

Feature Analysis and Price Prediction for Handsets

Data: https://docs.google.com/spreadsheets/d/1coMmkdfPQWYJzjCF_2njVVdcpGbKBW3fw0yE62p32Ws/edit?gid=457862218#gid=457862218

Project Scenario

You are working for a leading organization specializing in the sale of handsets. With the rising competition in the market, the organization seeks to refine its pricing strategy. Your task is to build a predictive model that estimates handset prices based on their features. Additionally, you will analyze the features that significantly influence handset pricing to provide actionable insights to improve the organization's marketing and pricing strategies.

Project Description

The dataset contains detailed information about various handsets, including their specifications such as memory, RAM, battery capacity, camera features, processor type, and the price. Your primary objectives are:

1. Build a predictive model to estimate handset prices.
 2. Identify and analyze the features that most significantly affect handset pricing.
-

Project Tasks

1. Data Exploration

- **Objective:** Understand the dataset's structure, data types, and feature distributions.
 - **Steps:**
 - Load and inspect the dataset for missing values, outliers, and inconsistencies.
 - Generate summary statistics and visualize the distributions of features.
 - Explore relationships between features and the target variable (price) using correlation analysis and scatter plots.
-

2. Data Preprocessing

- **Objective:** Clean and prepare the data for analysis and modeling.
- **Steps:**

- Handle missing values using imputation methods.
 - Identify and treat outliers using statistical techniques (e.g., Z-score, IQR).
 - Encode categorical variables (e.g., color, processor type) using techniques like one-hot encoding or label encoding.
 - Normalize or scale numerical features to ensure consistency.
-

3. Feature Analysis and Extraction

- **Objective:** Identify the most relevant features influencing handset prices.
 - **Steps:**
 - Perform correlation analysis to measure feature relevance to price.
 - Apply feature selection techniques such as:
 - Recursive Feature Elimination (RFE).
 - Mutual Information Scores.
 - Lasso Regression.
 - Use dimensionality reduction techniques (e.g., PCA) if necessary to reduce feature complexity.
-

4. Model Building

- **Objective:** Build a predictive model for handset pricing.
 - **Steps:**
 - Split the dataset into training and testing sets (e.g., 80-20 split).
 - Experiment with multiple algorithms, such as:
 - Linear Regression.
 - Decision Trees.
 - Random Forests.
 - Gradient Boosting Models (e.g., XGBoost, LightGBM).
 - Use grid search or randomized search for hyperparameter tuning to optimize the model's performance.
-

5. Model Evaluation

- **Objective:** Evaluate the performance of the predictive model.
- **Steps:**

- Use metrics such as:
 - Mean Absolute Error (MAE).
 - Root Mean Squared Error (RMSE).
 - R^2 Score.
 - Compare the performance of different algorithms and select the best-performing model.
-

6. Feature Importance Analysis

- **Objective:** Validate and analyze the importance of features in the predictive model.
 - **Steps:**
 - Extract feature importance scores from tree-based models (e.g., Random Forests).
 - Visualize feature importance using bar charts or summary plots.
 - Confirm alignment of feature importance results with the findings from the feature analysis phase.
-

7. Report and Visualization

- **Objective:** Summarize the entire project, including findings, visualizations, and recommendations.
 - **Steps:**
 - Document the workflow, including data exploration, preprocessing, feature analysis, model building, and evaluation.
 - Use visualizations to present:
 - Feature correlations and distributions.
 - Actual vs. predicted prices.
 - Feature importance rankings.
 - Highlight key insights and findings.
-

8. Recommendations

- **Objective:** Provide actionable insights based on the project findings.
- **Steps:**
 - Identify the features with the strongest influence on pricing.
 - Recommend strategies for incorporating these features into pricing and marketing decisions.

- Suggest areas for further exploration, such as additional features or advanced modeling techniques.
-

Deliverables

1. Project Report or Presentation:

- A detailed report summarizing the entire process, from data exploration to model building.
- Insights and visualizations highlighting key findings.

2. Code and Scripts:

- Well-documented Python scripts and notebooks used for data preprocessing, analysis, and modeling.

3. Visualizations:

- Graphs and charts showcasing feature importance, model performance, and data insights.

4. Recommendations:

- Practical recommendations for pricing strategy and marketing improvements based on the findings.
-

Technology Stack

- **Data Analysis and Visualization:** Python, Pandas, NumPy, Matplotlib, Seaborn.
 - **Machine Learning:** Scikit-learn, XGBoost, LightGBM for building and evaluating predictive models.
 - **Feature Engineering:** Feature selection and extraction libraries in Python.
 - **Reporting Tools:** Jupyter Notebook, PowerPoint for presentation.
-

Resources:

- **Feature Extraction Techniques:** <https://towardsdatascience.com/feature-extraction-techniques-d619b56e31be>
- **Machine Learning Feature Engineering Guide :** <https://www.analyticsvidhya.com/blog/2021/04/guide-for-feature-extraction-techniques/>

- [Machine Learning Feature Extraction Techniques: https://medium.com/nerd-for-tech/machine-learning-feature-selection-and-extraction-with-examples-80e3e2c2e1a1](https://medium.com/nerd-for-tech/machine-learning-feature-selection-and-extraction-with-examples-80e3e2c2e1a1)
- [Guide to Encoding Categorical Data : https://towardsdatascience.com/all-about-categorical-variable-encoding-305f3361fd02](https://towardsdatascience.com/all-about-categorical-variable-encoding-305f3361fd02)

Deadline: 22 Dec, 2024