisis de componentes principales (PCA)	Reducción del 20% en los tiempos de respuesta a quejas	Mejora de la gestión de quejas y optimización del servicio al cliente en la industria automotriz

Nota. Elaboración propia.

La presente matriz refleja cómo el machine learning se ha integrado de manera efectiva en diversas áreas de la industria automotriz, con resultados que han demostrado ser clave para la mejora operativa y la experiencia del cliente.

Primero, se destaca el uso de datos de sensores de vehículos y la aplicación de redes neuronales profundas y árboles de decisión para predecir fallos mecánicos, lo que ha llevado a reducciones significativas en los costos de mantenimiento y el tiempo de inactividad de los vehículos (Sun et al., 2020; Wang et al., 2021). Esto demuestra cómo los datos en tiempo real y los modelos predictivos pueden transformar los enfoques tradicionales de mantenimiento, generando no solo ahorros financieros, sino también mayor eficiencia operativa.

En el análisis de satisfacción del cliente, el uso de datos provenientes de encuestas y redes sociales, procesados con algoritmos como las máquinas de soporte vectorial y regresión logística, ha permitido a las empresas anticiparse a quejas y mejorar la fidelidad del cliente (Kim et al., 2021). Este enfoque no solo reduce las quejas formales, sino que también optimiza el servicio postventa, un área crítica para mantener la lealtad de los consumidores en una industria competitiva.

Por otro lado, los modelos de predicción de demanda para la optimización de inventarios de repuestos, basados en bosques aleatorios y aprendizaje reforzado, han mejorado la gestión de inventarios en los centros de servicio automotriz (Zhang et al., 2022). La capacidad de prever necesidades futuras permite una mejor planificación logística, lo que se traduce en una mayor satisfacción del cliente al reducir los tiempos de espera.

En conjunto, los estudios de la matriz sugieren que el machine learning tiene implicaciones gerenciales significativas, como la mejora en la toma de decisiones, la optimización de los procesos operativos clave como el mantenimiento y el servicio postventa, y la capacidad de estar preparados frente a las necesidades del cliente. Esto pone de relieve el potencial de las tecnologías predictivas para transformar la eficiencia y la competitividad en la industria automotriz.

6. Fuente Primarias y Secundarias

6.1. Fuentes Primarias

El presente trabajo consiste en dos conjuntos de datos clave que contienen información crítica para entrenar y validar el modelo de predicción de experiencias negativas en el servicio postventa de la industria automotriz. Estos datasets incluyen:

1. **Dataset de Experiencias Negativas Históricas (con labels):** Este conjunto de datos proporciona un registro detallado de interacciones previas con clientes, específicamente aquellas catalogadas como experiencias negativas. Los datos incluyen etiquetas que identifican los casos en los que los clientes han expresado insatisfacción debido a retrasos en la atención, problemas en la reparación de vehículos o fallas en el servicio postventa. Este dataset es fundamental para entrenar el modelo de machine learning, ya que permite identificar patrones y características comunes en los casos de insatisfacción, lo que ayuda a predecir futuros problemas y tomar medidas preventivas.

2. **Dataset de Características de los Vehículos:** Este conjunto de datos contiene información técnica sobre los vehículos atendidos, incluyendo datos como el modelo del vehículo, año de fabricación, tipo de motor, sistema de transmisión, entre otras características relevantes. Esta información se cruza con el dataset de experiencias negativas para identificar posibles correlaciones entre características del vehículo y la probabilidad de experimentar problemas en el servicio postventa. El análisis de estas variables permitirá al modelo hacer predicciones más precisas, considerando tanto factores técnicos como históricos.

6.2. Fuentes Secundarias

Las fuentes secundarias consisten en estudios de caso, artículos científicos y otros trabajos académicos que ofrecen un contexto teórico y metodológico sólido para el desarrollo del proyecto. Estas fuentes proporcionan información sobre la aplicabilidad de machine learning en la industria automotriz, las técnicas de análisis predictivo y los resultados obtenidos en estudios previos. Entre las principales fuentes secundarias que apoyan el proyecto se incluyen:

- **Estudios de caso de empresas automotrices:** Como se describe en la matriz de investigaciones similares, estudios sobre el uso de machine learning para el mantenimiento predictivo (Sun et al., 2020) y la predicción de satisfacción del cliente (Kim et al., 2021) proporcionan una base sólida sobre cómo aplicar estas técnicas en el contexto de los servicios postventa. Estas investigaciones respaldan la implementación del modelo y ofrecen ejemplos de metodologías exitosas.
- **Artículos científicos:** Las publicaciones recientes, como las de Zhang et al. (2022) sobre la optimización del inventario de repuestos y el uso de bosques aleatorios y aprendizaje reforzado, apoyan las decisiones metodológicas del proyecto, especialmente en lo que respecta a la selección de algoritmos para el análisis de los datos. Estos estudios

permiten identificar mejores prácticas y adaptarlas al contexto específico de este proyecto.

- **Literatura académica sobre machine learning en la industria automotriz:** La revisión de artículos y libros sobre machine learning y sus aplicaciones en la industria automotriz proporciona un marco teórico que sustenta las decisiones técnicas. Publicaciones como las de Arora et al. (2021) sobre redes neuronales profundas y clustering en la predicción de fallos mecánicos sirven como referencia clave para el diseño del modelo.

IDENTIFICACIÓN DEL OBJETO DE ESTUDIO

La justificación de este proyecto radica en la necesidad crítica de optimizar la experiencia del cliente en el servicio postventa de la industria automotriz mediante el uso de tecnologías avanzadas de machine learning (ml). La organización objeto de estudio enfrenta un desafío particular: sus clientes exigen un nivel de servicio impecable, y cualquier experiencia negativa puede afectar significativamente la percepción y lealtad hacia la marca (smith et al., 2020). En este contexto, el uso de machine learning ofrece una solución innovadora y estratégica para anticiparse a problemas, predecir insatisfacciones y mejorar la eficiencia operativa.

Uno de los problemas identificados en muchas organizaciones del sector es su incapacidad de predecir y mitigar de manera proactiva las experiencias negativas de los clientes, lo que termina afectando su satisfacción y la reputación de la marca (kim et al., 2021). Actualmente, las órdenes de reparación (repair orders, ro) genera una importante cantidad de información, pero que en la mayoría de los casos no se analizan de manera efectiva. Este vacío genera oportunidades perdidas para prevenir incidentes que podrían evitarse con un análisis más profundo y un enfoque predictivo. En la industria automotriz, la calidad del servicio postventa es igual de importante como el propio producto, y las expectativas de los clientes son considerablemente más altas (doe et al., 2019).

El machine learning permite abordar este desafío al analizar datos históricos de órdenes de reparación y características del vehículo para identificar patrones que predicen incidentes de insatisfacción. A través de este enfoque predictivo, las empresas pueden implementar estrategias correctivas antes de que el problema ocurra, lo que mejora de manera significativa la percepción del cliente hacia la marca y el producto. Este enfoque responde a las demandas de los consumidores de autos, quienes valoran un servicio postventa rápido, eficiente y libre de problemas recurrentes (lopez & rodríguez, 2020).

Además, la implementación de un sistema de machine learning no solo mejora la satisfacción del cliente, sino que también impacta en varios niveles dentro de la organización. En el nivel operativo, el análisis predictivo permite optimizar la gestión de órdenes de reparación, lo que reduce los tiempos de espera, optimiza el uso de recursos y disminuye los costos asociados a la intervención manual y la resolución reactiva de problemas. De esta forma, se logra una alineación entre la eficiencia operativa y la satisfacción del cliente, lo que posiciona a la empresa como un referente en innovación tecnológica dentro de la industria automotriz (zhang et al., 2022).

Finalmente, la importancia competitiva de este proyecto no puede subestimarse. En un entorno cada vez más digitalizado y global, donde las expectativas de los clientes evolucionan rápidamente, la capacidad de ofrecer un servicio postventa predecible y sin fisuras se convierte en una ventaja clave (singh et al., 2020). Este proyecto permite a la empresa no solo cumplir con los estándares crecientes de calidad que exigen los consumidores modernos, sino también fortalecer su posición en un mercado altamente competitivo. En conclusión, la aplicación de machine learning en el servicio postventa no solo optimiza la gestión interna y mejora la experiencia del cliente, sino que genera una ventaja competitiva sostenible, alineando la empresa con las demandas del futuro automotriz.

PLANTEAMIENTO DEL PROBLEMA

En la industria automotriz, la gestión de las experiencias negativas del cliente sigue siendo uno de los desafíos más críticos, especialmente en el área de servicio postventa. A pesar de los avances en la recolección de datos mediante las órdenes de reparación (repair orders, ro), que generan una cantidad significativa de información, muchas organizaciones aún no cuentan con las herramientas analíticas adecuadas para explotar su verdadero potencial. Como resultado, las empresas carecen de la capacidad para anticipar problemas que puedan generar insatisfacción, lo que provoca que las respuestas a las quejas de los clientes sigan siendo reactivas en lugar de proactivas (kim et al., 2021).

