DATS 6202

Machine Learning I

Team 11: Mobile Price Classification

Data Science, Columbian College of Arts & Sciences

Problem Statement

- Bob has started his own mobile company and wants to compete with big players like Apple and Samsung. However, he's unsure how to price his mobile phones competitively. To tackle this, he gathers sales data from various companies to understand the relationship between different features (e.g., RAM, internal memory) and selling prices. Bob needs assistance in analyzing this data and building a model to predict price ranges for his mobile phones based on their specifications. By doing so, Bob aims to make informed pricing decisions and establish his company in the competitive mobile phone market.

Solution

Our aim is to predict the price range of mobile phones based on their specifications. We'll explore different classification models to find the one that best predicts phone prices accurately. By comparing models using metrics like accuracy, we'll identify the most effective one for this task.

Implementation

The Pipeline for Regression and Classification:

- Loading the data
- Splitting the data
- Handling uncommon features
- Handling identifiers
- Handling date time variables
- Handling missing data
- Encoding the data
- Splitting the feature and target
- Scaling the data
- Feature selection & feature engineering

Data Description

- 21 columns
  - ✓ **battery_power**: The total energy a battery can store in milliampere-hours (mAh).
  - ✓ **blue**: Indicates the presence of Bluetooth (1 if present, 0 if absent).
  - ✓ **clock_speed**: The speed at which the microprocessor executes instructions, measured in gigahertz (GHz).
  - ✓ **dual_sim**: Indicates whether the device supports dual SIM cards (1 if supported, 0 if not supported).
  - ✓ **fc**: Front Camera megapixels.
  - ✓ **four_g**: Indicates the presence of 4G connectivity (1 if present, 0 if absent).
  - ✓ **int_memory**: Internal Memory capacity in gigabytes (GB).
  - ✓ **m_dep**: Mobile Depth in cm.
  - ✓ **mobile_wt**: Weight of mobile phone in grams.
  - ✓ **n_cores**: Number of cores of the processor.

