

Proyecto Data Science: Automotora Anaconda

GRUPO:

- PAULO GONZÁLEZ
- JORGE GUERRERO
- DANIEL MARDONES
- GONZALO ROJAS
- ESTEBAN SÁNCHEZ

PROFESOR:

- DANIEL BEFFERMANN

TUTOR:

- JHON POMA

PRESENTACIÓN 1:

- HITO 1

FECHA:

- 26/01/2023

Equipo de trabajo: roles para el desarrollo del proyecto

Líder de equipo:
Paulo González /
Jorge Guerrero

Analista de datos:
Esteban Sánchez /
Gonzalo Rojas

Control de calidad y
validación de datos:
Esteban Sánchez /
Gonzalo Rojas

Ingeniero de
modelamiento =
Daniel Mardones /
Jorge Guerrero

Visualización de
datos = Daniel
Mardones / Jorge
Guerrero

Documentador =
Daniel Mardones /
Paulo González

Selección del tema y motivación

Tema seleccionado: Automotora Anaconda

Motivación: generar una herramienta de apoyo a la toma de decisiones de los equipos de compra y venta, para obtener las mejores opciones existentes en el mercado, mejorando la rentabilidad del negocio.

Planificación de la investigación: Requerimientos

Objetivos

Generar una herramienta que ayude al equipo de compras a encontrar las mejores oportunidades.

- Modelo de clasificación con técnicas de Machine Learning para encontrar mejor rentabilidad

Definir una propuesta de inversión (lista de automóviles convenientes) para la Automotora.

- A través de las métricas obtenidas se define una propuesta de inversión relacionada a los vehículos que estén por debajo del precio de venta promedio, estableciendo un margen de venta

Definir una estrategia de venta por gama de vehículos

- Se crea una nueva columna que clasifique la marca del vehículo por generalista (0) o premium (1), para posteriormente generar un análisis del comportamiento en la venta de vehículos y así desarrollar una estrategia

Encontrar oportunidades de compra y venta entre ciudades/estados.

- A través de un EDA, podemos determinar en que ciudades / estados se encuentran las mejores oportunidades de compra y determinar las mejores ciudades / estados para vender estos vehículos, maximizando la utilidad.

Planificación de la investigación

Hipótesis:

- Nuestra hipótesis es poder determinar, mediante un modelo clasificatorio si es rentable la compra de un vehículo, de acuerdo a sus especificaciones, para una posterior venta

Vector objetivo:

- El vector objetivo es la variable “Price”.

Estrategias analíticas a nivel descriptivo:

- Se desarrollará un Análisis Exploratorio de Datos (EDA), lo cual permitirá visualizar la distribución de vehículos de acuerdo: marca, modelos, año, ciudad, estado, millas y valores de vehículos.

Planificación de la investigación: Control de calidad, validación y análisis de datos

Control de calidad, validación y análisis de datos



```
graph TD; A[Control de calidad, validación y análisis de datos] --> B[Enriquecimiento del data set a través de webscrapping para obtener la mayor cantidad de información del vehículo. Se realizará mediante la columna "VIN"]; B --> C[Creación de base de datos para almacenamiento de información]; C --> D[Normalización y validación de datos]; D --> E[Creación de nueva columna con información de la gama del vehículo: generalista (0) o premium (1)];
```

Enriquecimiento del data set a través de webscrapping para obtener la mayor cantidad de información del vehículo. Se realizará mediante la columna "VIN"

Creación de base de datos para almacenamiento de información

Normalización y validación de datos

Creación de nueva columna con información de la gama del vehículo:
generalista (0) o premium (1)

Planificación de la investigación: Modelación y predicción de trabajo

Implementación estándar de la librería LazyPredict para conocer de forma preliminar los mejores clasificadores para esta solución.

Selección de los 5 mejores modelos a partir del paso anterior, e implementación de ellos con hiperparámetros a definir.

Debido a que se trata de un problema de clasificación, las métricas a utilizar para seleccionar el mejor modelo serán: precisión, recall y f1-score.

Para almacenamiento de resultados, utilizaremos una base de datos Postgres, lo que permite obtener mediante queries la conveniencia de la compra de un vehículo.

Consideraciones del proyecto