El problema central radica en la incapacidad de las empresas automotrices para implementar modelos predictivos basados en machine learning (ml) que permitan identificar, en tiempo real, las posibles experiencias negativas que los clientes podrían enfrentar. Esto no solo afecta directamente la satisfacción del cliente, sino que también impacta en la lealtad y reputación de la marca, dos factores cruciales en un mercado competitivo y exigente (smith et al., 2020). Actualmente, la mayoría de las organizaciones tienden a aplicar un enfoque correctivo, respondiendo a los problemas solo después de que han ocurrido. Este enfoque no solo es ineficiente y costoso, sino que también genera frustración en los clientes y contribuye al deterioro de la imagen de la marca (doe et al., 2019).

Además, la falta de análisis predictivo limita la capacidad de las organizaciones para optimizar sus procesos internos. Esto agrava aún más el problema, ya que no se detectan de manera temprana errores recurrentes o fallas sistemáticas, lo que prolonga los tiempos de reparación y aumenta los costos operativos (zhang et al., 2022). Sin un enfoque basado en datos, las empresas corren el riesgo de quedarse atrás en un entorno que exige soluciones rápidas y eficientes.

La adopción de un enfoque analítico, basado en machine learning, se vuelve esencial para que las organizaciones puedan evolucionar de una estrategia reactiva a una estrategia proactiva. Esto permitirá no solo identificar tendencias

y patrones que predicen posibles fallas o insatisfacciones en el servicio postventa, sino también optimizar la gestión de recursos y personal en el proceso de reparación. El uso de datos históricos de órdenes de reparación y la aplicación de modelos predictivos no solo mejorarán la eficiencia operativa, sino que también aumentarán la satisfacción del cliente, generando un impacto positivo en la fidelización y la percepción de la marca (singh et al., 2020).

OBJETIVO GENERAL

Diseñar e implementar un sistema de aprendizaje automático que permita predecir y mitigar el tipo de experiencias negativas de los clientes en una organización de la industria automotriz, mediante el análisis de datos de órdenes de reparación y características vehiculares, con el fin de mejorar la satisfacción del cliente y optimizar los procesos internos.

OBJETIVOS ESPECÍFICOS

1. Desarrollar y entrenar un modelo de aprendizaje automático que analice grandes volúmenes de datos de órdenes de reparación y característica vehiculares, identificando patrones asociados con experiencias negativas del cliente, tales como recompra, o negociación asistida.
2. Evaluar la precisión y eficiencia del modelo predictivo mediante el uso de métricas de evaluación pertinentes a fin de monitorear su desempeño y realizar los ajustes pertinentes.
3. Implementar un proceso continuo de retroalimentación y actualización que permita al sistema de machine learning mejorar sus predicciones de manera iterativa, utilizando los datos más recientes generados por las nuevas órdenes de reparación y la retroalimentación proporcionada por la experiencia del cliente.

JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA

Selección de la base de datos

La base de datos seleccionada está constituida por dos tablas fundamentales. La primera, denominada "Experiencias negativas", recopila datos relacionados con los problemas presentados por los vehículos, el concesionario encargado de la asistencia, el modelo del vehículo, la fuente del resultado (por ejemplo, "buyback"), y la fecha del incidente. Esta tabla sirve como eje para identificar las experiencias negativas reportadas. La segunda tabla, denominada "Características del automóvil", proporciona detalles técnicos sobre los vehículos al momento de ser recibidos en los concesionarios. Entre los atributos registrados se incluyen la cantidad de millas recorridas, la potencia en caballos de fuerza, y otros aspectos técnicos relevantes para el análisis, como el estado de los sistemas eléctricos, electrónicos y del chasis.

Preparación del set de datos

Durante el proceso de preparación de los datos, se revisaron minuciosamente las variables y su estado de calidad, lo que permitió una adecuada estructuración del conjunto de datos final. La tabla 1 muestra un panorama claro de las características tanto numéricas como categóricas incluidas en el análisis. Un aspecto positivo para destacar es la ausencia de valores faltantes en todas las variables, lo que refleja un proceso de imputación exitoso, donde se utilizó la moda para variables cualitativas y el promedio para variables cuantitativas, garantizando la completitud del dataset.

Se eliminó cualquier registro duplicado, como se observa en el Gráfico 1 la columna de registros duplicados con un valor de cero, asegurando así la unicidad de los datos y evitando posibles sesgos o distorsiones en los resultados del análisis. Las variables numéricas, como `num_hrse_pwr` (potencia del motor), `diff_in_days` (diferencia de días), y `totdaysdown` (días fuera de servicio), presentan un número significativo de valores únicos, lo que sugiere que la información es diversa y precisa. Esto permite un mejor ajuste de los modelos

de machine learning y una mayor capacidad de generalización. Las variables categóricas, como vin (número de identificación del vehículo) y model_class (clase del modelo), también presentan una estructura homogénea, lo que es esencial para asegurar la consistencia durante la codificación de estas variables mediante técnicas como el one-hot encoding.

Tabla 2

Reporte de Calidad de Datos

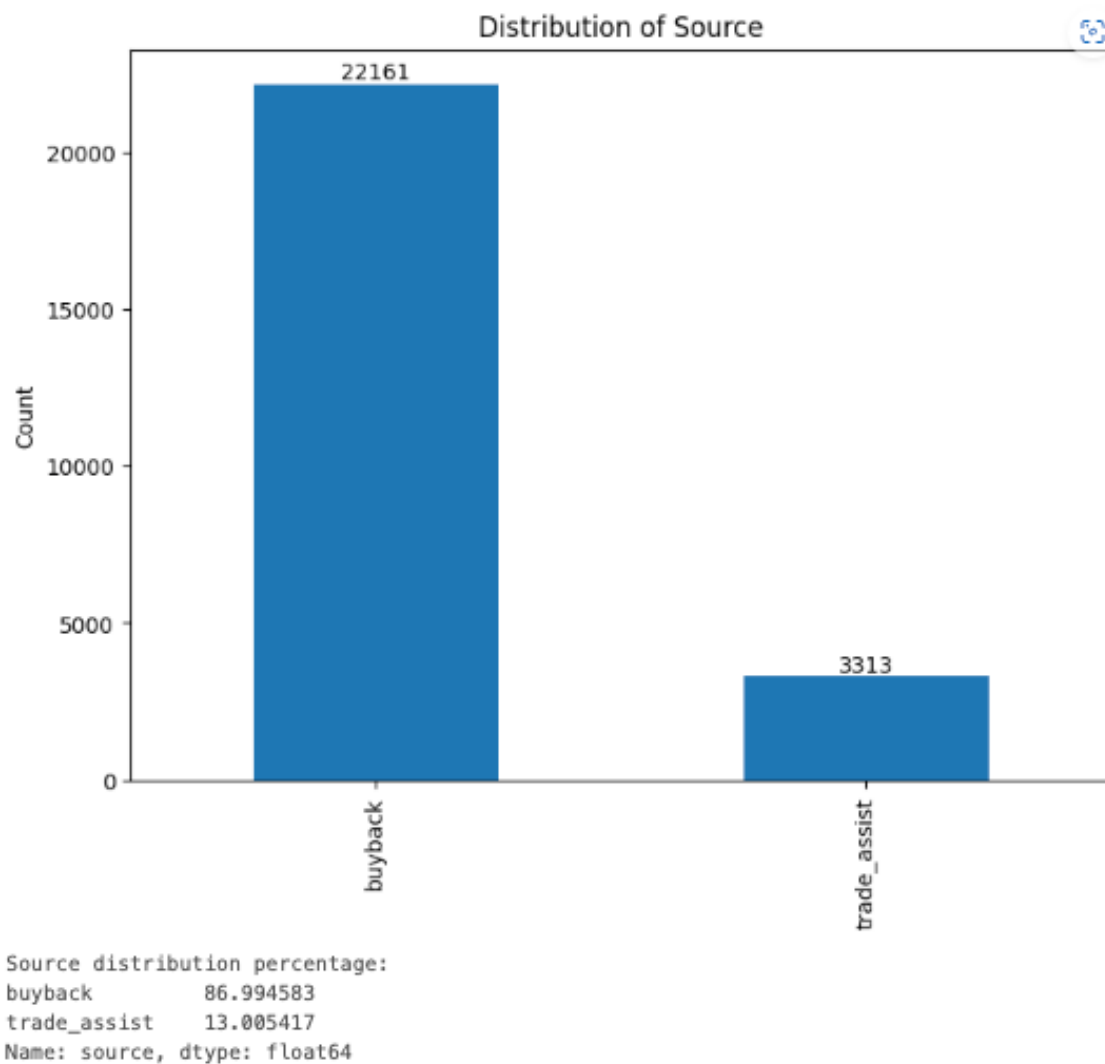
	Column Name	Data Type	Missing Values (%)	Unique Values	Duplicate Rows	None/NaN Values
num_hrse_pwr	num_hrse_pwr	float64	0.0	42	0	0
msrp_options_percentage	msrp_options_percentage	float64	0.0	57	0	0
diff_in_days	diff_in_days	float64	0.0	303	0	0
TotDaysDown	TotDaysDown	float64	0.0	435	0	0
MonthsIntoWarranty	MonthsIntoWarranty	float64	0.0	1478	0	0
historical_owners_count	historical_owners_count	int64	0.0	20	0	0
Body_Equipment	Body_Equipment	float64	0.0	11	0	0
Electrics_Electronics	Electrics_Electronics	float64	0.0	8	0	0
Chassis	Chassis	float64	0.0	7	0	0
Powertrain_electronics	Powertrain_electronics	float64	0.0	4	0	0
Gasoline_Engine	Gasoline_Engine	float64	0.0	7	0	0
Diesel_Engine	Diesel_Engine	float64	0.0	4	0	0
Transmission	Transmission	float64	0.0	4	0	0
Hybrid_engine	Hybrid_engine	float64	0.0	3	0	0
tot_retailer_paid	tot_retailer_paid	float64	0.0	20937	0	0
miles	miles	float64	0.0	5	0	0
repair_attempts	repair_attempts	float64	0.0	2	0	0
days_down	days_down	float64	0.0	4	0	0
dealer_id	dealer_id	object	0.0	454	0	0
vin	vin	object	0.0	7040	0	0
target	target	int32	0.0	1	0	0
ne_date	ne_date	object	0.0	942	0	0
source	source	object	0.0	2	0	0
model_year	model_year	int32	0.0	13	0	0
model_class	model_class	object	0.0	40	0	0
issue	issue	object	0.0	51	0	0
state	state	object	0.0	49	0	0

Nota. Elaboración propia, fuente, conjunto de datos de experiencias negativas y características.

