

Primera entrega del proyecto IA

Por:

Esteban Caro Pelaez
Julian David Olaya Restrepo
Isabela Pareja Gil

Materia:

Introducción a la inteligencia artificial

Profesor:

Raul Ramos Pollan



UNIVERSIDAD DE ANTIOQUIA

1 8 0 3

**UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERÍA INDUSTRIAL
MEDELLÍN 2023**

1. Descripción del problema predictivo

Una de las formas de adquisición de vehículos en los concesionarios es mediante la compra de automóviles en subasta. Muchos de estos vehículos tienen fallas mecánicas recurrentes que pueden estar relacionadas con diferentes factores; problemas de autenticación del vendedor, modificación del tacómetro para el kilometraje, fallas mecánicas, entre otros. La compra de dichos vehículos llegan a ser inversiones riesgosas, puesto que en muchas ocasiones, la cantidad de dinero necesario para solventar dichas fallas superan la ganancia esperada por el concesionario.

Teniendo en cuenta lo anterior, se despierta un especial interés por tener un modelo basado en datos que permita predecir si el auto que se compró tiene un alto riesgo de ser una mala compra teniendo en cuenta las diferentes características propias del vehículo, tales como la marca, el año de fabricación y el tipo de automóvil.

2. Descripción del dataset

El dataset a utilizar fue obtenido de la competición en Kaggle “Don’t Get Kicked!”, al cual se puede acceder mediante el enlace: https://www.kaggle.com/competitions/DontGetKicked/data?select=Carvana_Data_Dictionary.txt. El dataset cuenta con 121.692 registros de autos, de los cuales el 60% es destinado para el training y el 40% para testing. El dataset cuenta con el 40% de los datos categóricos y el 60% numéricos.

Descripción de las variables

RefId: Identificador del vehículo dentro de la base de datos

IsBadBuy: variable binaria que representa si el vehículo fue una mala compra (1) o no (0).

PurchDate: fecha en la que se compró el vehículo en la subasta

Auction: proveedor a quien se le compró el vehículo en la subasta

VehYear: año de fabricación del vehículo

VehicleAge: años transcurridos desde el año de fabricación del vehículo

Make: marca del fabricante del vehículo

Model: modelo del vehículo

Trim: versión del modelo con una configuración particular

SubModel: submodelo del vehículo

Color: color del vehículo

Transmission: tipo de transmisión del vehículo (automática o manual)

WheelTypeID: identificador del tipo de llanta (Covers 2 - Alloy 1)

WheelType: tipo de llanta (Covers, Alloy)

VehOdo: kilometraje del vehículo

Nationality: nacionalidad del fabricante del vehículo

Size: categoría del tamaño del vehículo

TopThreeAmericanName: identifica si el vehículo hace parte del top 3 de fabricantes estadounidenses.

WarrantyCost: El precio de la garantía con un términos de 36 meses y y 36 mil millas.

MMRAcquisitionAuctionAveragePrice: Precio de adquisición para este vehículo en estado medio en el momento de la compra

MMRAcquisitionAuctionCleanPrice: Precio de adquisición para este vehículo en un estado superior a la media en el momento de la compra

MMRAcquisitionRetailAveragePrice: Precio de adquisición de este vehículo en el mercado minorista en estado medio en el momento de la compra

MMRAcquisitionRetailCleanPrice: Precio de adquisición de este vehículo en el mercado minorista en un estado superior a la media en el momento de la compra.

MMRCurrentAuctionAveragePrice: Precio de adquisición de este vehículo en estado medio en el día actual

MMRCurrentAuctionCleanPrice: Precio de adquisición de este vehículo en el estado anterior en el día actual

MMRCurrentRetailAveragePrice: Precio de adquisición de este vehículo en el mercado minorista en estado medio en el día actual.

MMRCurrentRetailCleanPrice: Precio de adquisición para este vehículo en el mercado minorista en condiciones superiores a la media en el día actual.

PrimeUnit: Identifica si el vehículo tendría una demanda mayor que una compra estándar

AcquisitionType: Identifica cómo se adquirió el vehículo (subasta, intercambio, etc.).

AUCGUART: El nivel de garantía proporcionado por la subasta para el vehículo (Luz verde - Garantizado/interceder, Luz amarilla - precaución/problema, luz roja - vendido tal cual)

KickDate: Fecha de devolución del vehículo a la subasta

BYRNO: Número único asignado al comprador que adquirió el vehículo

VNZIP: Código postal donde se compró el automóvil

VNST: Estado en el cual el automóvil fue comprado

VehBCost: Costo de adquisición pagado por el vehículo en el momento de la compra

IsOnLineSale: identifica si el vehículo se compró originalmente en línea.

3. Métricas de desempeño

Como métrica de desempeño de machine learning vamos a utilizar el Accuracy dado que es un problema de clasificación binaria (1 y 0). Es deseable que el Accuracy tenga un valor alto dado que esto dará confiabilidad al modelo puesto que significa con qué porcentaje de acierto cuenta el modelo al momento de predecir malas compras.

$$Accuracy = \frac{\text{Number of correct prediction}}{\text{Total number of predictions}}$$

Como métrica de negocio se podría utilizar el dinero libre que se obtiene después de ponerlo en óptimas condiciones para el cliente cuando una compra es correcta, dentro del cual se puede tener en cuenta el porcentaje de dinero que se gana al confiar en el modelo.

4. Desempeño deseable en producción

Se medirá el desempeño del modelo teniendo en cuenta el porcentaje de ganancia por cada vehículo vendido, cuando éste porcentaje sea mayor a un 15%. Dado que si no cumple con este porcentaje sería una mala compra, dado que no alcanzaría la meta del concesionario.

5. Bibliografía

- Don't Get Kicked! - Predict if a car purchased at auction is a lemon | Kaggle. (2011) Retrieved 5 january 2012, from <https://www.kaggle.com/competitions/DontGetKicked/overview>
- www.DeepL.com/Translator
- Introducción al enmarcado de problemas del aprendizaje automático | (2022) Retrieved 27 september, from <https://developers.google.com/machine-learning/problem-framing?hl=es-419> .