

Análisis de Datos y Modelado Predictivo: Devaluación de Autos Usados

Introducción

La toma de decisiones basada en datos requiere identificar y modelar las variables críticas que gobiernan un sistema. Este proyecto aplica estos principios al mercado de autos usados en Argentina, tratando la depreciación de vehículos como un proceso multivariable dependiente de la **antigüedad (t)** y el **desgaste operativo**, representado por los kilómetros de uso, (x).

El objetivo fue diseñar un sistema integral capaz de recolectar masivamente datos crudos, depurarlos mediante filtros estadísticos, modelar matemáticamente su comportamiento y diseñar un dashboard interactivo para visualizar los resultados. Se busca predecir el valor de mercado con alta precisión, caracterizar la devaluación futura del vehículo y así tomar una decisión de inversión.

Metodología de Análisis de Datos

Para caracterizar el sistema, se implementó un pipeline de procesamiento de datos masivos (*High-Throughput Data Processing*).

- 1. Adquisición:** Se automatizó la recolección de cientos de puntos de datos de múltiples fuentes web, estructurando información no homogénea.
 - 2. Filtrado de Señal:** Los datos crudos contienen “ruido”. Se aplicaron técnicas estadísticas de **Caldado de Sigma** ($Z\text{-Score} > 1.5$) para separar outliers (en general, publicaciones con valores artificialmente bajos, debido a condiciones excepcionalmente malas o precios más altos de lo que el mercado convalida) y asegurar que los datos representen fielmente las realidades del mercado.
-

Modelado Matemático

Se propuso describir la depreciación del activo mediante funciones que correlacionan el valor (P , el precio de mercado) con las variables de estado (t , x).

Aproximación Lineal

$$P(t, x) = at + bx + c$$

Asume una tasa de depreciación constante. A pesar de parecer reduccionista, resultó efectiva para modelar el comportamiento algunos vehículos con demanda relativamente inelástica como la **Ford Ecosport** ($R^2 = 0.918$) y el **Volkswagen T-Cross** ($R^2 = 0.784$).

Modelo de Decaimiento Exponencial

$$P(t, x) = e^{\alpha t + \beta x + \gamma}$$

Asume que el tiempo y el uso generan la pérdida de una fracción del valor del auto (y no una cantidad fija). Este modelo captura la rápida pérdida de valor inicial y la estabilización asintótica en el tiempo, característico de bienes de consumo masivo. Este modelo, por sí solo fue efectivo para una gama mucho más amplia de vehículos.

Optimización por Ensamble

Para maximizar la correlación (R^2), se diseñó un algoritmo que pondera las salidas de ambos modelos:

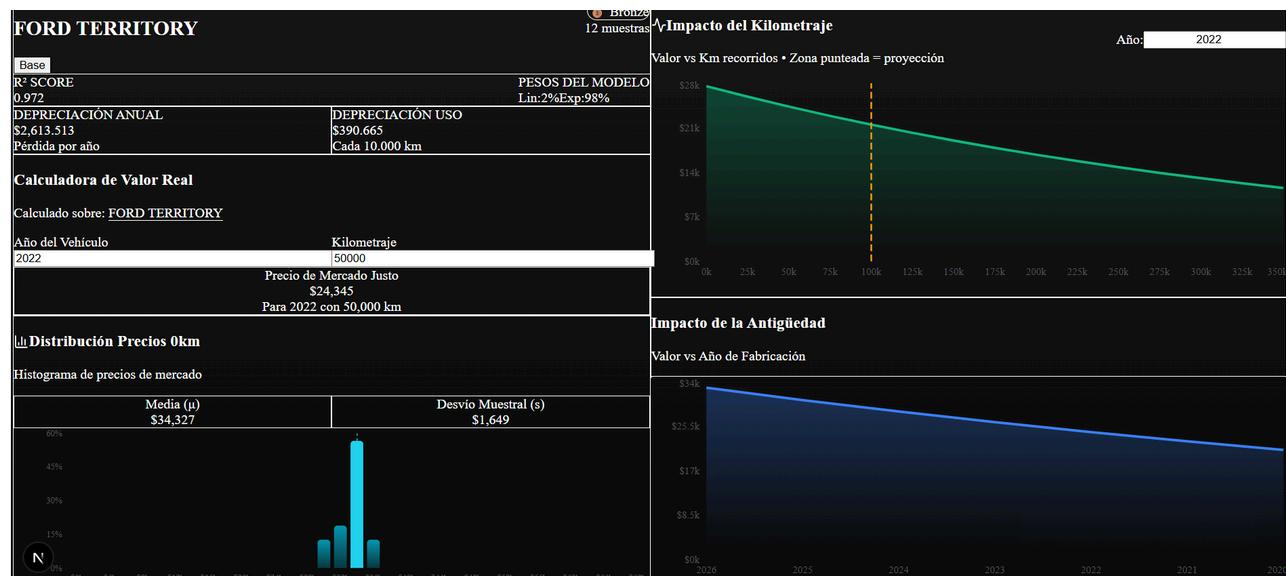
$$P_{ensamble}(t, x) = L \cdot P_{lineal}(t, x) + E \cdot P_{exponencial}(t, x)$$

