



# MODELO DE PREDICCIÓN DE PRECIOS DE VEHÍCULOS

Análisis de Datos Machine Learning Valuación Automotriz



# Contexto del Mercado y Problemática

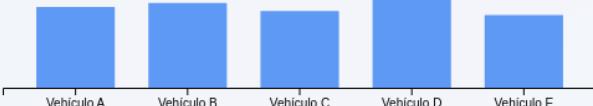
## Contexto del Mercado

⚠️ Factores responsables de la variabilidad: historial de vehículo, antigüedad, kilometraje, condiciones de uso, características técnicas, ubicación y tipo de vendedor

👥 Esto genera falta de transparencia, pérdida de valor, negociaciones extensas y riesgos financieros tanto para consumidores como para empresas del sector

⌚ Tiempos prolongados de negociación debido a la falta de referencias actualizadas

Variabilidad de Precios en Vehículos Similares



## Desafíos Actuales

### ⚠️ Inconsistencias

Precios variables para vehículos similares

### ☒ Pérdidas Económicas

Precios incorrectos derivan en pérdidas

### ☒ Tiempo Perdido

Negociaciones prolongadas

### 👤 Baja Satisfacción

Clientes insatisfechos con precios injustos

# Definición del Problema de Negocio



## Problemas Identificados

### Valoración Objetiva

- Valoración instantánea y objetiva
- Mayor confianza comercial

### Reducción de fricciones

- Reducción de fricciones en negociación
- Estandarización interna de precios

### Integración

- Plataformas de e-commerce automotriz
- Dealers, aseguradoras, sistemas internos de tasación

### Soporte para decisiones

- Soporte para decisiones de compra, venta y financiamiento
- Apps móviles para consumidores

### Impacto en el Negocio

Un modelo estandarizado y automatizado provee confianza por partes interesadas y soporte al negocio. La automatización estandariza los criterios internos y permite que los precios sean objeto de predicción objetiva, replicable y basada en datos reales.

# Solución Propuesta

## Modelo de Machine Learning

Desarrollo de un modelo de Machine Learning para predecir el precio óptimo de un vehículo usado, permitiendo una valoración basada en datos históricos y patrones del mercado.

### Rápida

Valoración instantánea al recibir un vehículo

### Confiable

Resultados consistentes basados en patrones reales

### Basada en Datos

Ánalysis de miles de vehículos históricos

### Escalable

Adaptación a distintos escenarios de negocio

## Flujo del Modelo



Datos de Entrada → Procesamiento → Modelo ML → Predicción → Precio Resultante

# Relevancia y Objetivos del Negocio

Este modelo ofrece soluciones estructuradas para transformar información histórica en predicciones teóricas, optimizando la sustentabilidad empresarial y soportando decisiones estratégicas sustentables en entornos empresariales complejos.



## Eficiencia Operativa

Reduce riesgo de sobreprecio al identificar precios inflados mediante análisis estadístico de variación de precios y rampas.



## Reducción de Pérdidas

Proporciona transparencia para compradores y vendedores basados en datos históricos y cronológicos, eliminando subjetivismo.



## Satisfacción del Cliente

Optimiza la valoración para escenarios de inventario distribuido permitiendo tasación uniforme, estimación de tiempos de demanda y detección de tendencias.



## Estandarización de Criterios

Mejora la experiencia de usuario al ofrecer tasación inmediata y accesible a todos los niveles y transmitiendo la información clave para decisiones autónomas.



Genera ventajas competitivas constantes con capacidades escalables de soporte a decisiones estratégicas y ofrecimiento en timeframes mensuales y trimestrales.

# Conjunto de Datos y Variables Clave

## Dataset Estructurado

Para construir este modelo, utilizamos un dataset robusto compuesto por:

Más de 3,500 registros reales de vehículos

ColumnTransformer con StandardScaler y OneHotEncoder

Feature engineering para optimización



## Variables Clave

### Marca y Modelo

Distintas marcas y modelos de vehículos

### Año

Fecha de fabricación y antigüedad del vehículo

### Precio

Valor del vehículo en condiciones buen estado

### Kilometraje

Recorrido total del vehículo

### Tipo de Combustible

Gasolina, diésel, eléctrico y híbridos

### Transmisión

Manual, automática, semiautomática

### Tipo de Vendedor

Privado, concesionario, revendedor, etc.

### Número de Dueños

Historial de propietarios del vehículo

### Edad del Vehículo

Años reportados desde la fabricación

### Km/Año (feature diseñado)

Features creados para modelado predictivo

# Beneficios Corporativos del Modelo



## Estandarización del Pricing

Elimina variabilidad subjetiva entre valuadores, creando precios consistentes para vehículos similares.

- Precios justos y consistentes



## Decisiones Más Informadas

Precios basados en patrones reales del mercado, no en criterios subjetivos o experiencia individual.