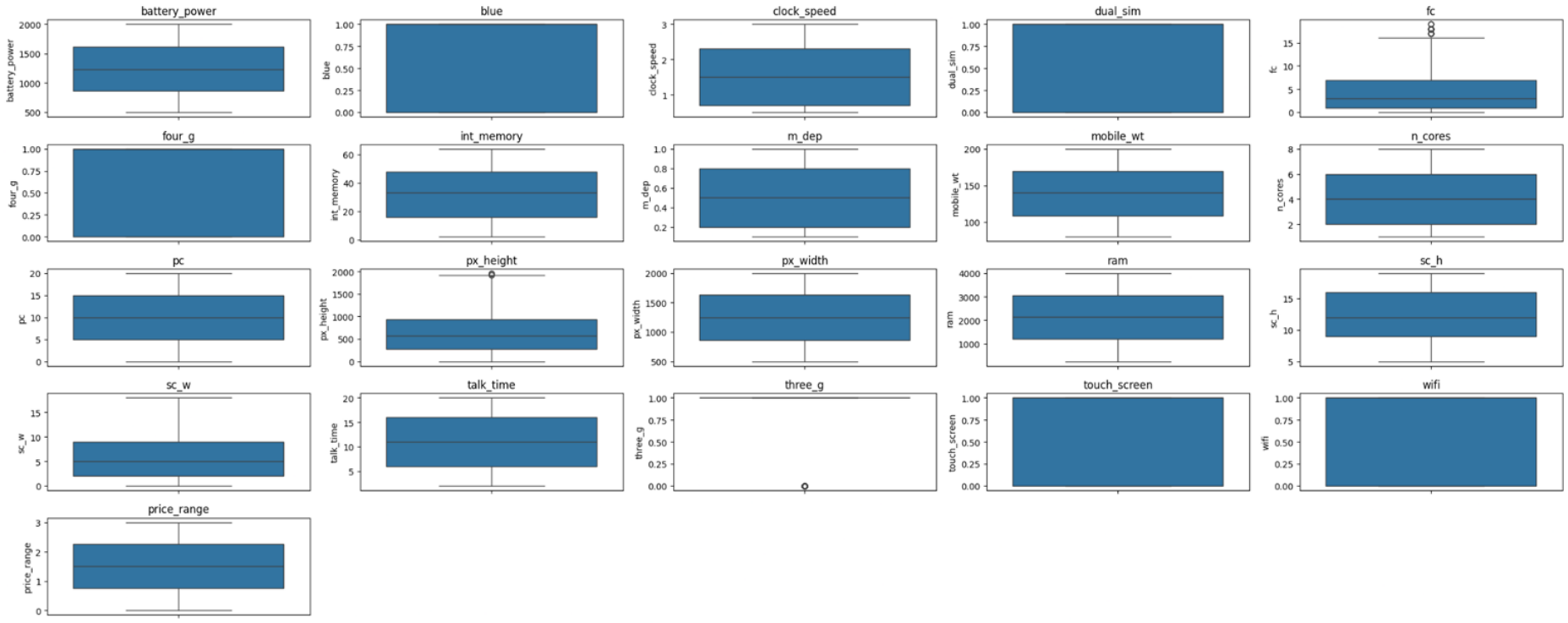
Data Description

- 21 columns
  - ✓ **pc**: Primary Camera megapixels.
  - ✓ **px_height**: Pixel Resolution Height.
  - ✓ **px_width**: Pixel Resolution Width.
  - ✓ **ram**: Random Access Memory capacity in megabytes (MB).
  - ✓ **sc_h**: Screen Height of mobile in cm.
  - ✓ **sc_w**: Screen Width of mobile in cm.
  - ✓ **talk_time**: The longest time that a single battery charge will last when you are on a call.
  - ✓ **three_g**: Indicates the presence of 3G connectivity (1 if present, 0 if absent).
  - ✓ **touch_screen**: Indicates whether the device has a touch screen (1 if present, 0 if absent).
  - ✓ **wifi**: Indicates whether the device supports Wi-Fi (1 if supported, 0 if not supported).
  - ✓ **price_range**: This is the target variable with values indicating the price range of the mobile phone.

