

Primera entrega IA para ingenieros (Don't Get Kicked!)

José Alejandro Urrego Pabón  
Mateo Toro Molina

Inteligencia artificial para ingenieros

Raúl Ramos Pollan  
Docente



25 de septiembre de 2023

Facultad de ingeniería  
Medellin, Antioquia  
Colombia

# Tabla de contenido

<b>1. Descripción del problema</b>	<b>3</b>
<b>2. Descripción del dataset</b>	<b>3</b>
<b>3. Métrica de desempeño</b>	<b>3</b>
3.1. Precisión . . . . .	4
3.2. Accuracy . . . . .	4
<b>4. Criterio de desempeño deseable</b>	<b>4</b>
<b>5. Referencias</b>	<b>4</b>

# 1. Descripción del problema

Los concesionarios que compran autos en gran cantidad durante las subastas son propensos a comprar autos en mal estado o que necesitan gran cantidad de reparaciones debido precisamente a que esos autos se camuflan con los que están en buen estado, los concesionarios y la comunidad automovilista llama a estas desafortunadas compras “Kicks” o patadas.

Nosotros, como estudiantes de ingeniería mecánica, hemos detectado que con un pronóstico más preciso usando técnicas de Machine Learning se le podría dar herramientas a los concesionarios sobre posibles autos averiados y, de esta forma, poder ahorrar mucho dinero en compras “patada”.

Los autos “pateados” o comprados con alguna avería son aquellos que cuentan con tacómetros manipulados, problemas en la carrocería que son imperceptibles pero que ponen en riesgo la vida del conductor, problemas con el traspaso de la propiedad del vehículo y muchos problemas más, estos percances hacen que los concesionarios pierdan mucho dinero en tratar de reparar las averías para revender el vehículo, porque inclusive muchas veces la propia avería es más costosa que el precio del auto.

# 2. Descripción del dataset

El dataset implementado se tomó de una competencia realizada en Kaggle donde reportan 121756 autos recopilados entre los años 2009 y 2010; dicho dataset cuenta con 33 columnas de variables que permiten la correcta identificación y clasificación del auto.

El 60 % del dataset fue destinado para el proceso de entrenamiento y el otro 40 % para el proceso de testeo, además, cerca del 40 % de las variables son categóricas y el otro 60 % son variables numéricas, por ultimo, se tienen muchos valores faltantes o nulos en por lo menos 5 columnas del dataset.

El dataset cuenta con 4 archivos, el primero donde se explican cada una de las variables que conforman dichos archivos (*Carvana\_Data\_Dictionary.txt*), un archivo destinado al entrenamiento del modelo (*Training.csv*), otro archivo destinado al testeo de este (*test.csv*) y un archivo final con las entradas para el testeo (*Example\_entry.csv*) [1]

# 3. Métrica de desempeño

En este problema como solo se define si la futura compra es mala o buena (posible patada o no), por lo tanto, se tiene un problema binario donde 1 correspondería a si y 0 correspondería a no, por lo que se pretenden calcular dos métricas de desempeño.

### 3.1. Precisión

Esta métrica de desempeño nos indica que porcentaje de valores que calificaron como positivos son realmente positivos, es decir, cuales de los autos que son predichos como posibles patadas en realidad lo son. [2, 3]

$$Presicion = \frac{Verdaderopositivo}{Verdaderopositivo + Falsopositivo} = \frac{TP}{TP + FN}$$

### 3.2. Accuracy

Esta métrica es muy similar a la anterior solo que en este caso se cuantifica la totalidad de los valores que han sido fielmente clasificados. [2, 3]

$$Accuracy = \frac{Verdaderopositivo + Verdaderonegativo}{Verdaderopositivoynegativo + Falsopositivoynegativo} = \frac{TP + TN}{TP + TN + FN + FP}$$

## 4. Criterio de desempeño deseable

Con este modelo se augura que podrán aumentar en un 10 % las ganancias por cada auto vendido, dicho porcentaje es bastante conservador por tratarse de los altos precios en cuanto a la compra y venta de los vehículos, por lo que, si se logra este porcentaje, la ganancia neta por la venta de cada vehículo será verdaderamente redituable para el concesionario.

## 5. Referencias

[1] <https://www.kaggle.com/competitions/DontGetKicked/overview>

[2] <https://www.themachinelearners.com/metricas-de-clasificacion>

[3] [https://rramosp.github.io/ai4eng.v1/content/LAB %201.02 %20- %20METRICS.html](https://rramosp.github.io/ai4eng.v1/content/LAB%201.02%20-%20METRICS.html)