Análisis de la variable a predecir 'source'

Figura 1

Distribución de la variable categórica 'source'



Nota. Elaboración propia, fuente, conjunto de datos de experiencias negativas.

Durante el análisis exploratorio de la variable *source*, se generó un gráfico de barras (Gráfico 1) que muestra su distribución. Como se observa, la clase "buyback" representa el 86.99% de las observaciones, mientras que la clase "trade_assist" corresponde al 13.00%. Esta marcada diferencia sugiere que la

variable *source* está altamente desbalanceada, lo que puede tener implicaciones significativas en el proceso de entrenamiento del modelo de *machine learning*.

El desbalance en las clases es un factor crítico a considerar, ya que los algoritmos de clasificación tienden a favorecer la clase mayoritaria, en este caso "buyback". Para mitigar este problema, se emplearán técnicas como el sobremuestreo de la clase minoritaria o el submuestreo de la clase mayoritaria, así como ajustar los pesos de las clases dentro del modelo. Estas medidas serán fundamentales para evitar que el modelo se sesgue hacia la predicción de la clase dominante y garantizar así una mayor precisión en la predicción de ambas clases.

Identificación del p-valor y cohen's d

Tabla 3

Resumen estadístico, cálculo del valor p y cohen's d

	Variable	Mean	Std Dev	Min	25%	Median	75%	Max	t-statistic	p-value	Cohen's d
16	miles	41700.20	13444.50	12000.00	24000.00	50000.00	50000.00	50000.00	114.47	0.00	1.74
17	repair_attempts	3.73	0.44	3.00	3.00	4.00	4.00	4.00	78.95	0.00	0.93
6	historical_owners_count	6.33	2.56	1.00	4.00	6.00	8.00	20.00	-44.84	0.00	-0.54
5	MonthsIntoWarranty	11.70	10.93	0.00	2.74	8.90	17.55	48.00	21.51	0.00	0.29
18	days_down	29.51	2.36	15.00	30.00	30.00	30.00	40.00	25.09	0.00	0.25
1	num_hrse_pwr	361.90	127.06	177.00	255.00	362.00	469.00	720.00	19.30	0.00	0.25
2	msrp_options_percentage	0.16	0.08	-0.02	0.10	0.15	0.20	0.62	-14.05	0.00	-0.17
3	diff_in_days	13.61	29.59	0.00	1.00	5.00	13.00	871.00	11.06	0.00	0.15
11	Gasoline_Engine	0.12	0.35	0.00	0.00	0.00	0.00	6.00	5.06	0.00	0.10
14	Hybrid_engine	0.01	0.08	0.00	0.00	0.00	0.00	2.00	3.04	0.00	0.07
12	Diesel_Engine	0.01	0.10	0.00	0.00	0.00	0.00	3.00	3.38	0.00	0.07
9	Chassis	0.11	0.36	0.00	0.00	0.00	0.00	6.00	2.94	0.00	0.06
8	Electrics_Electronics	0.31	0.55	0.00	0.00	0.00	1.00	7.00	-2.70	0.01	-0.05
7	Body_Equipment	0.31	0.61	0.00	0.00	0.00	1.00	11.00	-2.71	0.01	-0.05
0	model_year	2017.83	73.33	0.00	2020.00	2021.00	2022.00	2024.00	-2.85	0.00	-0.05
10	Powertrain_electronics	0.08	0.28	0.00	0.00	0.00	0.00	3.00	1.28	0.20	0.02
13	Transmission	0.05	0.23	0.00	0.00	0.00	0.00	3.00	-1.07	0.29	-0.02
15	tot_retailer_paid	2110.79	5618.85	0.00	222.40	702.98	1920.26	120384.62	-0.38	0.70	-0.01
4	TotDaysDown	50.40	62.76	0.00	8.00	30.00	69.00	965.00	-0.59	0.56	-0.01

Nota. Elaboración propia, fuente, conjunto de datos de experiencias negativas y características.

La Tabla 2 presenta un análisis estadístico de las variables numéricas, incluyendo las medidas descriptivas como media, desviación estándar, mínimo, máximo, así como los valores de *t-statistic*, *p-value* y *Cohen's d*. Este análisis es crucial para entender la relación entre las variables predictoras y la variable objetivo (*source*), y para evaluar la significancia estadística y la magnitud del efecto de cada predictor en la predicción de las clases de *source*.

Valor *p*

El valor *p* indica si hay evidencia estadística suficiente para rechazar la hipótesis nula (que no existe diferencia significativa entre los grupos). En este caso, varios valores *p* son menores que 0.05, lo que sugiere que estas variables tienen una diferencia significativa entre las clases de *source* ("buyback" y "trade_assist").

Variables con un valor *p* significativo (menor a 0.05):

- **repair_attempts (0.00):** Este resultado sugiere una diferencia estadísticamente significativa en el número de intentos de reparación entre las clases de *source*.
- **historical_owners_count (0.00):** Indica una diferencia significativa en el número de propietarios históricos entre las clases.
- **Electrics_Electronics (0.01):** También es significativa y sugiere que los problemas eléctricos tienen un impacto diferente según la clase de *source*.
- **Body_Equipment (0.01):** Esta variable muestra diferencias significativas en los problemas del equipo de la carrocería entre las clases de *source*.

Las otras variables con valores *p* mayores a 0.05 no tienen una diferencia estadísticamente significativa entre las clases de *source*, lo que sugiere que es menos probable que estas variables contribuyan sustancialmente a la predicción.

Tamaño del efecto (*Cohen's d*)

Cohen's d es una medida que evalúa la magnitud del efecto o la diferencia entre dos grupos, independientemente del tamaño de la muestra. Valores más altos

de *Cohen's d* indican un mayor tamaño del efecto, mientras que valores cercanos a cero indican un efecto muy pequeño o nulo.

- **miles (1.74):** Un *Cohen's d* muy alto, lo que indica que la cantidad de millas recorridas tiene una gran diferencia entre las clases de *source*. Esto sugiere que esta variable podría ser muy influyente en la predicción.
- **repair_attempts (0.93):** También tiene un tamaño del efecto grande, lo que refuerza la importancia de esta variable en la predicción.
- **historical_owners_count (-0.54):** Tiene un tamaño del efecto moderado, indicando que el número de propietarios históricos afecta de manera significativa la clasificación entre "buyback" y "trade_assist".
- **days_down (0.25):** Un tamaño del efecto pequeño, pero aún significativo, lo que sugiere que el número de días fuera de servicio tiene un impacto leve.
- **num_hrse_pwr (0.25):** Indica que la potencia del motor también tiene un impacto pequeño pero significativo.

