**Mobile Price Prediction: Comprehensive Report**

* 1. **Introduction**

**Objective**

The objective of this project is to analyze mobile phone features and predict their prices using machine learning models. The findings aim to assist organizations in understanding which features significantly impact pricing, thereby enabling better pricing strategies and marketing decisions.

**Dataset Overview**

The dataset consists of various mobile phone specifications, including memory, RAM, battery capacity, AI lens, height, and processor type. The target variable for prediction is **Price**.

**2. Data Exploration**

**Dataset Information**

* The dataset contains 376 columns, including numerical and categorical features.
* Some columns have missing values, which were handled using appropriate imputation techniques.
* Duplicate rows were checked and removed if found.

**Data Cleaning & Preprocessing**

* Missing numerical values were replaced with the median.
* Categorical missing values were replaced with the mode.
* Outliers were identified and removed using the **Interquartile Range (IQR) method**.
* Categorical features were converted into numerical values using **one-hot encoding**.

**Correlation Analysis**

A heatmap was generated to analyze feature correlations with price:

* Features such as **RAM, Battery, and Processor type** showed high correlation with price.
* Some features had little to no impact and were removed to improve model performance.

**3. Feature Selection & Dimensionality Reduction**

**Top Features Selection (SelectKBest Method)**

Using the **SelectKBest** method with the **f\_regression** scoring function, the top 10 most influential features were selected:

* **RAM**
* **Battery Capacity**
* **Processor Type**
* **Mobile Height**
* **Memory Storage**

**Principal Component Analysis (PCA)**

* **PCA with 5 components** was applied for dimensionality reduction.
* The first few components explained the majority of variance in the dataset.

**4. Model Building & Evaluation**

The dataset was split into **training (80%) and testing (20%) sets**. Several regression models were trained and evaluated:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **MAE** | **MSE** | **RMSE** | **R-squared** |
| Linear Regression | 9.3041 | 4.3318 | 2.08129 | -1.2359 |
| Decision Tree | 8.0335 | 3.8 | 1.94937 | 8.9157 |
| Random Forest | 1.16351 | 4.0419 | 2.010472 | 8.846753 |
| Gradient Boosting | 1.240755 | 4.11938 | 2.029625 | 8.824674 |

* **Decision Tree** performed the best with an **R-squared score of 0.89**.
* **Random Forest Regression** also showed high accuracy, making it a strong contender for price prediction.

**5. Feature Importance Analysis**

Feature importance was analyzed for tree-based models, revealing that:

* **RAM** had the highest impact on price.
* **Battery capacity and Processor Type** were also major contributing factors.
* Less significant features were identified and could be excluded to optimize the model.

**6. Recommendations**

Based on the analysis, the following recommendations can be made to the organization:

1. **Prioritize RAM & Processor Quality:** These features significantly influence pricing and should be a focus for both product development and marketing campaigns.
2. **Battery Performance Matters:** Customers tend to pay more for phones with higher battery capacity, so highlighting battery performance can increase product value.
3. **Optimize Feature Set for Budget Phones:** Excluding less impactful features can reduce production costs without significantly affecting price perception.
4. **Target Marketing Strategies:** Models with high RAM and advanced processors can be marketed as premium devices, while mid-range devices should balance affordability and performance.