

# Documentation: Mobile Phone Price Prediction Analysis

## 1. Introduction

The mobile phone industry in India has witnessed significant fluctuations in pricing, influenced by multiple factors such as features, brand reputation, and market demand. The purpose of this study was to analyse mobile phone datasets and identify the key causes behind price increases. To achieve this, we built and evaluated machine learning models, focusing on performance accuracy and interpretability.

## 2. Data Preparation and Libraries Used

### Objective

To load the dataset and prepare the necessary libraries for analysis.

#### Explanation

- **Data Source:** A pre-cleaned dataset of mobile phones containing features such as company, display size, RAM, battery, and price.
- **Libraries Imported:**
  - **Pandas / NumPy** for data handling.
  - **Matplotlib / Seaborn** for visualization.
  - **scikit-learn** for preprocessing, model training, and evaluation.
- The dataset was confirmed to be clean and ready for further processing.

## 3. Data Preprocessing

### Objective

To transform the dataset into a usable format for machine learning.

#### Steps Taken

- Encoded categorical variables (such as company name) into numerical form.
- Separated features (X) and target (y) variables, with y being the price.
- Standardized and scaled data for better model performance.
- Split the dataset into training and testing sets to validate performance fairly.

#### Summary

The preprocessing ensured the dataset was well-prepared for model training, reducing the chances of bias or inconsistency.

## 4. Model Building and Evaluation

### Model Chosen

**Random Forest Regressor** was selected due to its:

- Ability to handle both categorical and numerical data.
- Robustness against overfitting.

- High interpretability in terms of feature importance.

#### **Model Evaluation (before cross-validation)**

- **Training  $R^2$  Score:** *0.97*
- **Test  $R^2$  Score:** *0.88*
- **Mean Absolute Error (MAE):** *0.22*
- **Mean Squared Error (MSE):** *0.10*
- **Root Mean Squared Error (RMSE):** *0.31*

**Interpretation:** The model performed strongly on unseen test data, with minimal error.

#### **5. Cross-Validation Results**

Cross-validation was applied to confirm the stability of the model.

- **Cross-Validation  $R^2$  Score:** *0.87* (average across folds)

#### **Interpretation:**

The cross-validation results show the model generalizes well, with consistent performance across different splits of the dataset.

## **6. Model Saving and Best Parameters**

The final trained model was saved locally for future reuse, avoiding the need for retraining.

- **Best Parameters Selected:** Default Random Forest performed effectively without major hyperparameter tuning.
- **Final Model CV  $R^2$  Score:** *0.87*

#### **Conclusion:**

The Random Forest model proved to be effective in capturing the complex relationships between mobile features and their pricing. This analysis provides strong evidence that technical specifications, combined with brand, significantly influence price increases in the Indian mobile phone market.