Mobile Phones: Price Estimation

Ludovica Di Silvestre, 14-01-21

Table of contents

Mission

As a **mobile phone company**, price definition is extremely important step **before launching**

Goals of **Machine Learning**

- Predict the **price range** each phone should be sold
- Relationships between **features** (ex. battery, ram etc.) and **price range**

Company Strategy

Price Range (y)

	Values	Profitability
Very Low	0	$
Low	1	$$
Medium	2	$$$
High	3	$$$$

Mobile Phone Features (X)

Battery

Total energy a battery can store in one time

Internal Memory

Internal Memory(storage space installed)

Ram

Random Access Memory (that means more RAM can let more apps run in the background without slowing your phone down)

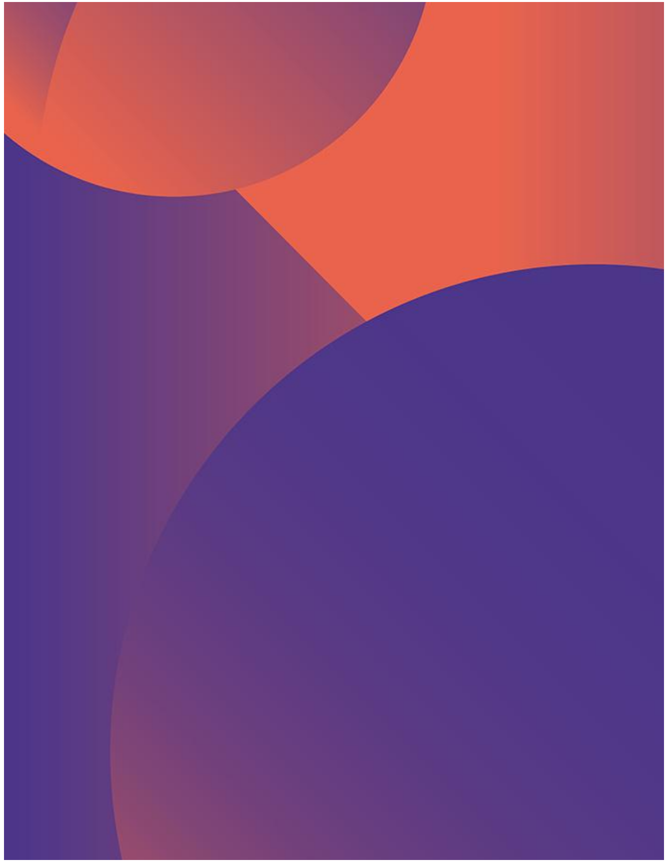
Clock Speed

Speed at which microprocessor executes instructions

If Prices Predicted are lower than Real Prices → company loses potential profits $

	Actual Price	Predicted Price	Business Result (Profitability)
Predicted Prices = Actual Prices	Medium $$$	Medium $$$	✓
Predicted Prices > Actual Prices	Low $	High $$$$	✓
Actual Prices > Predicted Prices	Medium $$$	Low $	✗