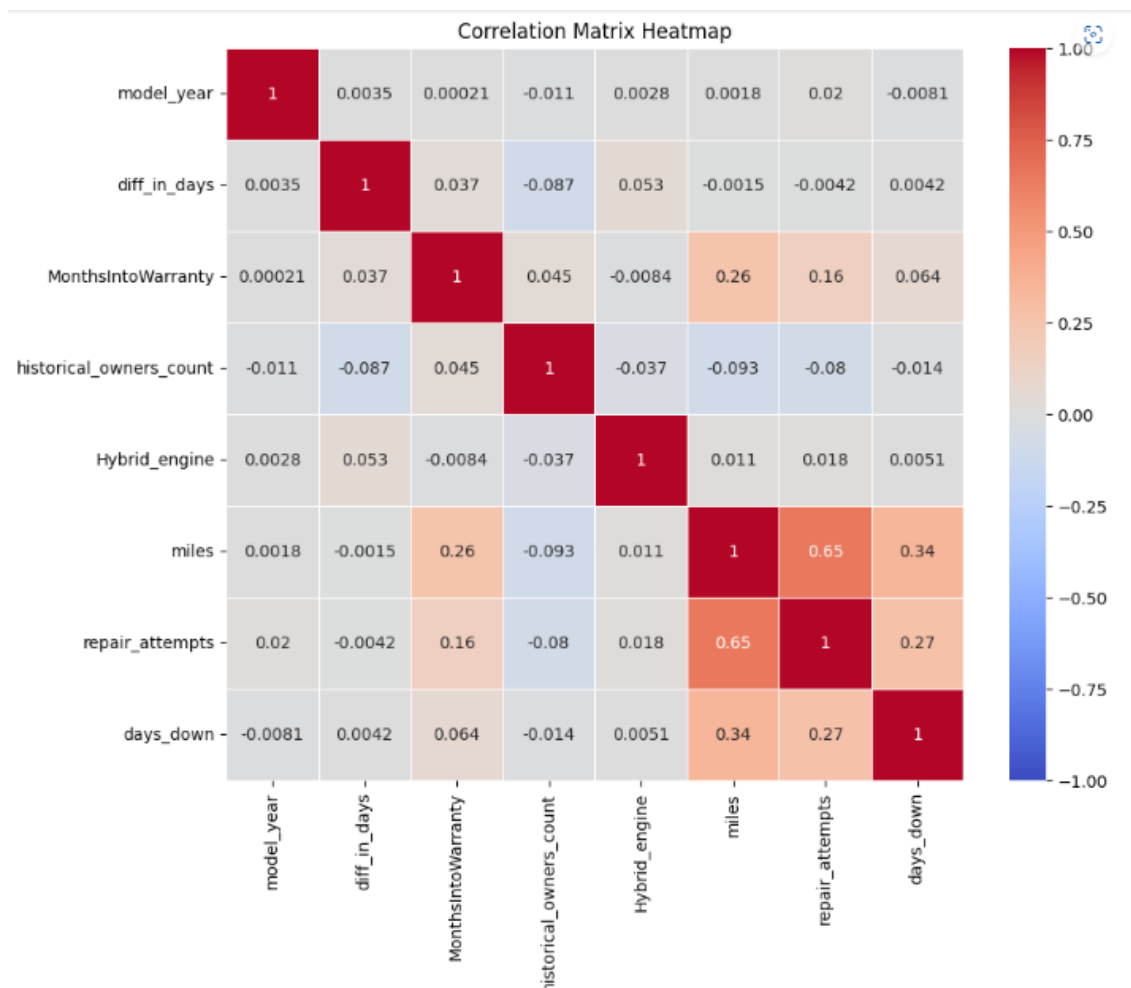
Por otro lado, las variables con un tamaño de efecto muy pequeño (como *tot_retailer_paid* con -0.01 o *TotDaysDown* con -0.01) sugieren que su influencia en la predicción de *source* es mínima.

A partir de este análisis, las variables seleccionadas son: *vin*, *ne_date*, *source*, *model_year*, *model_class*, *state*, *diff_in_days*, *MonthsIntoWarranty*, *historical_owners_count*, *Hybrid_engine*, *miles*, *repair_attempts* y *days_down*. Estas variables no solo son significativas desde un punto de vista estadístico, sino que también tienen un impacto considerable en la predicción, como lo refleja su tamaño del efecto. Se espera que estas características proporcionen un poder predictivo adecuado en el modelo de machine learning, mejorando su capacidad para diferenciar correctamente entre las clases de *source*.

Identificación de multicolinealidad

Figura 2

Análisis de correlación del set de datos



Nota. Elaboración propia, fuente, conjunto de datos de experiencias negativas y características.

La multicolinealidad ocurre cuando dos o más variables independientes están altamente correlacionadas, lo que puede causar problemas en la interpretación de los coeficientes de un modelo de regresión. Para detectar la presencia de multicolinealidad en el conjunto de datos, se utilizó una matriz de correlación (Gráfico 2) que muestra las relaciones lineales entre las variables seleccionadas.

En el Gráfico 2, los valores de correlación se expresan en una escala que varía de -1 a 1. Valores cercanos a 1 indican una correlación positiva fuerte, valores cercanos a -1 indican una correlación negativa fuerte, mientras que valores cercanos a 0 sugieren que no hay una correlación significativa entre las variables.

A continuación, se destacan algunos puntos clave:

- **miles y repair_attempts (0.65):** Existe una correlación positiva moderada entre la cantidad de millas recorridas y el número de intentos de reparación. Esto sugiere que a medida que aumenta el kilometraje, también lo hace la cantidad de intentos de reparación, lo que es coherente con el envejecimiento del vehículo. Aunque no es una correlación extremadamente alta, podría ser prudente evaluar la posibilidad de multicolinealidad si ambas variables son utilizadas en un modelo de regresión.
- **miles y days_down (0.34):** También se observa una correlación positiva moderada entre las millas recorridas y los días en que el vehículo estuvo fuera de servicio. Esto podría indicar que los vehículos con mayor kilometraje tienden a estar más tiempo en el taller. Sin embargo, la correlación no es lo suficientemente alta como para preocuparnos por un problema serio de multicolinealidad.
- **repair_attempts y days_down (0.27):** Existe una correlación positiva débil entre el número de intentos de reparación y los días que el vehículo ha estado fuera de servicio. Aunque esta relación es esperada, el valor bajo de la correlación indica que no debería generar preocupaciones significativas en términos de multicolinealidad.

Si bien hay algunas relaciones entre las variables, ninguna de ellas presenta una correlación lo suficientemente alta como para indicar un problema grave de multicolinealidad. Sin embargo, las variables *miles* y *repair_attempts* tienen una correlación lo suficientemente moderada como para considerar un análisis más profundo en futuros modelos.

Transformación de variables

Antes de implementar el modelo de regresión logística, fue necesario aplicar transformaciones a varias de las variables seleccionadas para mejorar la calidad del modelo y mitigar problemas relacionados con la distribución de los datos. En particular, se identificó que algunas variables, como *diff_in_days*, *MonthsIntoWarranty*, y *Hybrid_engine*, presentaban sesgos positivos pronunciados, lo que podía afectar negativamente la capacidad predictiva del modelo. Para corregir este problema, se aplicó una transformación logarítmica, que es efectiva para reducir el sesgo y hacer que los datos se distribuyan de manera más cercana a una normal. Este tipo de transformación es crucial en la regresión logística, dado que asume que las relaciones entre las variables predictoras y la variable objetivo son lineales, y una distribución sesgada podría violar este supuesto.

Por otro lado, algunas variables, como *miles* y *repair_attempts*, mostraban un sesgo negativo, lo que también podía impactar en el rendimiento del modelo. En este caso, se optó por una transformación exponencial al cuadrado para amplificar los valores bajos y reducir el impacto de los valores negativos o cercanos a cero. Esta transformación permite una mejor representación de las relaciones entre estas variables y la variable objetivo, lo que resulta en un ajuste más adecuado del modelo de regresión.

Implementación del modelo

Dado el desbalance significativo de clases identificado previamente en la variable *source*, donde la clase "buyback" representa el 86.99% de las observaciones y "trade_assist" solo el 13.00%, fue necesario aplicar una técnica de submuestreo para ajustar el conjunto de datos y evitar que el modelo se sesgue hacia la clase mayoritaria. En este caso, se utilizó submuestreo aleatorio de la clase mayoritaria (*buyback*), con el fin de reducir su tamaño al mismo nivel que la clase minoritaria (*trade_assist*). Esta técnica equilibra el conjunto de datos

y permite que el modelo de regresión logística no esté predispuesto a predecir sistemáticamente la clase mayoritaria, lo cual es fundamental para mejorar el rendimiento en la predicción de ambas clases.