Donde los coeficientes de peso L, E se calculan minimizando el error cuadrático medio (RMSE). Para algunos modelos que no presentaban comportamientos netamente lineales o exponenciales, combinar la información obtenida por ambos modelos logró un mejor ajuste. Tal es el caso para, por ejemplo, para la **Jeep Renegade** ($L = 0.43, E = 0.57, R^2 = 0.850$) o para la **Hyundai Tucson** ($L = 0.36, E = 0.64, R^2 = 0.861$).

Visualización y Dashboard de Control

Para la interfaz de usuario, se diseñó un **Dashboard Interactivo** que busca facilitar la toma de decisiones mediante indicadores claros de valoración. Permite:

- Visualizar las curvas de depreciación.
- Acceder a valores numéricos de pérdida de valor anual y por 10.000 km.
- Comparar métricas de dispersión y precisión del modelo.
- Comparar los valores de ejemplares usados con los de 0km.
- Calcular el valor de mercado predicho por el modelo para una cierta antigüedad y kilometraje.



Validación de Resultados

El modelo se sometió a un test de validación con un conjunto de control de 75 unidades reales de 17 modelos distintos.

- **Resultados:** Se alcanzó una **Precisión Global del 89.44%**, con picos del 94% en modelos de alta representatividad estadística.
- **Conclusión:** El sistema valida la capacidad de aplicar principios del modelado matemático para transformar grandes volúmenes de datos dispersos en información

de alto valor predictivo.

Reporte de modelos analizados:

Modelo	Proporción (Lin/Exp)	R ²	Accuracy	Desgaste Temporal (USD/año)	Desgaste Operativo (USD/10k KM)
Nissan Kicks	0% / 100%	0.487	94.09%	-\$562	-\$274
Toyota C. Cross	0% / 100%	0.522	93.14%	-\$1,048	-\$491
Honda HR-V	0% / 100%	0.783	92.38%	-\$1,077	-\$269
Ford EcoSport	100% / 0%	0.918	91.99%	-\$644	-\$330
Toyota SW4	0% / 100%	0.748	91.77%	-\$1,580	-\$333
Ford Territory	1.5% / 98.5%	0.972	91.64%	-\$2,614	-\$391
Jeep Renegade	43% / 57%	0.850	91.55%	-\$977	-\$318
Jeep Compass	0% / 100%	0.719	91.54%	-\$1,180	-\$447
VW Tiguan	0% / 100%	0.708	90.69%	-\$1,086	-\$498
Chevrolet Tracker	0% / 100%	0.843	89.93%	-\$624	-\$292
VW T-Cross	100% / 0%	0.784	87.49%	-\$806	-\$412

Modelo	Proporción (Lin/Exp)	R ²	Accuracy	Desgaste Temporal (USD/año)	Desgaste Operativo (USD/10k KM)
Hyundai Tucson	36% / 64%	0.861	84.05%	-\$884	-\$877
VW Nivus	4% / 96%	0.450	83.79%	-\$404	-\$464
Honda CR-V	1% / 99%	0.983	81.86%	-\$1,703	-\$287
Kia Sorento	0% / 100%	0.855	80.85%	-\$1,444	-\$304
Renault Duster	0% / 100%	0.987	79.07%	-\$815	-\$118