AFA COMPANY

By: - Asma Alsubaie - Fatimah Alghamdi - Alanoud Alqubaysi

Abstract

The goal of our project is to take advantage and use the available datasets in some online shopping stores and train a model to predict the price range of our product. This kind of prediction will help company estimate price of mobiles to give tough competition to other mobile manufacturers. Also it will be useful for Consumers to verify that they are paying best price for a mobile. The dataset obtains through web scraping which is used to extract content and data from the chosen websites. The plan is using a Linear Regression model/algorithm to train our dataset which will predict the price range.

Design

The data is provided by Gsmarena website by scraping, and then taking the data that come from 2010 until now to be appropriate and match nowadays prices, And we predict the prices by using the linear model that aid to clarify the Influencing characteristics.

Data

In this project, data was brought from the Gsmarena website. It was collected 50 columns about mobile phones specifications such as price, size, likes, ... etc. Also, it was collected 4508 raws that describes every device specification. The data was transferred to numerical data in order to deal with it.

Algorithms

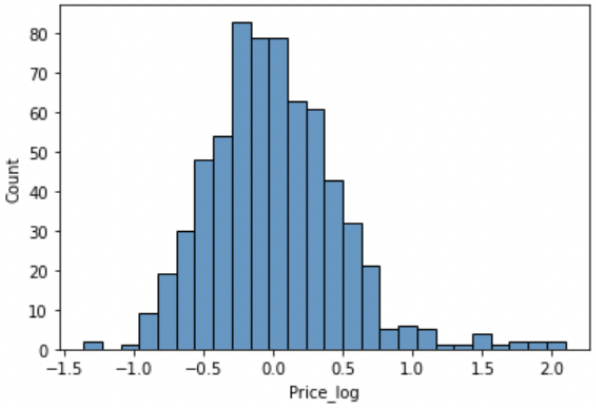
1 We access to page “http://www.gsmarena.com/makers.php3” and store all the data that contains the brand names of all available gadget information provided by GSMArena.
2 Choose the most important feature that effect the price and help us to predict it.
3 Filter and clean data from 2010 to 2021 to be appropriate and match nowadays prices.
4 Convert data into numeric data.
5 Drop outliers and show the correlation between target and features.
6 Split data into train and test.
7 Build Linear Regression Model.
8 Check the RMSE (root mean squared error) .
9 Choose the best model.

Tools

-Numpy and Pandas for data manipulation.
-BeautifulSoup, Webdriver and Jupyter.
-Matplotlib , Seaborn, Sklearn and Numpy.

Communication

First Linear Regression Model




```
2.043954296573887
```