



Primera entrega de proyecto

Por:

Luis David Morales

Materia:

Introducción a la inteligencia artificial

Profesor:

Raul Ramos Pollan

UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERÍA

MEDELLÍN 2022

1. Planteamiento del problema.

Para entrar en un mercado y ser competitivo, las empresas deben realizar estudios sobre el producto que quieren comercializar en determinada región. Estos estudios se basan en las necesidades que deben suplir al cliente con los artículos que ofrecen, y al mercado existente en la zona por parte de otras empresas. Para que una empresa de automóviles extranjera pueda entrar al comercio de autos local, deben entender los factores que afectan el precio de los autos locales ya que se pueden presentar diferencias significativas con los factores propios. Para esto deben identificar las variables más significativas para poder predecir el precio de un vehículo. Una vez se identifica esto, se ajusta un valor adecuado al producto con el fin de garantizar una ganancia al fabricante y a su vez generar competitividad en el mercado.

2. Dataset

El dataset a utilizar se encuentra en la plataforma Kaggle (https://www.kaggle.com/datasets/hellbuoy/car-price-prediction?select=CarPrice_Assignment.csv) con el título de “Car price prediction multiple linear regression”. Este dataset consiste en información de las variables más relevantes que se deben tener en cuenta para definir el precio de un vehículo, contando con 26 columnas y 205 filas. Las columnas se identifican como sigue:

'car_ID', 'symboling', 'CarName', 'fueltype', 'aspiration', 'doornumber', 'carbody', 'drivewheel', 'engine location', 'wheelbase', 'carlength', 'carwidth', 'carheight', 'curbweight', 'enginetype', 'cylindernumber', 'enginesize', 'fuelsystem', 'bore ratio', 'stroke', 'compressionratio', 'horsepower', 'peakrpm', 'citympg', 'highwaympg', 'price'.

Dentro de los archivos que se pueden obtener de Kaggle, se encuentra uno llamado "CarPrice_Assignment.csv", en el cual se encuentra toda la información necesaria para abrir y utilizar el dataset en pandas.

3. Métricas

Una de las métricas a emplear es el “Recursive Feature Elimination” (RFE). Esta métrica sirve para seleccionar las características más relevantes en un conjunto de datos, algo necesario para este dataset y lo que se quiere predecir. Otra de las métricas seleccionadas es el “Root Mean Squared Logarithmic Error” (RMSLE), seleccionada para indicar el error porcentual de acierto entre los datos reales y predichos. Para el precio de los vehículos, una empresa requiere que el error de predicción sea igual o inferior al 30%, por lo que el precio puede ser 30% superior o inferior al original. Como métrica de negocio se espera que el error en la predicción sea el más bajo posible, y superior al valor original, para garantizar ganancia y competitividad en el mercado de vehículos local.

4. Desempeño

Se espera entonces que las predicciones de los datos que se evalúen con el modelo, no tengan un error en pérdida superior al 30%, ya que esto garantiza que por lo menos se suple el precio de producción de los vehículos. Aunque si más del 10% de los datos predichos tienen un porcentaje de error superior al 30%, no habrá un beneficio económico que motive el desarrollo del negocio y por lo tanto no se recomienda realizar llevar a cabo el negocio.

5. Bibliografía

- Car Price Prediction Multiple Linear Regression. (2022). Retrieved from https://www.kaggle.com/datasets/hellbuoy/car-price-prediction?select=CarPrice_Assignment.csv
- Metrics - MSE, R^2 , RMSLE. (2022). Retrieved from <https://hrngok.github.io/posts/metrics/>
- Brownlee, J. (2022). Recursive Feature Elimination (RFE) for Feature Selection in Python. Retrieved 3 July 2022, from <https://machinelearningmastery.com/rfe-feature-selection-in-python/>