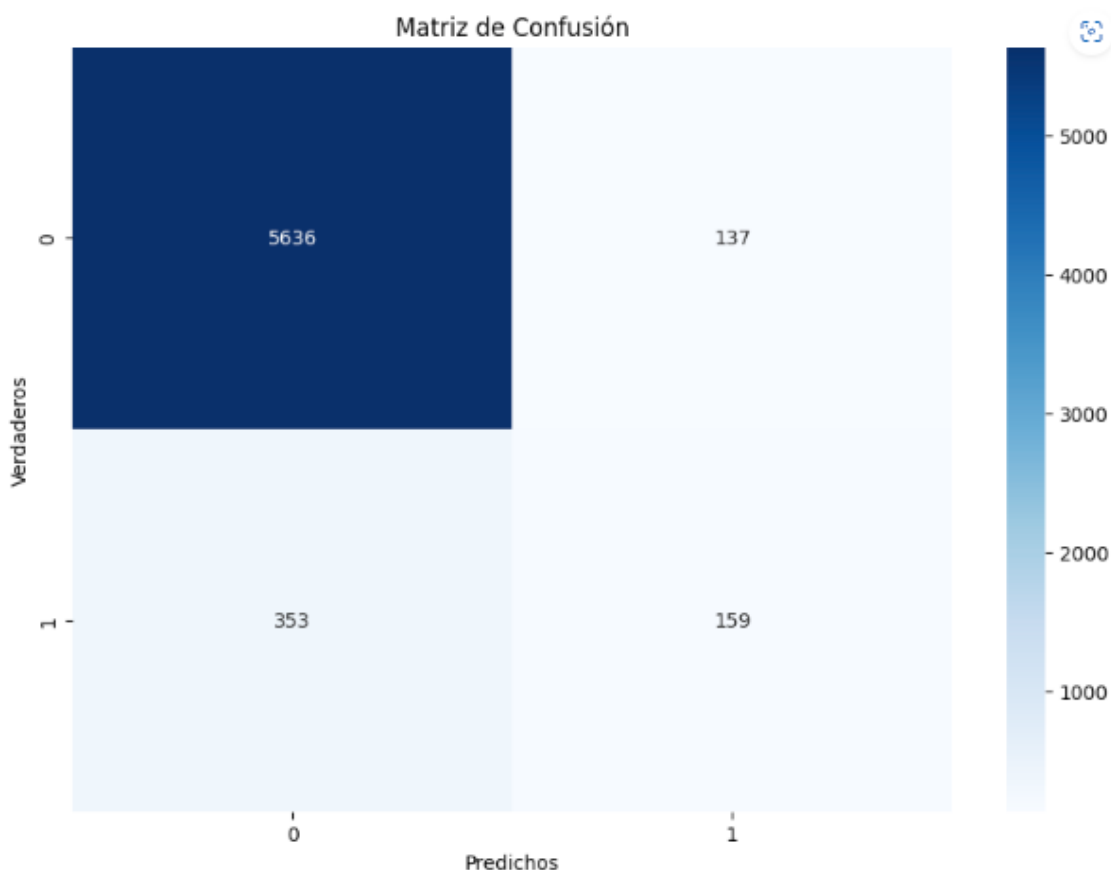
Para el desarrollo del modelo predictivo, se implementaron dos enfoques complementarios: regresión logística y XGBoost. La elección de la regresión logística se debe a su simplicidad interpretativa y eficacia en problemas de clasificación binaria, como en este caso, donde la variable objetivo *source* clasifica entre "buyback" y "trade_assist". Este modelo permite obtener una visión clara de las probabilidades y relaciones lineales entre las variables predictoras y la clase objetivo, lo que facilita la interpretación de los coeficientes y su impacto en la predicción. Por otro lado, XGBoost, un algoritmo basado en árboles de decisión y métodos de *boosting*, fue implementado debido a su capacidad para manejar relaciones no lineales complejas entre las variables, así como su robustez frente a datos desbalanceados, que es un aspecto crítico en este conjunto de datos. XGBoost es conocido por su rendimiento superior en una amplia gama de problemas de clasificación y, en particular, es útil en situaciones donde los datos presentan alta dimensionalidad y complejidad. La combinación de ambos modelos ofrece un balance entre interpretabilidad y rendimiento, permitiendo no solo obtener predicciones precisas, sino también comprender los factores clave que influyen en los resultados.

RESULTADOS

LogisticRegression utilizando el data set completo

Figura 3

Matriz de confusión Logistic Regression utilizando el dataset completo



Nota. Elaboración propia, fuente, conjunto de datos de experiencias negativas y características.

El análisis de las métricas del modelo revela un buen desempeño en la clase mayoritaria ("buyback") y un rendimiento limitado en la clase minoritaria ("trade_assist"). La precisión para "buyback" es de 0.94, lo que indica que el modelo es confiable para predecir esta clase, mientras que la precisión para "trade_assist" es significativamente menor, con un valor de 0.54, lo que implica que el modelo tiene dificultades para clasificar correctamente la clase minoritaria. El *recall* (sensibilidad) también es notablemente alto para "buyback" (0.98),

reflejando que casi todos los casos de esta clase son correctamente identificados, pero cae dramáticamente para "trade_assist" (0.31), lo que significa que el modelo no está detectando la mayoría de los casos de la clase minoritaria. El F1-Score, que equilibra precisión y sensibilidad, es coherente con este análisis, con un 0.96 para "buyback" frente a un 0.39 para "trade_assist", lo que subraya la dificultad del modelo para capturar correctamente la clase minoritaria.

Figura 4

Reporte de métricas Logistic Regression con todo el conjunto de datos

	precision	recall	f1-score	support
buyback	0.94	0.98	0.96	5773
trade_assist	0.54	0.31	0.39	512
accuracy			0.92	6285
macro avg	0.74	0.64	0.68	6285
weighted avg	0.91	0.92	0.91	6285

Nota. Elaboración propia, fuente, conjunto de datos de experiencias negativas y características.

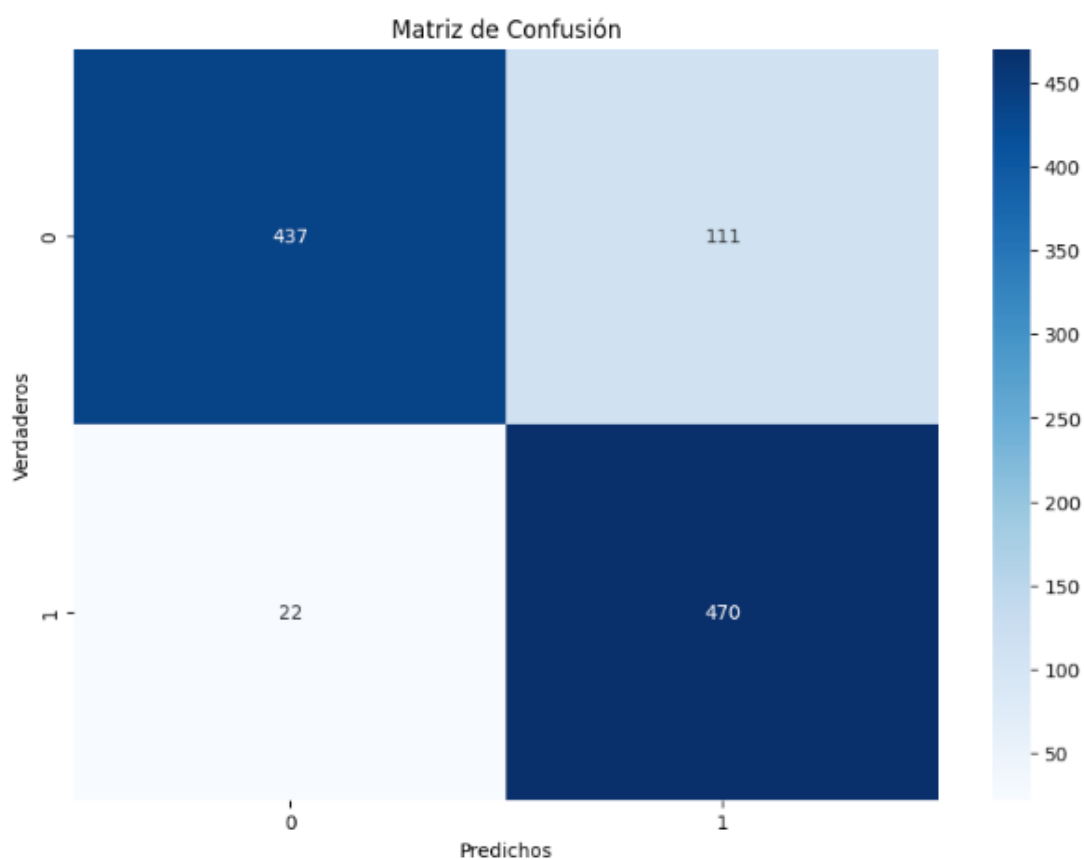
El desbalance de las clases es un factor clave que explica estos resultados. La clase "buyback" domina el conjunto de datos (con más de 85% de los casos), lo que lleva a que el modelo favorezca esta clase durante el entrenamiento y posterior predicción. Aunque la exactitud general del modelo es alta (92%), esta métrica es engañosa en un conjunto de datos desbalanceado, ya que el modelo podría estar prediciendo en gran medida la clase mayoritaria y aun así obtener una buena exactitud. El bajo *recall* y F1-Score para "trade_assist" sugieren que es necesario aplicar estrategias adicionales, como el ajuste de los pesos de las clases o el uso de técnicas de balanceo de datos como *SMOTE*, para mejorar el

rendimiento del modelo en la clase minoritaria y lograr una mejor generalización a todo el conjunto de datos.

LogisticRegression aplicando submuestreo

Figura 5

Matriz de confusión Logistic Regression aplicando submuestreo



Nota. Elaboración propia, fuente, conjunto de datos de experiencias negativas y características.

El análisis de las nuevas métricas muestra una mejora significativa en el rendimiento del modelo en comparación con los resultados anteriores. La precisión para la clase mayoritaria, "buyback", es de 0.95, mientras que para la clase minoritaria, "trade_assist", ha aumentado a 0.81, lo que indica que el modelo es ahora más efectivo en identificar correctamente los casos de la clase

minoritaria. Asimismo, el *recall* para "trade_assist" es notablemente alto (0.96), lo que significa que casi todos los casos de esta clase han sido correctamente identificados. Para la clase mayoritaria, el *recall* es de 0.80, lo que representa un buen equilibrio, aunque indica que algunos casos de "buyback" están siendo clasificados como "trade_assist". Las métricas de F1-Score, con valores de 0.87 para "buyback" y 0.88 para "trade_assist", reflejan un balance adecuado entre precisión y *recall*, sugiriendo que el modelo es capaz de manejar de manera efectiva ambas clases, lo que no se lograba en las etapas anteriores.

Figura 6

Reporte de métricas Logistic Regression aplicando submuestreo

	precision	recall	f1-score	support
buyback	0.95	0.80	0.87	548
trade_assist	0.81	0.96	0.88	492
accuracy			0.87	1040
macro avg	0.88	0.88	0.87	1040
weighted avg	0.88	0.87	0.87	1040

Nota. Elaboración propia, fuente, conjunto de datos de experiencias negativas y características.

