A Classification Based approach for predicting Smartphone Price Categories

Sayan Das - B2430035 **Raihan Uddin** - B2430070

November 25, 2024

Outline

Background

- The smartphone industry experiences continuous technological innovations, with manufacturers introducing advanced features.
- Multiple global players, such as Apple, Samsung, and Xiaomi, vie for market share, leading to frequent product launches and **pricing battles**.
- Consumers demand **value for money**, with preferences shifting toward devices offering high performance at competitive prices.

Motivation

We hope our model will help:

- Simplify pricing strategies for **manufacturers**
- Increase pricing transparency for **consumers**

Motivation

We hope our model will help:

- Simplify pricing strategies for **manufacturers**
- Increase pricing transparency for **consumers**

Objectives

- Develop a Robust Classification Model
- Determine the most influential factors behind smartphone pricing
- Compare algorithm performances
- Practical Applicability
- Methodological Contribution

Objectives

- Develop a Robust Classification Model
- Determine the most influential factors behind smartphone pricing
- Compare algorithm performances
- Practical Applicability
- Methodological Contribution

Objectives

- Develop a Robust Classification Model
- Determine the most influential factors behind smartphone pricing
- Compare algorithm performances
- Practical Applicability
- Methodological Contribution

Objectives

- Develop a Robust Classification Model
- Determine the most influential factors behind smartphone pricing
- Compare algorithm performances
- Practical Applicability
- Methodological Contribution

Objectives

- Develop a Robust Classification Model
- Determine the most influential factors behind smartphone pricing
- Compare algorithm performances
- Practical Applicability
- Methodological Contribution

Source

Link - https://www.kaggle.com/datasets/iabhishekofficial/smartphone-price-classification

The dataset is publicly available and contains **2000 smartphone entries** with **19 feature variables** and **1 target variable** representing `price_range`.

Features

Feature Name	Description	Type
battery_power	battery capacity in mAh	Numerical
clock_speed	speed at which processor executes instructions	Numerical
fc	front Camera Megapixels	Numerical
pc	primary Camera Megapixels	Numerical
int_memory	internal Memory capacity	Numerical
m_dep	smartphone Depth in cm	Numerical
mobile_wt	weight of the smartphone	Numerical
n_cores	number of cores in processor	Numerical
px_height	pixel Resolution Height	Numerical
px_width	pixel Resolution Width	Numerical
ram	RAM in MB	Numerical
sc_h	screen Height in cm	Numerical
sc_w	screen Width in cm	Numerical
talk_time	longest time that a single battery charge will last over a call	Numerical
blue	has bluetooth or not	Categorical
dual_sim	has dual sim support or not	Categorical
four_g	has 4G or not	Categorical
three_g	has 3G or not	Categorical
wifi	has wifi or not	Categorical
touch_screen	has touch screen or not	Categorical

Data Cleaning

Handling Missing Values: A check for missing values revealed no missing values in the dataset.

Handling Duplicate Values: Also, the dataset was checked for duplicate entries and none were found.

Handling Invalid Values: The dataset was checked for negative entries and none were found in the dataset.

There are some features which can not be zero, like `battery_power`, `ram`, etc. So, we checked for zero values in these features. We saw `px_height` and `sc_w` have 2 and 180 zero values respectively. We replaced these zero values with the mean of the respective features.

Data Cleaning

Handling Missing Values: A check for missing values revealed no missing values in the dataset.

Handling Duplicate Values: Also, the dataset was checked for duplicate entries and none were found.

Handling Invalid Values: The dataset was checked for negative entries and none were found in the dataset.

There are some features which can not be zero, like `battery_power`, `ram`, etc. So, we checked for zero values in these features. We saw `px_height` and `sc_w` have 2 and 180 zero values respectively. We replaced these zero values with the mean of the respective features.

Data Cleaning

Handling Missing Values: A check for missing values revealed no missing values in the dataset.

Handling Duplicate Values: Also, the dataset was checked for duplicate entries and none were found.

Handling Invalid Values: The dataset was checked for negative entries and none were found in the dataset.

There are some features which can not be zero, like `battery_power`, `ram`, etc. So, we checked for zero values in these features. We saw `px_height` and `sc_w` have 2 and 180 zero values respectively. We replaced these zero values with the mean of the respective features.

Acknowledgement

References

Thank You!