- Fundamentado en datos reales



## Reducción de Errores Humanos

Minimiza sobrevaluaciones o subvaluaciones típicas de la evaluación manual.

- Mayor precisión en valuaciones



## Escalabilidad Automatizada

Permite procesar múltiples unidades de forma automática, reduciendo tiempo y recursos.

- Procesamiento rápido y eficiente

💡 Estos beneficios traducen directamente en mayor eficiencia operativa, satisfacción del cliente y competitividad en el mercado.

# Casos de Uso y Aplicaciones



## Concesionarios

Cotización instantánea al recibir un vehículo en parte de pago, permitiendo una evaluación rápida y objetiva para establecer el precio de compra.

- ✓ Integración con procesos de compra/venta

## E-commerce Automotriz

Mostrar precios sugeridos o rangos recomendados para vehículos, mejorando la experiencia del cliente y la transparencia en los procesos de compra.

- ✓ Mejora la confianza del comprador

## Compradores Particulares

Obtener una referencia confiable antes de negociar, permitiendo a los compradores potenciales tomar decisiones informadas y justas.

- ✓ Soporte para la negociación

## Equipos de Pricing

Analizar tendencias y definir estrategias de mercado, proporcionando datos objetivos para la toma de decisiones estratégicas.

- ✓ Base para estrategias de mercado

El modelo sirve tanto para herramientas internas como para aplicaciones orientadas al cliente final

# Metodología de Desarrollo



## Evaluación de Algoritmos

RandomForestRegressor  
Mejor R<sup>2</sup> y RMSE más estable

GradientBoostingRegressor  
Alto desempeño pero menos estable

XGBoostRegressor  
Menor MAE pero mayor complejidad

LinearRegression  
Interpretable pero menos flexible

SVR  
Mayor sensibilidad a outliers

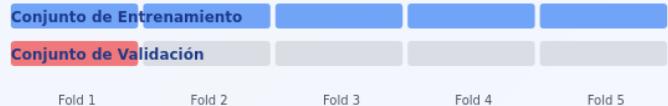
## Pipeline y Validación

### Pipeline Automatizado

- StandarScaler para variables numéricas
- OneHotEncoder para variables categóricas
- Particionamiento en Train/Validate/Test

### Validación por GridSearchCV

- 5-fold cross-validation
- Evaluá cientos de combinaciones de hiperparámetros
- Selección optimum para cada algoritmo



# Feature Engineering y Optimización



## Edad del Vehículo

Cálculo directo: Año actual menos año de fabricación

Ejemplo:

$$2023 - 2018 = 5 \text{ años}$$



## Codificación One-Hot

Manejo de variables categóricas con alta cardinalidad

Marca (Ford, Toyota...)    Combustible (Gasolina, Diesel...)

Transmisión (Manual, Automática...)



## Kilómetros Promedio

Análisis de kilometraje por periodo de uso

Fórmula: **Km totales / Años de uso**



## Transformaciones Estadísticas

Estabilización del comportamiento del target variable

Técnicas: **Log, Box-Cox, Standardization**



## Escalado de Variables

Normalización de variables numéricas para mejor desempeño

Métodos:

**Min-Max, Z-score, Robust**

# Desafíos Técnicos Resueltos



## Alta Cardinalidad

Categorías con muchos niveles, especialmente modelos de auto

### Solución Implementada:

- Grouping de modelos similares
- Feature hashing para codificación
- Regularización en modelos



## Distribución Sesgada

Precios con distribución asimétrica que afecta modelos

### Solución Implementada:

- Transformaciones logarítmicas
- Box-Cox para estabilizar varianza
- Winsorización de extremos



## Datos Faltantes

Valores ausentes que requerían tratamiento

### Solución Implementada:

- Imputación automatizada
- Modelos robustes a valores faltantes
- Generación de features de m Completo



## Pipeline Reproducible

Necesidad de un sistema automatizado y confiable

### Solución Implementada:

- ColumnTransformer para manejo mixed
- Parámetros versionados
- Automatización de despliegue

# Métricas de Desempeño del Modelo

Las métricas permiten evaluar la calidad de las predicciones y la generalización del modelo.



**MAE**

Error Absoluto Medio



**RMSE**

Error Cuadrático Medio



**R<sup>2</sup>**

Coeficiente de Determinación



## 💡 Conclusión

Estas métricas evidencian que el modelo generaliza bien y ofrece predicciones confiables.

# Valor Estratégico y Ventajas Competitivas

## 👑 Activo Estratégico

El modelo de predicción desarrollado representa un activo estratégico para la organización, capaz de transformar la información histórica en predicciones accionables.

