Recell Your Cellphones, Tablets, and More!

SLF Recell Project and Supervised Learning- Foundations

Rae Downen
August 4, 2023

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- EDA Results
- Data Preprocessing
- Model Performance Summary
- Appendix

Executive Summary

If the data was also the inventory:

- I would increase inventory of newer models, so that the largest part of the inventory was not your oldest phones.
- I would increase inventory of Apple products. Android is 93% of the current inventory.
- To increase the Apple inventory, offer specials that encourage people to sell their devices.

For the actual modeling:

- I would scale the numerical data.
- I would use IQR to address the outliers.
- To prevent data leakage, I would not use the target variable during exploration.

Business Problem Overview and Solution Approach

- Recell needs a dynamic pricing strategy, using machine learning, for used and refurbished devices, so that the company will make a large profit.
- A solution was found by analyzing the data and identifying various factors that influence price, I was able to build a linear regression model to predict the price of the used devices.

EDA Results

- Univariate Exploration: Android was top os (93.1%).
- Bivariate Exploration: Battery (.81) and weight (.83) were highly correlated.
- Release year, as expected, influenced the price the newer the phone, the higher the price.

Link to Appendix slide on data background check

Data Preprocessing

- Duplicate value check: The data had no duplicate observations.
- Missing value treatment: The missing data was imputed with the median.
- Outlier check (treatment if needed): There were outliers.
- Feature engineering: The column ‘release_year” was transformed by subtracting it from the year the data was collected.
- Data preparation for modeling: The categorical features were encoded. The data was split into train and test.

Model Performance Summary

- The model is able to explain 84% of the variation in the data.
- The train/test RMSE is pretty similar and low, so it is not overfitting.
- The MAPE on the test set suggests that we can predict with 4.5% of the prices.
- So, all in all, it is a pretty good model for prediction.

RMSE	MAE	R-squared	Adj. R squared	MAPE
0.231426	0.181502	0.842797	0.84122	4.344607

RMSE	MAE	R-squared	Adj. R squared	MAPE
0.240476	0.186263	0.839667	0.835864	4.511603

Link to Appendix slide on model assumptions

APPENDIX

Data Background and Contents

- The data was collected in 2021. In its original form the data had 3454 observations with 15 different features listed here:

- brand_name: Name of manufacturing brand
- os: OS on which the device runs
- screen_size: Size of the screen in cm
- 4g: Whether 4G is available or not
- 5g: Whether 5G is available or not
- main_camera_mp: Resolution of the rear camera in megapixels
- selfie_camera_mp: Resolution of the front camera in megapixels
- int_memory: Amount of internal memory (ROM) in GB
- ram: Amount of RAM in GB
- battery: Energy capacity of the device battery in mAh
- weight: Weight of the device in grams
- release_year: Year when the device model was released
- days_used: Number of days the used/refurbished device has been used
- normalized_new_price: Normalized price of a new device of the same model in euros
- normalized_used_price: Normalized price of the used/refurbished device in euros

Model Assumptions

- MULTICOLLINEARITY: Tested by using VIF. Dropped screen_size, weight, and years_since_released.
- LINEARITY AND INDEPENDENCE: Tested by by making a plot of fitted values vs residuals and checking for patterns. No pattern was found, so it checked out.
- NORMALITY: Tested by checking distribution of residuals, Q-Q plot, Shapiro Wilkes test. The results were affirmative: the residuals were normally distributed, Q-Q plot showed a straightish line, and the Shapiro-Wilkes Test had a p-value (4.22) greater than 0.05.
- HOMOSCEDASTICITY: Tested by Goldfeldquandt test. Its p-value (0.15) was greater that 0.05

Great Learning

Happy Learning !