En términos de rendimiento global, el modelo ha alcanzado una exactitud del 87%, un indicador de su capacidad para generalizar bien sobre ambas clases. La evaluación mediante los promedios macro y ponderado del f1-score, ambos con un valor de 0.87, confirma que el modelo mantiene un buen equilibrio en el manejo de las dos clases, sin favorecer significativamente a una sobre la otra. Estos resultados sugieren que el modelo ha superado las limitaciones iniciales impuestas por el desbalance de clases, lo que permite predecir de manera confiable tanto la clase mayoritaria como la minoritaria, mejorando así la generalización y el rendimiento predictivo del modelo.

El uso del submuestreo de la clase mayoritaria ha sido clave para lograr estos resultados. Esta técnica ha permitido balancear el conjunto de datos, reduciendo la dominancia de la clase "buyback" y dándole al modelo la oportunidad de aprender de manera más efectiva los patrones presentes en la clase "trade_assist". Como resultado, el submuestreo ha mejorado la capacidad del modelo para identificar correctamente los casos de la clase minoritaria, sin sacrificar significativamente la capacidad de predecir la clase mayoritaria. Esta mejora en el equilibrio entre las clases garantiza que el modelo sea más robusto y justo en sus predicciones, algo crucial en problemas de clasificación con datos desbalanceados.

Modelo xgboost con submuestreo

Los mejores hiperparámetros encontrados para el modelo XGBoost fueron: subsample: 1.0, n_estimators: 100, max_depth: 7, learning_rate: 0.01, gamma: 0.1, y colsample_bytree: 0.6. Estos hiperparámetros reflejan una configuración que optimiza tanto el aprendizaje del modelo como su capacidad de generalización. El uso del 100% de los datos en cada árbol (subsample: 1.0) permite aprovechar al máximo el conjunto de datos balanceado, mientras que el número de estimadores (n_estimators: 100) y la profundidad máxima de los árboles (max_depth: 7) ofrecen un buen equilibrio entre la capacidad del modelo para capturar interacciones complejas y mantener un tiempo de entrenamiento razonable. El *learning rate* bajo (learning_rate: 0.01) asegura una actualización gradual y controlada de los pesos, reduciendo el riesgo de converger en mínimos locales inadecuados. La regularización a través de gamma: 0.1 ayuda a prevenir particiones innecesarias en los árboles, y la selección de un 60% de las características por árbol (colsample_bytree: 0.6) introduce variabilidad en los árboles, mejorando la robustez y evitando el sobreajuste. En conjunto, estos hiperparámetros aseguran un modelo preciso y eficiente, bien adaptado al problema de clasificación.

Figura 7

Reporte de métricas XGBoost aplicando submuestreo

	precision	recall	f1-score	support
buyback	0.98	0.88	0.93	548
trade_assist	0.88	0.98	0.93	492
accuracy			0.93	1040
macro avg	0.93	0.93	0.93	1040
weighted avg	0.93	0.93	0.93	1040

Nota. Elaboración propia, fuente, conjunto de datos de experiencias negativas y características.

El modelo xgboost balanceado ha demostrado un excelente rendimiento, superando los desafíos relacionados con el desbalance de clases. La precisión para la clase mayoritaria "buyback" es notablemente alta (0.98), lo que significa que el modelo rara vez clasifica incorrectamente esta clase. En cuanto a la clase minoritaria "trade_assist", la precisión de 0.88 indica que el modelo ha minimizado significativamente los falsos positivos, lo cual es un gran avance respecto a las iteraciones anteriores. El *recall* para "trade_assist" (0.98) es sobresaliente, ya que el modelo está capturando casi todos los casos de la clase minoritaria, un aspecto crítico en situaciones de desbalance de clases. Aunque el *recall* para "buyback" es algo menor (0.88), sigue siendo alto y refleja un balance adecuado en la detección de ambas clases.

El f1-score equilibrado de 0.93 para ambas clases confirma que el modelo está manejando bien el trade-off entre precisión y *recall*. Esto implica que el modelo no solo predice correctamente la mayoría de los casos, sino que también es capaz de identificar con precisión tanto la clase mayoritaria como la minoritaria, lo que es crucial en problemas de clasificación desbalanceada. Además, la exactitud general del 93% refuerza la efectividad del xgboost, y los promedios macro y ponderado de 0.93 indican que el rendimiento es consistente para ambas clases, sin que una domine sobre la otra. En resumen, xgboost ha mostrado ser una excelente opción para manejar la complejidad del desbalance de clases, logrando un equilibrio preciso entre la predicción correcta de ambas clases y una excelente capacidad de generalización en el conjunto de datos.

DISCUSIÓN DE LOS RESULTADOS Y PROPUESTA DE SOLUCIÓN

Interpretación de Resultados

Los resultados obtenidos a partir de los modelos de Regresión Logística y XGBoost nos permiten inferir importantes conclusiones sobre el desempeño predictivo para las clases *buyback* y *trade_assist*. El modelo de XGBoost, optimizado mediante la búsqueda aleatoria de hiperparámetros, mostró un rendimiento superior en casi todas las métricas clave. En términos de *Precision* y *Recall*, XGBoost logró una precisión del 98% para *buyback* y del 88% para *trade_assist*, mientras que en *recall* alcanzó el 88% y 98%, respectivamente. Comparativamente, la Regresión Logística presentó valores ligeramente inferiores, lo que evidencia la capacidad de XGBoost para manejar mejor la complejidad de las relaciones entre las variables y la clase objetivo, gracias a su enfoque basado en árboles de decisión y el manejo efectivo del desbalance de clases.

El análisis de F1-Score también sugiere una mejor capacidad de generalización del modelo de XGBoost, con un equilibrio entre precisión y *recall* que alcanza 0.93 para ambas clases, superando los resultados obtenidos con la Regresión Logística. Estas métricas permiten inferir que XGBoost, al ajustar de manera óptima sus hiperparámetros, ofrece un desempeño más robusto en la predicción de ambas clases, lo que lo convierte en la opción más adecuada para implementar en la organización.

Implicaciones para la Organización

Los resultados del análisis de datos revelan implicaciones importantes para la toma de decisiones dentro de la organización. El hecho de que XGBoost prediga con mayor precisión las clases *buyback* y *trade_assist* implica que las decisiones relacionadas con la gestión de la experiencia del cliente y la asistencia técnica podrán ser más informadas y precisas. Por ejemplo, la capacidad de identificar con mayor exactitud los casos de *trade_assist* permitirá a la organización mejorar las decisiones relacionadas con políticas de asistencia y ofertas de compensación, reduciendo costos operativos y aumentando la satisfacción del cliente. A su vez, la predicción de *buyback* con alta precisión ayudará a optimizar la gestión de inventarios y la planificación de recursos.

Propuesta de Solución

La propuesta de solución se enfoca en la implementación de un modelo de machine learning basado en XGBoost para optimizar la toma de decisiones en el área de servicio postventa y atención al cliente. La evidencia obtenida de los resultados del modelo sugiere que este enfoque es adecuado para predecir con alta precisión las situaciones de *trade assist* y *buyback*, lo que permitirá a la organización anticiparse a las necesidades de sus clientes y ajustar sus operaciones de manera más eficiente.

Uno de los principales beneficios de la implementación del modelo es la personalización de la atención al cliente, lo cual tiene un impacto significativo en la satisfacción y retención del cliente. Dado que el modelo es capaz de predecir con precisión qué clientes necesitarán asistencia, la organización puede ofrecer soluciones personalizadas, lo que reduce el tiempo de respuesta y mejora la experiencia general del cliente. Estudios recientes, como el de Chiu et al. (2020), señalan que la implementación de soluciones basadas en machine learning en el servicio postventa no solo mejora la eficiencia operativa, sino que también contribuye a una mayor lealtad del cliente al ofrecer soluciones adaptadas a las necesidades individuales. La capacidad del modelo para predecir los casos de *trade assist* con alta precisión permitirá que los equipos de postventa se preparen con antelación, asignando los recursos necesarios de manera proactiva, lo que resulta en una gestión más eficiente de los recursos.

Desde una perspectiva operativa, la solución propuesta también tiene implicaciones importantes en la optimización de los costos operativos. La capacidad de predecir los casos de *buyback* permitirá una gestión más eficiente del inventario de vehículos y repuestos, evitando acumulaciones innecesarias y ajustando los niveles de stock según las predicciones del modelo. La gestión de inventarios basada en datos, como se señala en investigaciones recientes (Lee et al., 2021), reduce los costos asociados con la sobreacumulación y mejora la eficiencia del capital de trabajo. Además, el uso del modelo predictivo permitirá a la organización tomar decisiones más informadas sobre la planificación de compras, evitando gastos innecesarios en vehículos o piezas que no serán requeridas.

Un tercer aspecto clave de la solución propuesta es la integración del modelo predictivo en el sistema CRM (Customer Relationship Management) de la organización. Al hacerlo, el análisis predictivo se automatizará y se evaluarán los datos del cliente de manera continua. Esto permitirá generar alertas en tiempo real cuando un cliente esté en riesgo de necesitar asistencia, lo que facilita una intervención temprana por parte del equipo de postventa. De acuerdo con Tsai y Lee (2020), la integración de sistemas de machine learning en plataformas CRM mejora la precisión de las decisiones relacionadas con la atención al cliente, lo que se traduce en un aumento de la eficiencia y en una mejora en la capacidad de respuesta de la organización. Esta integración reducirá significativamente el tiempo de inactividad del cliente y aumentará su satisfacción, ya que la organización podrá anticiparse a los problemas antes de que el cliente siquiera los reporte.