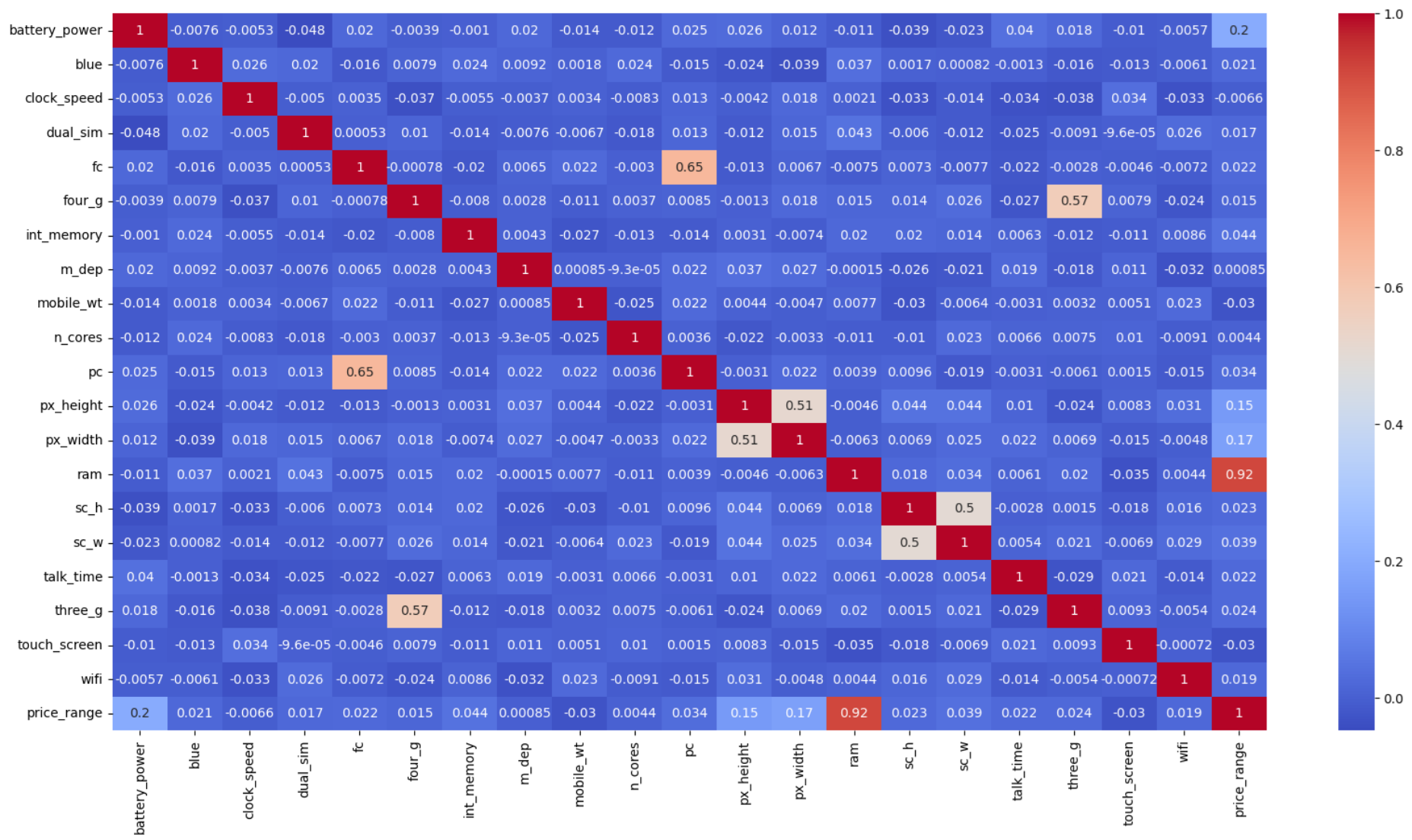
Data Description

- 21 columns
  - ✓ **pc**: Primary Camera megapixels.
  - ✓ **px_height**: Pixel Resolution Height.
  - ✓ **px_width**: Pixel Resolution Width.
  - ✓ **ram**: Random Access Memory capacity in megabytes (MB).
  - ✓ **sc_h**: Screen Height of mobile in cm.
  - ✓ **sc_w**: Screen Width of mobile in cm.
  - ✓ **talk_time**: The longest time that a single battery charge will last when you are on a call.
  - ✓ **three_g**: Indicates the presence of 3G connectivity (1 if present, 0 if absent).
  - ✓ **touch_screen**: Indicates whether the device has a touch screen (1 if present, 0 if absent).
  - ✓ **wifi**: Indicates whether the device supports Wi-Fi (1 if supported, 0 if not supported).
  - ✓ **price_range**: This is the target variable with values indicating the price range of the mobile phone.

EDA


battery_power
blue
clock_speed
dual_sim
fc
four_g
int_memory
m_dep
mobile_wt
n_cores
pc
px_height
px_width
ram
sc_h
sc_w
talk_time
three_g
touch_screen
wifi
price_range

