**📘 Project Report: Mobile Price Range Prediction**

**🧠 Objective**

Build a classification system that predicts the price range of mobile phones based on their technical specifications. The target variable price\_range has four classes:

* 0: Low cost
* 1: Medium cost
* 2: High cost
* 3: Very high cost

**📂 Dataset Overview**

The dataset includes 21 features describing a mobile phone’s specifications (battery power, RAM, internal memory, camera megapixels, resolution, etc.) and one target variable price\_range.

**Key Features:**

* **Numerical**: battery\_power, ram, px\_height, px\_width, int\_memory, pc, fc
* **Binary**: blue, dual\_sim, four\_g, three\_g, wifi, touch\_screen

**🔍 Exploratory Data Analysis (EDA)**

* **RAM** is the most influential feature with a clear increasing trend across price categories.
* **Pixel resolution** (width & height) and **battery power** also strongly impact pricing.
* Correlation heatmaps and feature importance from Random Forest confirmed these insights.
* Class distribution is **balanced**, allowing fair comparison across all categories.

**🧪 Model Building & Evaluation**

**Models Used:**

* **Logistic Regression** (Linear Classifier)
* **Random Forest** (Ensemble Tree-Based Classifier)

**📊 Performance Comparison**

| **Model** | **Accuracy** | **Strengths** | **Weaknesses** |
| --- | --- | --- | --- |
| **Logistic Regression** | **96.5%** | High performance on all classes | Might underfit complex data |
| **Random Forest** | 88.0% | Good for non-linear data | Lower precision/recall for classes 1 & 2 |

**📈 Classification Reports (Highlights)**

**✅ Logistic Regression:**

| **Class** | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- |
| 0 | 0.99 | 0.98 | 0.98 |
| 1 | 0.96 | 0.96 | 0.96 |
| 2 | 0.95 | 0.94 | 0.94 |
| 3 | 0.96 | 0.98 | 0.97 |

**🌲 Random Forest:**

| **Class** | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- |
| 0 | 0.95 | 0.96 | 0.96 |
| 1 | 0.83 | 0.83 | 0.83 |
| 2 | 0.81 | 0.80 | 0.80 |
| 3 | 0.93 | 0.93 | 0.93 |

**🧠 Analysis**

* **Logistic Regression** significantly outperforms Random Forest in this case.
* The high performance of Logistic Regression suggests that the relationship between features and pricing is **well-separated and linear**, especially due to powerful features like **RAM**, **pixel resolution**, and **battery\_power**.
* **Random Forest** underperforms slightly, likely due to:
  + Not being hyperparameter tuned.
  + Overfitting noise in mid-range classes (1 and 2).
* Features such as wifi, touch\_screen, and dual\_sim had **low importance**, indicating they contribute little to pricing decisions.

**📌 Best Evaluation Metric**

* **Accuracy** is reliable due to **balanced class distribution**.
* **F1-score** is important to ensure that no class is being predicted poorly.
* **Macro Avg F1** confirms **consistently high performance** across all categories for Logistic Regression.

**✅ Conclusion**

* **Logistic Regression** is the best model for this problem:
  + Simpler, more interpretable
  + Highest accuracy (96.5%)
  + Strong generalization to all classes

**🚀 Next Steps (Optional Improvements)**

* Hyperparameter tuning (e.g., GridSearchCV for Random Forest or XGBoost)
* Try other classifiers like:
  + XGBoost
  + LightGBM
  + SVM with RBF kernel
* Use SHAP or LIME for explainability