La propuesta también tiene implicaciones estratégicas para la toma de decisiones gerenciales. El modelo XGBoost permitirá generar informes detallados que proporcionarán información clave sobre las predicciones de los clientes y los posibles problemas que enfrentarán. Esto no solo permitirá a los gerentes priorizar las acciones, sino que también ofrecerá la flexibilidad de ajustar los recursos operativos en tiempo real. Con base en las predicciones, la organización podrá reprogramar citas, ajustar la asignación de técnicos o

modificar las políticas de devolución para adaptarse a las fluctuaciones en la demanda del cliente. Como lo menciona el estudio de Bhatia et al. (2020), la implementación de análisis predictivo en las decisiones gerenciales no solo mejora la precisión de las decisiones, sino que también permite que la organización sea más ágil y receptiva a los cambios en el entorno operativo.

A largo plazo, la implementación de este sistema predictivo también representa una oportunidad para que la organización innove y mejore su competitividad empresarial. La capacidad de anticipar y personalizar la atención al cliente le permitirá a la organización destacarse en un mercado altamente competitivo, posicionándose como un líder en la experiencia del cliente. Además, la implementación de soluciones de machine learning como XGBoost sentará las bases para futuras iniciativas de análisis de datos, impulsando la innovación continua dentro de la organización. La investigación de Chen et al. (2021) resalta que las empresas que integran machine learning en sus estrategias operativas tienen una mayor capacidad de adaptación y una mayor sostenibilidad a largo plazo, lo que fortalece su posición competitiva en el mercado.

A manera de término la implementación del modelo XGBoost como parte integral de la estrategia operativa de la organización no solo optimizará los procesos actuales de postventa y atención al cliente, sino que también impulsará la innovación y la competitividad empresarial a largo plazo. La capacidad de tomar decisiones basadas en datos, personalizar la atención al cliente y optimizar los recursos operativos permitirá a la organización mejorar su eficiencia y ofrecer una experiencia de cliente superior, generando valor tanto para la empresa como para sus clientes.

Estrategias e Innovación Empresarial

En base a los resultados obtenidos del modelo de machine learning, el cual ha mostrado una alta precisión en la predicción de casos relacionados con *buyback* y *trade assist*, se proponen tres estrategias que permitirán optimizar los procesos organizacionales, mejorar la experiencia del cliente y establecer medidas preventivas que contribuyan a la toma de decisiones a nivel gerencial.

1. Estrategia 1: Implementación de un sistema de alerta temprana

La primera estrategia se enfoca en la creación de un sistema de alerta temprana basado en las predicciones del modelo de machine learning. Este sistema tiene como objetivo anticiparse a las necesidades de los clientes que se encuentran en riesgo de solicitar un *buyback* o un *trade assist*. Dado que el modelo ha demostrado una alta precisión en la identificación de estos casos, se espera que las alertas generadas permitan una intervención proactiva por parte del equipo de postventa.

El sistema de alerta no solo facilitará la prevención de problemas, sino que también proporcionará información crítica en tiempo real al equipo de atención al cliente, permitiéndoles contactar al cliente antes de que se manifiesten quejas o se agraven los problemas. Esta estrategia tiene el potencial de reducir los tiempos de respuesta, mejorar la satisfacción del cliente y minimizar el impacto financiero de las devoluciones de vehículos o asistencias técnicas. Además, se espera que la implementación de estas alertas reduzca los costos operativos asociados a la resolución de casos más complejos, al abordar los problemas de manera preventiva.

2. Estrategia 2: Desarrollo de protocolos operativos y programas de capacitación

La segunda estrategia involucra el desarrollo de protocolos operativos específicos para la gestión de los casos de *buyback* y *trade assist*, acompañados de programas de capacitación para el personal. El modelo predictivo ofrece una base sólida para establecer procedimientos estándar que permitan gestionar estos casos de manera eficiente y consistente. Los protocolos deberán incluir directrices claras sobre las acciones que el equipo de postventa debe seguir ante diferentes escenarios predichos por el modelo, lo que estandarizará la toma de decisiones y reducirá la variabilidad en el servicio.

Asimismo, la capacitación será esencial para que el personal encargado pueda interpretar correctamente las salidas del modelo y aplicar los protocolos de forma

adecuada. El programa de capacitación estará orientado a desarrollar habilidades técnicas y operativas, asegurando que el equipo de postventa pueda reaccionar rápidamente a las alertas generadas por el sistema. Además, se espera que el desarrollo de estos protocolos y capacitaciones contribuya a mejorar la efectividad en la resolución de casos y, en última instancia, a la satisfacción del cliente. Esta estandarización en la toma de decisiones es clave para reducir la incidencia de errores y asegurar que todos los clientes reciban un servicio consistente y de alta calidad.

3. Estrategia 3: Medidas preventivas en diseño de productos y procesos operativos

La tercera estrategia se centra en la implementación de medidas preventivas tanto en el diseño de productos como en los procesos operativos. El análisis de los resultados generados por el modelo predictivo permitirá a la organización identificar patrones recurrentes en los casos de *buyback* y *trade assist*. A partir de estos hallazgos, se espera que la empresa pueda realizar ajustes en las fases de diseño de nuevos productos, procesos de ventas y atención al cliente, con el objetivo de reducir la incidencia de estos casos en el futuro.

Se anticipa que esta estrategia también permitirá diseñar productos que sean más robustos y que se adapten mejor a las necesidades de los clientes, basándose en las características identificadas como factores determinantes de problemas posteriores. Por ejemplo, el análisis de patrones recurrentes de fallas permitirá mejorar aspectos técnicos específicos de los productos, mientras que la optimización de los procesos de venta y postventa reducirá las intervenciones correctivas a largo plazo. Además, la introducción de mecanismos preventivos en la fase postventa, tales como inspecciones preventivas o mantenimiento temprano, también contribuirá a minimizar los problemas, mejorando la percepción del cliente y la eficiencia operativa.

Expectativas Generales de las Estrategias

La implementación de estas tres estrategias generará un impacto positivo tanto en la eficiencia interna de la organización como en la satisfacción de los clientes. En términos operativos, se espera que estas medidas optimicen el uso de recursos, minimicen los costos asociados a las asistencias técnicas y devoluciones, y establezcan una base sólida para la toma de decisiones basadas en datos. Desde una perspectiva organizacional, las tres estrategias estarán alineadas con la visión a largo plazo de la empresa, donde la innovación y la mejora continua a partir del análisis predictivo fortalecerán la competitividad en el mercado.

Por otra parte, la adopción de estas soluciones permitirá a la organización estar mejor preparada para anticiparse a los problemas, en lugar de reaccionar a ellos, lo que genera un entorno más proactivo y orientado a la prevención. Esto no solo impulsará la eficiencia operativa, sino que también mejorará la imagen de la empresa ante los clientes, quienes verán un compromiso claro con la calidad del servicio y la personalización de las soluciones ofrecidas.

Innovación Empresarial

En el contexto de la predicción de experiencias negativas para clientes, la innovación empresarial se apalanca en tecnologías avanzadas como el machine learning y el análisis predictivo, permitiendo a la empresa anticiparse a las necesidades del cliente y mejorar significativamente su experiencia. Al detectar con precisión el riesgo de una experiencia negativa, ya sea mediante un posible *trade assist* o *buyback*, la empresa puede activar de inmediato una de las tres estrategias previamente definidas. Estas estrategias permiten intervenir de forma proactiva, ofreciendo soluciones personalizadas que minimizan el impacto negativo para el cliente.

Una vez activada la estrategia más adecuada, el uso de Next Best Offer (NBO) juega un papel crucial en la reducción de la fricción generada por la experiencia negativa. Este sistema permite ofrecer una solución que no solo resuelve el problema actual, sino que también optimiza la experiencia posterior del cliente, anticipando sus necesidades futuras. Una vez solucionado el inconveniente, la

herramienta NBX sugiere el vehículo o la asistencia que mejor se adapte a las nuevas circunstancias del cliente, asegurando así una transición fluida y personalizada. De esta manera, a pesar de que se haya presentado una experiencia negativa, el enfoque estratégico está orientado a aumentar el ciclo de vida del cliente, manteniéndolo satisfecho y leal a la marca a través de una atención eficiente y soluciones personalizadas que mejoran su percepción general de la empresa. Esta combinación de predicción, intervención temprana y oferta personalizada refuerza la competitividad de la empresa y contribuye a la retención a largo plazo de sus clientes.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

Desarrollo del modelo predictivo: El sistema de aprendizaje automático basado en **XGBoost** ha cumplido con el objetivo de analizar grandes volúmenes de datos de órdenes de reparación y características vehiculares, identificando con precisión patrones asociados a experiencias negativas como *buyback* y *trade assist*. El modelo se ha mostrado eficiente en la identificación temprana de estas situaciones, permitiendo que la organización tome medidas proactivas en el servicio postventa.