Heat Map


battery_power 1 -0.0076 -0.0053 -0.048 0.02 -0.0039 -0.001 0.02 -0.014 -0.012 0.025 0.026 0.012 -0.011 -0.039 -0.023 0.04 0.018 -0.01 -0.0057 0.2
blue -0.0076 1 0.026 0.02 -0.016 0.0079 0.024 0.0092 0.0018 0.024 -0.015 -0.024 -0.039 0.037 0.0017 0.00082 -0.0013 -0.016 -0.013 -0.0061 0.021
clock_speed -0.0053 0.026 1 -0.005 0.0035 -0.037 -0.0055 -0.0037 0.0034 -0.0083 0.013 -0.0042 0.018 0.0021 -0.033 -0.014 -0.034 -0.038 0.034 -0.033 -0.0066
dual_sim -0.048 0.02 -0.005 1 0.00053 0.01 -0.014 -0.0076 -0.0067 -0.018 0.013 -0.012 0.015 0.043 -0.006 -0.012 -0.025 -0.0091 -9.6e-05 0.026 0.017
fc 0.02 -0.016 0.0035 0.00053 1 -0.00078 -0.02 0.0065 0.022 -0.003 0.65 -0.013 0.0067 -0.0075 0.0073 -0.0077 -0.022 -0.0028 -0.0046 -0.0072 0.022
four_g -0.0039 0.0079 -0.037 0.01 -0.00078 1 -0.008 0.0028 -0.011 0.0037 0.0085 -0.0013 0.018 0.015 0.014 0.026 -0.027 0.57 0.0079 -0.024 0.015
int_memory -0.001 0.024 -0.0055 -0.014 -0.02 -0.008 1 0.0043 -0.027 -0.013 -0.014 0.0031 -0.0074 0.02 0.02 0.014 0.0063 -0.012 -0.011 0.0086 0.044
m_dep 0.02 0.0092 -0.0037 -0.0076 0.0065 0.0028 0.0043 1 0.00085 -9.3e-05 0.022 0.037 0.027 -0.00015 -0.026 -0.021 0.019 -0.018 0.011 -0.032 0.00085
mobile_wt -0.014 0.0018 0.0034 -0.0067 0.022 -0.011 -0.027 0.00085 1 -0.025 0.022 0.0044 -0.0047 0.0077 -0.03 -0.0064 -0.0031 0.0032 0.0051 0.023 -0.03
n_cores -0.012 0.024 -0.0083 -0.018 -0.003 0.0037 -0.013 -9.3e-05 -0.025 1 0.0036 -0.022 -0.0033 -0.011 -0.01 0.023 0.0066 0.0075 0.01 -0.0091 0.0044
pc 0.025 -0.015 0.013 0.013 0.65 0.0085 -0.014 0.022 0.022 0.0036 1 -0.0031 0.022 0.0039 0.0096 -0.019 -0.0031 -0.0061 0.0015 -0.015 0.034
px_height 0.026 -0.024 -0.0042 -0.012 -0.013 -0.0013 0.0031 0.037 0.0044 -0.022 -0.0031 1 0.51 -0.0046 0.044 0.044 0.01 -0.024 0.0083 0.031 0.15
px_width 0.012 -0.039 0.018 0.015 0.0067 0.018 -0.0074 0.027 -0.0047 -0.0033 0.022 0.51 1 -0.0063 0.0069 0.025 0.022 0.0069 -0.015 -0.0048 0.17
ram -0.011 0.037 0.0021 0.043 -0.0075 0.015 0.02 -0.00015 0.0077 -0.011 0.0039 -0.0046 -0.0063 1 0.018 0.034 0.0061 0.02 -0.035 0.0044 0.92
sc_h -0.039 0.0017 -0.033 -0.006 0.0073 0.014 0.02 -0.026 -0.03 -0.01 0.0096 0.044 0.0069 0.018 1 0.5 -0.0028 0.0015 -0.018 0.016 0.023
sc_w -0.023 0.00082 -0.014 -0.012 -0.0077 0.026 0.014 -0.021 -0.0064 0.023 -0.019 0.044 0.025 0.034 0.5 1 0.0054 0.021 -0.0069 0.029 0.039
talk_time 0.04 -0.0013 -0.034 -0.025 -0.022 -0.027 0.0063 0.019 -0.0031 0.0066 -0.0031 0.01 0.022 0.0061 -0.0028 0.0054 1 -0.029 0.021 -0.014 0.022
three_g 0.018 -0.016 -0.038 -0.0091 -0.0028 0.57 -0.012 -0.018 0.0032 0.0075 -0.0061 -0.024 0.0069 0.02 0.0015 0.021 -0.029 1 0.0093 -0.0054 0.024
touch_screen -0.01 -0.013 0.034 -9.6e-05 -0.0046 0.0079 -0.011 0.011 0.0051 0.01 0.0015 0.0083 -0.015 -0.035 -0.018 -0.0069 0.021 0.0093 1 -0.00072 -0.03
wifi -0.0057 -0.0061 -0.033 0.026 -0.0072 -0.024 0.0086 -0.032 0.023 -0.0091 -0.015 0.031 -0.0048 0.0044 0.016 0.029 -0.014 -0.0054 -0.00072 1 0.019
price_range 0.2 0.021 -0.0066 0.017 0.022 0.015 0.044 0.00085 -0.03 0.0044 0.034 0.15 0.17 0.92 0.023 0.039 0.022 0.024 -0.03 0.019 1
battery_power blue clock_speed dual_sim fc four_g int_memory m_dep mobile_wt n_cores pc px_height px_width ram sc_h sc_w talk_time three_g touch_screen wifi price_range
1.0 0.8 0.6 0.4 0.2 0.0