Project Timeline



ML Tasks

Explore Data

Define Metric

Based on the company profit **strategy**

$$\frac{y-\hat{y}}{\hat{y}+1}$$

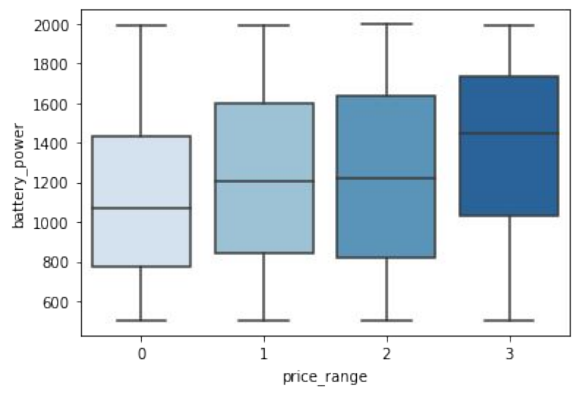
Scale Data

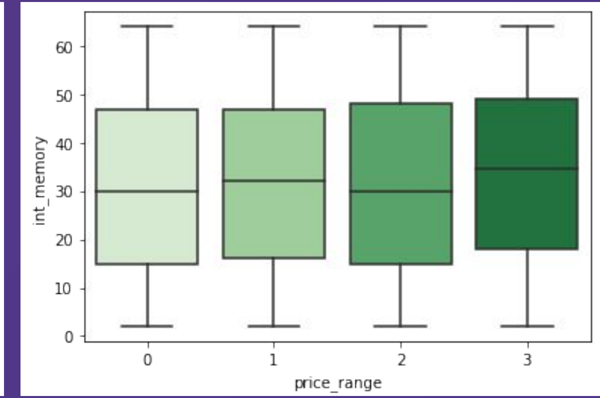
Models

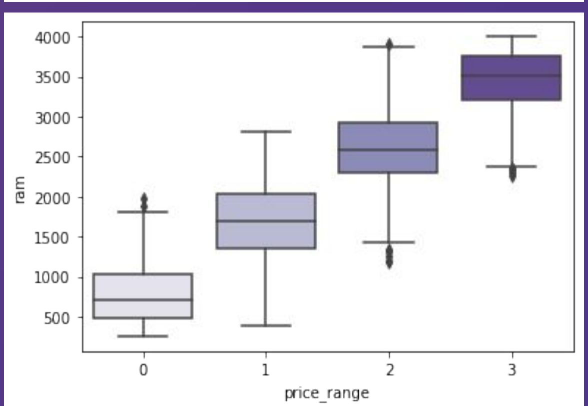
Summary

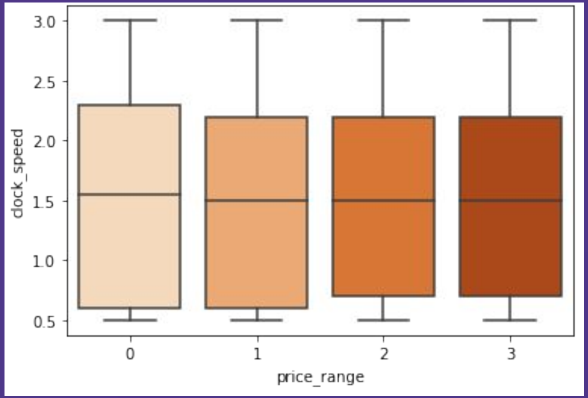
Exploratory Analysis

Relationship: Price and Features









Mobile phones with higher amount of ***RAM memory*** are clearly priced higher (**range=3**)

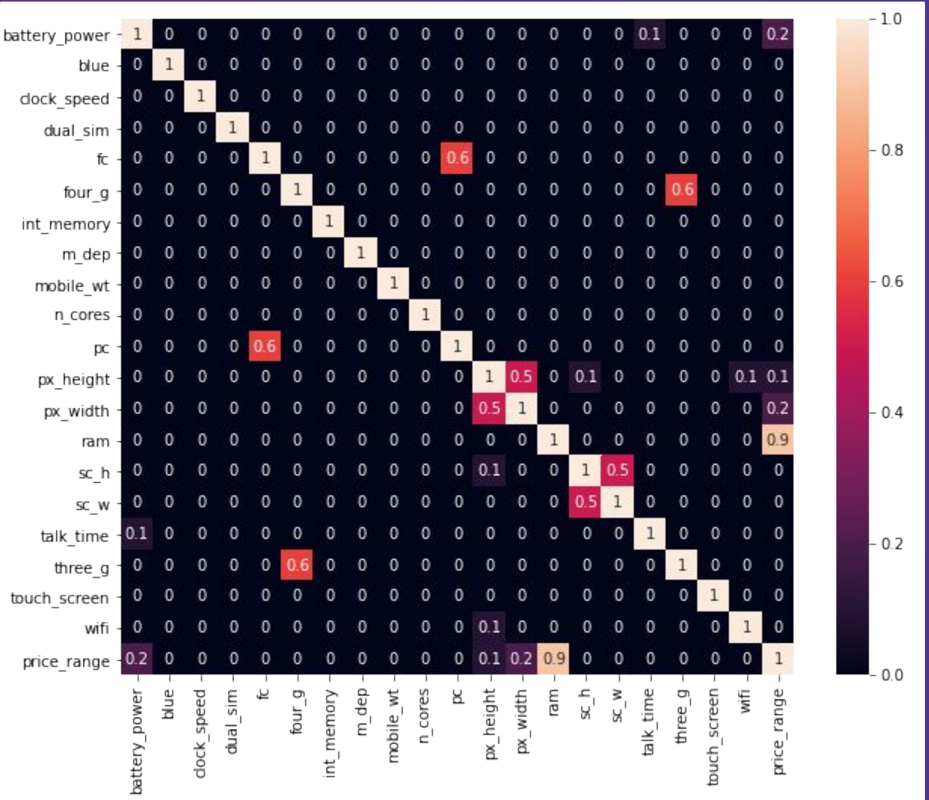
Similar pattern with ***battery power*** even if it looks that other features are more important in influcing the price range

Clock speed and *internal memory* are not key factor to influence the price range