- Estandarización de la valuación de vehículos
- Reducción de incertidumbre en procesos comerciales
- Aumento de la eficiencia operativa
- Base para digitalización del negocio

## Cadena de Valor Estratégico



## 🏆 Ventajas Competitivas

### 🎯 Precisión Mejorada

Estimaciones más precisas basadas en datos reales del mercado

### ⌚ Eficiencia Operativa

Procesos automatizados y toma de decisiones más rápidas

### 💡 Innovación Continua

Base para futuras aplicaciones de IA y mejoras tecnológicas

## 👉 Fundamento para el Futuro



Sistemas de  
recomendación



Análisis de demanda



Automatización completa  
del pricing

# Futuras Oportunidades de IA



## Modelo Actual

Predicción de precios de vehículos usados



### Sistemas de Recomendación

- ✓ Recomendación de vehículos según preferencias del cliente
- ✓ Similitudes con vehículos históricos valorados
- ✓ Personalización basada en comportamiento de compra



### Análisis de Demanda

- ✓ Identificación de tendencias de mercado por modelo
- ✓ Segmentación de clientes y preferencias
- ✓ Predicción de movimientos de inventario



### Automatización del Pricing

- ✓ Actualización automática de precios basada en mercado
- ✓ Procesos comerciales completamente automatizados
- ✓ Herramientas inteligentes para valuaciones en tiempo real

*"El modelo sentó las bases para una transformación digital completa de nuestros procesos"*

# Conclusiones y Próximos Pasos

## Impacto del Modelo

### Activo Estratégico

Herramienta clave para la digitalización del negocio y generación de ventajas competitivas

### Estandarización

Elimina variabilidad subjetiva entre valuadores y reduce incertidumbre en procesos comerciales

### Eficiencia Operativa

Mejora la toma de decisiones y reduce tiempos de respuesta en el proceso comercial

# Resultados Clave de los Modelos

## Evaluación de Modelos

Modelo	MAE	RMSE	R <sup>2</sup>
RandomForest	154,231	352,495	0.714
GradientBoosting	153,709	360,262	0.697
XGBoost	149,616	366,212	0.584
LinearRegression	179,358	384,487	0.541
SVR	216,278	532,267	0.120

## Interpretación de Resultados

### Precisión en Total de Variabilidad

RandomForest captura **aproximadamente el 71% de la variabilidad** en el mercado automotriz, mientras que los demás modelos varían entre el 58% y el 71%.

Esto indica una correlación sólida con el mercado real, dada la complejidad que presentan estos mercados y el gran número de factores no observables que impactan el precio final de los vehículos.

### Optimización de Métricas

#### MAE

XGBoost otorga el error promedio más bajo (149,616)

#### RMSE

RandomForest ofrece la mayor variabilidad controlada

La estabilidad relativamente constante del RandomForest contrasta con las fluctuaciones en otras métricas similares a XGBoost, lo que sugiere un modelo más constante ante diferentes escenarios.

### Propuesta Combinatoria

A pesar de que el RandomForest modela mejor el total de la variabilidad, otros modelos podrían ser más convenientes para escenarios específicos donde se deseé priorizar el menor impacto absoluto de los errores (XGBoost), reducir sesgo (SVR) o convertir en estándar de registro interno de tipo lineal (LinearRegression).

# Modelo Final Seleccionado: RandomForestRegressor

Tras una evaluación exhaustiva, finalmente seleccionamos **RandomForestRegressor** como modelo base para el desarrollo de la solución en producción por su equilibrio óptimo entre precisión y robustez.

Además, nuestro modelo integra un pipeline completo de preparación y transformación de datos para garantizar consistencia entre las etapas de entrenamiento y inferencia.

## Características Clave

- Equilibrio Óptimo**  
Mejora la predicción de precios con un buen balance entre modelos lineales (interpretables) y no lineales (precisos).
- Menor Sensibilidad**  
Menor sensibilidad al ruido y outliers en el set de datos comparado con modelos más sensibles como SVR.
- Interpretabilidad**  
Interpretarable relativa con técnicas de importancia de características que facilitan la explicación a diferentes stakeholders.
- Rendimiento Real**  
Excelente rendimiento en datos reales con patrones complejos y no lineales propios de esta variable dependiente.

## Características Técnicas del Pipeline

- Pipeline Completo**  
Integración de transformadores del ColumnTransformer, validación cruzada y normalización de variables financieras.
- Versionamiento de Modelos**  
Los modelos son versionados por fecha y registramos hiperparámetros esperados para cualquier contenedor Docker.
- Escalabilidad**  
Escalable tanto nivel de máquina (utilizando Dask) como nivel de set de datos (desde 100 registros hasta producción con millones).
- Uso Mínimo de Recursos**  
Dimensionado de float64 a float32 y memoria comprimida permiten modelos que requieren un mínimo de 4GB de memoria RAM.
- Backend Flexible**  
Se puede desplegar ejecutándose en motores DMV/DMD en backend y garantizaría un rendimiento superior con alto volúmenes de tráfico.