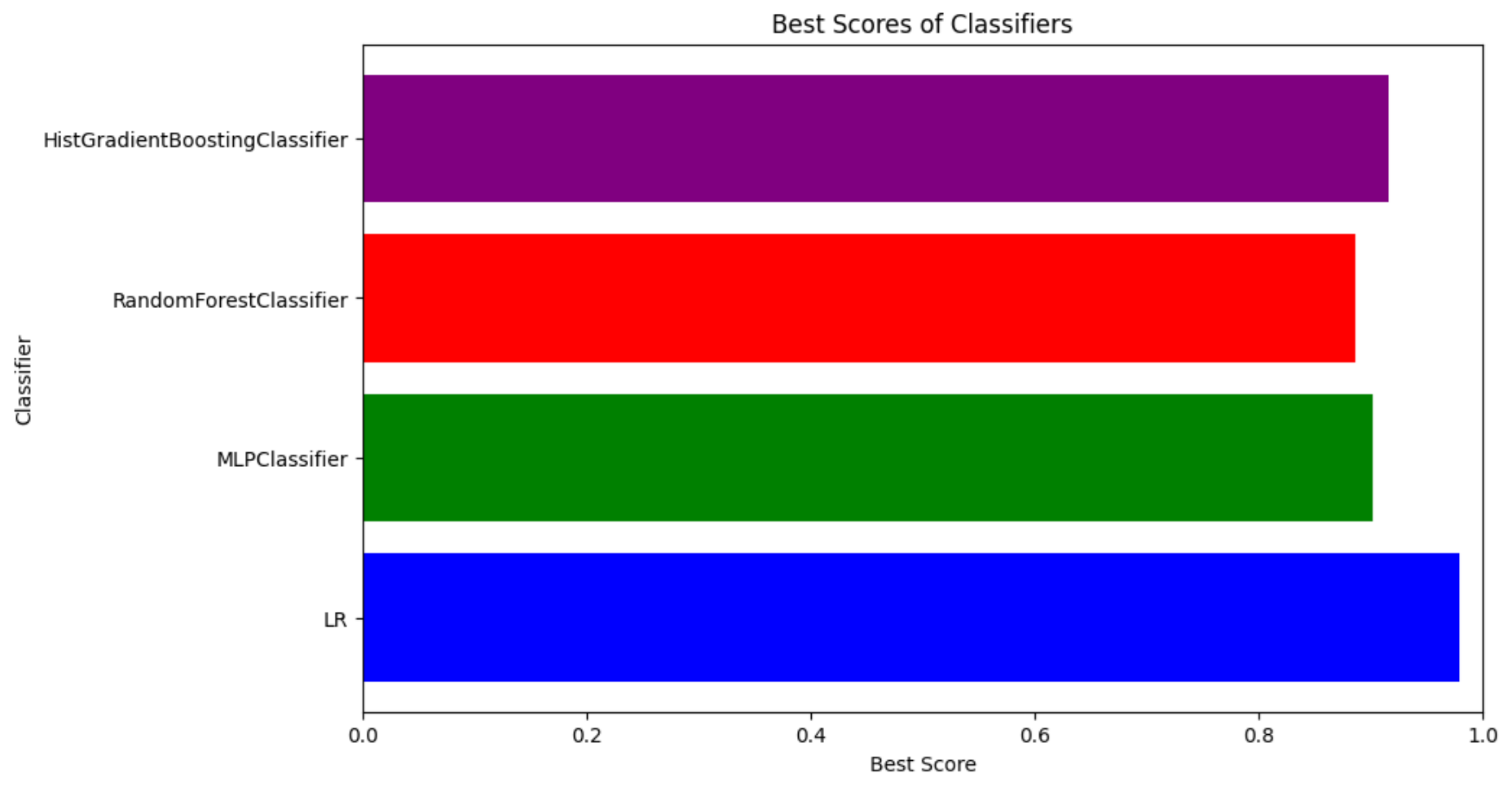
ML Algorithms

1. Logistic Regression
2. MLPClassifier
3. RandomForestClassifier
4. Hist GradientBoostingClassifier

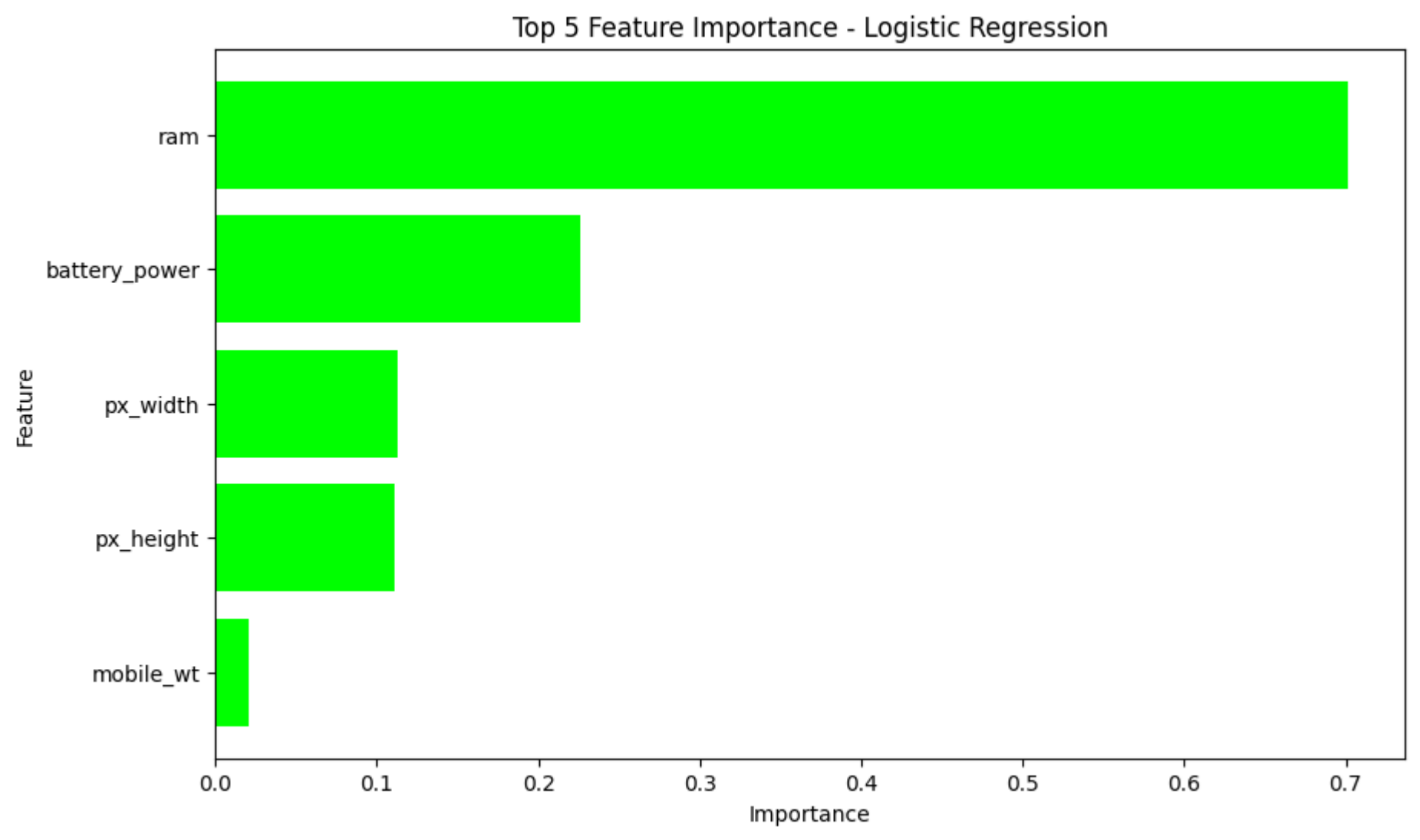
ML Algorithms - Best Score

1. Logistic Regression - **0.979462**
2. MLPClassifier - 0.901794
3. RandomForestClassifier - 0.886373
4. Hist GradientBoostingClassifier - 0.915545

ML Algorithms - Best Score



Feature Importance



Multivariate Visualization

price_range	ram	battery_power	px_height	px_width	mobile_wt
0	785.314	1116.902	536.408	1150.270	140.552
1	1679.490	1228.868	666.892	1251.908	140.510
2	2582.816	1228.320	632.284	1234.046	143.614
3	3449.232	1379.984	744.848	1369.838	136.320
4	2138.998	1248.510	627.121	1239.774	139.511

Conclusion

After evaluating various machine learning classification models for predicting phone price ranges based on specifications, we found that logistic regression emerged as the best fit for this task. With an accuracy of **0.979462**, *logistic regression* outperformed other models in accurately classifying phone prices. Therefore, we recommend utilizing logistic regression for predicting phone price ranges in future applications.

References

1. yuxiaohuang/teaching/p2 c2 s5 tree_based_models/case_study
2. Kaggle-Mobile Price Classification
3. Project layout - Referred to table 3 on page 20 of the lecture slides.

Thank You