Statistics: data are not scaled

	count	mean	std	min	25%	50%	75%	max
battery_power	2000.0	1238.51850	439.418206	501.0	851.75	1226.0	1615.25	1998.0
blue	2000.0	0.49500	0.500100	0.0	0.00	0.0	1.00	1.0
clock_speed	2000.0	1.52225	0.816004	0.5	0.70	1.5	2.20	3.0
dual_sim	2000.0	0.50950	0.500035	0.0	0.00	1.0	1.00	1.0
fc	2000.0	4.30950	4.341444	0.0	1.00	3.0	7.00	19.0
four_g	2000.0	0.52150	0.499662	0.0	0.00	1.0	1.00	1.0
int_memory	2000.0	32.04650	18.145715	2.0	16.00	32.0	48.00	64.0
m_dep	2000.0	0.50175	0.288416	0.1	0.20	0.5	0.80	1.0
mobile_wt	2000.0	140.24900	35.399655	80.0	109.00	141.0	170.00	200.0
n_cores	2000.0	4.52050	2.287837	1.0	3.00	4.0	7.00	8.0
pc	2000.0	9.91650	6.064315	0.0	5.00	10.0	15.00	20.0
px_height	2000.0	645.10800	443.780811	0.0	282.75	564.0	947.25	1960.0
px_width	2000.0	1251.51550	432.199447	500.0	874.75	1247.0	1633.00	1998.0
ram	2000.0	2124.21300	1084.732044	256.0	1207.50	2146.5	3064.50	3998.0
sc_h	2000.0	12.30650	4.213245	5.0	9.00	12.0	16.00	19.0
sc_w	2000.0	5.76700	4.356398	0.0	2.00	5.0	9.00	18.0
talk_time	2000.0	11.01100	5.463955	2.0	6.00	11.0	16.00	20.0
three_g	2000.0	0.76150	0.426273	0.0	1.00	1.0	1.00	1.0
touch_screen	2000.0	0.50300	0.500116	0.0	0.00	1.0	1.00	1.0
wifi	2000.0	0.50700	0.500076	0.0	0.00	1.0	1.00	1.0
price_range	2000.0	1.50000	1.118314	0.0	0.75	1.5	2.25	3.0

Multicollinearity



Machine Learning

Supervised Machine Learning Model for **categorical** variable (price range)

Machine Learning

Question

Can we predict the price range of mobile phones?

Input

Mobile Phones Features (Ram, Battery etc.)

Output

Price Range (Very Low, Low, Medium, High)

Machine Learning Models

Models Developed: Logistic Regressions, AdaBoost, K-Nearest Neighbors, Support Vector Machine, Gaussian Naive Bayes,Decision Tree, Random Forest, CatBoost

Evaluation: the most accurate models in predicting the prices are: **Catboost** and **SVM**

1. **SVM**: It has the lowest value of the score of the the *New Metric*, so it is the best (the metric was created based on the company profitability strategy→ where this assumption rules: **actual price (y)** **<=** **predicted price ŷ**

2. **CatBoost:** it has the second **highest scores** in **all** the standard evaluation metrics (excluding new metric),

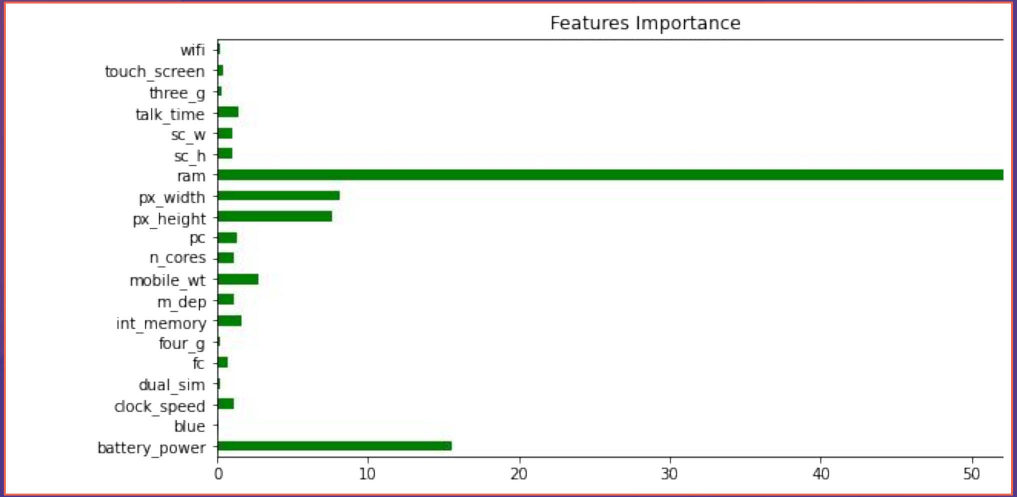
	Accuracy	Precision	Recall	F1	New Metric
LR	0.250375	0.250375	0.250375	0.250375	108.345827
KNN	0.250375	0.250375	0.250375	0.250375	108.345827
Ada	0.644678	0.644678	0.644678	0.644678	20.977011
NB	0.806597	0.806597	0.806597	0.806597	8.870565
DT	0.823088	0.823088	0.823088	0.823088	8.42079
RF	0.881559	0.881559	0.881559	0.881559	5.134933
Cat	0.946027	0.946027	0.946027	0.946027	2.373813
SVM	0.958021	0.958021	0.958021	0.958021	1.686657

$$\frac{y-\hat{y}}{\hat{y}+1}$$

Features Importance:

- **SVM**: The feature importance of a **nonlinear SVMs** can not be found out. The reason being is that, when the SVM is non-linear the dataset is mapped into a space of higher dimension

- **CatBoost:** the most relatively important feature for the prediction is **RAM** memory, followed by the **battery power** of the mobile phone.
  → NB: this match with the observation done during the **explaratory analysis**



Q&A