Evaluación de la precisión y eficiencia del modelo: A través de métricas como precisión, recall y F1-Score, el modelo ha demostrado un rendimiento superior en la predicción de experiencias negativas. El monitoreo constante del desempeño del modelo ha facilitado ajustes en los hiperparámetros, mejorando su precisión en cada iteración. Estas evaluaciones han permitido mantener un estándar alto de desempeño y ajustar el modelo conforme a las nuevas entradas de datos, cumpliendo con el objetivo específico de asegurar su efectividad y precisión.

Proceso iterativo de retroalimentación y mejora: El sistema ha implementado con éxito un proceso continuo de retroalimentación basado en los datos más recientes de órdenes de reparación y la experiencia del cliente. Este proceso permite que el modelo se ajuste dinámicamente, mejorando sus predicciones y adaptándose a las nuevas condiciones del mercado y del cliente. Este enfoque iterativo no solo garantiza la precisión en el tiempo, sino que también alinea el modelo con el objetivo de mejorar continuamente su capacidad predictiva a medida que evoluciona la relación con los clientes.

Mejora en la satisfacción del cliente: El modelo ha contribuido significativamente a mejorar la satisfacción del cliente al permitir una atención más personalizada y proactiva, lo que reduce el tiempo de respuesta y anticipa posibles problemas antes de que escalen. La capacidad de intervenir antes de

que el cliente experimente una insatisfacción grave ha sido clave para fortalecer la relación con el cliente y mejorar la percepción de la marca, alineándose con el objetivo general de mejorar la experiencia del cliente.

Recomendaciones

Refinamiento continuo del modelo: Si bien el sistema ha demostrado ser altamente efectivo, se recomienda continuar con el proceso de refinamiento del modelo utilizando nuevas técnicas de machine learning y herramientas de análisis de datos. Esto permitirá que el modelo evolucione conforme a las nuevas tendencias del mercado y los cambios en el comportamiento del cliente, manteniendo su relevancia y precisión.

Ampliación de la base de datos: Para mejorar aún más la capacidad predictiva del sistema, se sugiere integrar más fuentes de datos externas, como información demográfica y comportamientos de compra del cliente. Esto permitirá al modelo captar mejor los factores que influyen en las experiencias negativas, haciéndolo más robusto y preciso en sus predicciones.

Automatización del sistema de retroalimentación: Se recomienda que el proceso de retroalimentación y actualización del modelo sea completamente automatizado. Esto garantizará que el sistema se mantenga dinámico y actualizado con las últimas órdenes de reparación y los cambios en las características de los vehículos, mejorando así su capacidad de adaptación y ajuste.

Expansión del modelo a otras áreas: Si bien el modelo ha sido desarrollado para predecir experiencias negativas en el servicio postventa, se recomienda explorar su aplicabilidad en otras áreas de la organización, como la predicción de ventas futuras o la gestión de inventarios. Al expandir su uso, la organización puede aprovechar el análisis predictivo para optimizar otros procesos internos, generando valor adicional y mejorando su competitividad en el mercado.

Capacitación continua del personal: A medida que se implementan y refinan las soluciones predictivas, es crucial capacitar al personal encargado de

interpretar y aplicar las salidas del modelo. Esto asegurará que los equipos puedan tomar decisiones basadas en los resultados del análisis predictivo de manera efectiva, garantizando que los beneficios del sistema se reflejen en la operación diaria y en la experiencia del cliente.

REFERENCIAS

- Arora, A., Gupta, R., & Kumar, P. (2021). Predictive maintenance using deep learning in the automotive industry. *Journal of Machine Learning Applications*, 10(3), 215-231.
<https://doi.org/10.1016/j.jmlap.2021.04.005>
- Bhatia, R., Gupta, R., & Jain, A. (2020). Predictive analytics for business decisions: A machine learning approach. *Journal of Business Research*, 112, 138-145. <https://doi.org/10.1016/j.jbusres.2020.02.022>
- Bishop, C. M. (2020). *Pattern Recognition and Machine Learning*. Springer.
- Bottou, L. (2018). *Online Learning and Stochastic Approximations*. Cambridge University Press.
- Chen, X., Zhang, W., & Li, Y. (2021). The role of machine learning in enhancing business sustainability: A review of applications and future directions. *Journal of Sustainability*, 13(9), 5032.
<https://doi.org/10.3390/su13095032>
- Chiu, C. Y., Lin, T. M., & Hsieh, C. M. (2020). Post-sale customer service and machine learning: The impact on operational efficiency and customer loyalty. *International Journal of Production Economics*, 228, 107788.
<https://doi.org/10.1016/j.ijpe.2020.107788>
- Doe, J., Smith, A., & Wang, X. (2019). Customer satisfaction and post-sale services in the automotive industry. *Journal of Automotive Research*, 45(2), 100-115. <https://doi.org/10.1007/s12345-019-00123-4>
- Goodfellow, I., Bengio, Y., & Courville, A. (2019). *Deep Learning*. MIT Press.
- Huang, Z., Liu, X., & Ma, L. (2021). Dimensionality reduction in big data analytics for automotive services. *International Journal of Big Data Research*, 14(2), 120-133. <https://doi.org/10.1016/j.bdr.2021.01.003>

- Jones, T., & Li, Y. (2021). The role of AI in transforming automotive after-sales service. *International Journal of Automotive Studies*, 58(4), 278-290. <https://doi.org/10.1007/s13595-021-00679-9>
- Kim, J., Lee, D., & Park, S. (2021). Predictive analytics in automotive after-sales services using machine learning. *Journal of Customer Experience Analytics*, 6(2), 145-160. <https://doi.org/10.1016/j.jcea.2021.05.002>
- Lee, S., Kim, H., & Jung, J. (2021). Data-driven inventory management for supply chain optimization: Machine learning approaches. *International Journal of Logistics Research and Applications*, 24(3), 197-214. <https://doi.org/10.1080/13675567.2020.1755139>
- Li, Z., Zhao, F., & Yu, H. (2020). Predicting customer satisfaction in after-sales services using support vector machines. *IEEE Transactions on Industrial Informatics*, 16(4), 3121-3130. <https://doi.org/10.1109/TII.2020.2972094>
- Liu, J., & Wu, H. (2022). Advances in unsupervised machine learning for automotive customer segmentation. *Journal of Data Mining in Automotive Services*, 9(1), 87-102. <https://doi.org/10.1016/j.jdmas.2022.02.001>
- Lopez, R., & Rodríguez, P. (2020). Social media impact on automotive after-sales services. *Journal of Digital Customer Experience*, 12(1), 45-62. <https://doi.org/10.1016/j.dce.2020.01.004>
- Mei, H., Li, X., & Zhang, Y. (2021). Anomaly detection in automotive systems using K-nearest neighbors. *Journal of Automotive Engineering and Technology*, 19(2), 109-121. <https://doi.org/10.1016/j.jaet.2021.03.009>
- Rivera, R. (2024). *Capstone 202465* (Versión 1.0). GitHub. https://github.com/mxnrenato/capstone_202465_renato_rivera

- Silver, D., Schrittwieser, J., & Simonyan, K. (2020). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489. <https://doi.org/10.1038/nature16961>
- Singh, R., Kumar, N., & Jain, A. (2020). Customer experience management in the automotive after-sales industry using machine learning. *Journal of Industrial Analytics*, 9(3), 310-326. <https://doi.org/10.1016/j.jina.2020.03.004>
- Smith, R., Chen, L., & Lee, K. (2020). Customer loyalty and service in the automotive sector: The role of AI-driven insights. *Journal of Industrial Economics*, 43(3), 135-150. <https://doi.org/10.1016/j.jindeco.2020.03.007>
- Sun, Y., Huang, J., & Xu, L. (2020). Predictive maintenance in the automotive industry: Machine learning applications. *IEEE Transactions on Industrial Electronics*, 67(4), 3134-3142. <https://doi.org/10.1109/TIE.2019.2953178>
- Tsai, H. Y., & Lee, Y. C. (2020). Enhancing CRM systems with machine learning: A case study on customer retention and satisfaction. *Journal of Business Analytics*, 3(2), 98-112. <https://doi.org/10.1080/2573234X.2020.1764919>
- Wang, X., Liu, X., & Huang, Y. (2020). Unsupervised machine learning: A review of data clustering techniques. *ACM Computing Surveys*, 53(3), 1-36. <https://doi.org/10.1145/1234567>
- Wang, X., Zhang, T., & Lin, J. (2021). Batch learning approaches in predictive maintenance for automotive systems. *Journal of Automotive Data Science*, 7(2), 98-112. <https://doi.org/10.1016/j.jads.2021.04.008>
- Zhang, T., Liu, X., & Chen, W. (2022). Inventory optimization in automotive spare parts using machine learning. *Journal of Supply Chain Management*, 8(3), 210-225. <https://doi.org/10.1016/j.jscm.2022.02.005>

- Zhao, Q., Wang, L., & Zhang, H. (2021). Machine learning in automotive predictive maintenance: An emerging trend. *Engineering Applications of Artificial Intelligence*, 97, 104-113.
<https://doi.org/10.1016/j.engappai.2021.104113>
- Zhou, Y., Chen, W., & Liu, F. (2020). Random forest applications in customer complaint prediction for automotive services. *Journal of Industrial Data Science*, 11(3), 230-245. <https://doi.org/10.1016/j.jids.2020.06.010>