

# A Comprehensive Analysis of Mobile Phone Price Range Prediction Using Machine Learning and the CRISP-DM Framework

Neeharika Singh

September 23, 2023

## Abstract

Predicting the price range of mobile phones is a crucial task for various stakeholders in the industry. This paper presents a comprehensive study that employs the CRISP-DM methodology and machine learning algorithms for this task. Among the evaluated models, Logistic Regression stood out with an accuracy of 96.25%.

## 1 Introduction

Mobile phones have become an essential part of modern life. The market is highly competitive, with numerous features and specifications affecting the pricing. Predicting the price range can offer valuable insights for manufacturers, retailers, and consumers. This study aims to build a multiclass classification model for predicting mobile phone price ranges using machine learning algorithms and the CRISP-DM methodology.

## 2 Related Work

Various machine learning techniques have been applied to price prediction tasks, ranging from linear regression models to complex neural networks. However, the application of CRISP-DM in the mobile phone industry is relatively unexplored.

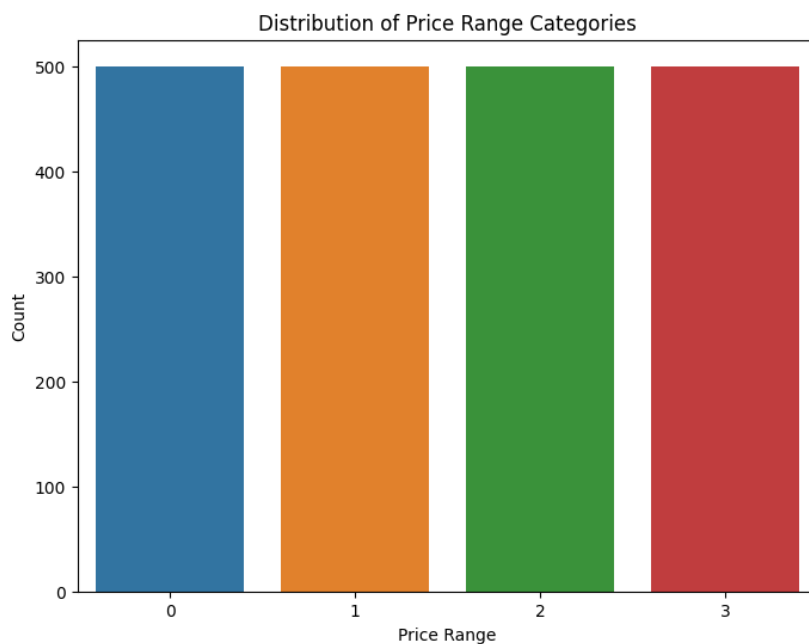
## 3 Methodology

### 3.1 CRISP-DM

The Cross-Industry Standard Process for Data Mining (CRISP-DM) consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

### 3.2 Dataset

The dataset consists of 2000 entries with 21 features, such as ‘battery power’, ‘RAM’, ‘4G’, etc. The target variable is ‘price range’, categorized from 0 to 3, representing different pricing categories.



### 3.3 Data Preparation

The features were scaled using Standard Scaling to bring them onto a comparable scale. The dataset was then divided into training and test sets.

## 4 Modeling and Evaluation

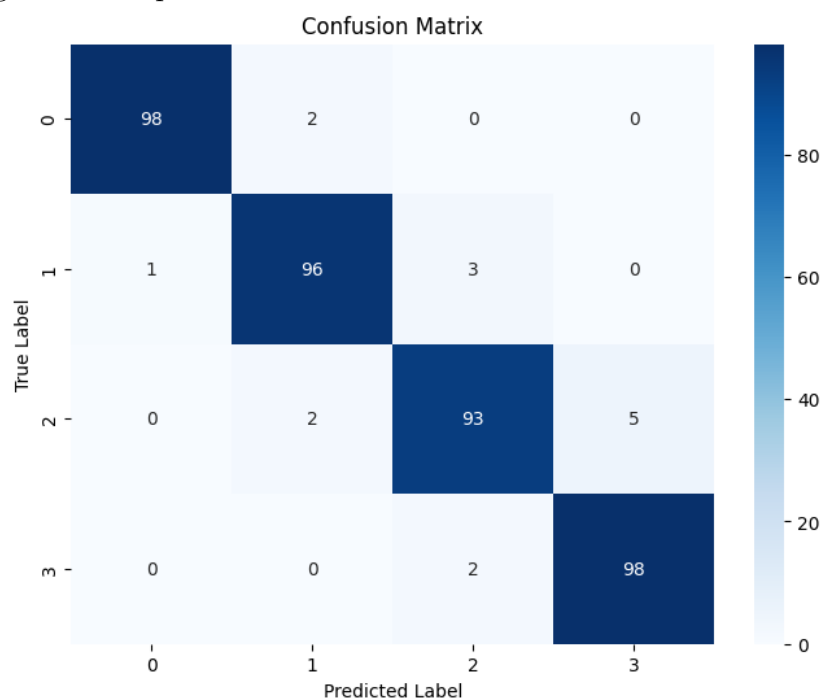
### 4.1 Modeling

Four machine learning models were evaluated: Logistic Regression, K-Nearest Neighbors, Support Vector Classifier, and Random Forest.

- Logistic Regression: Achieved the highest accuracy of 96.25%.
- K-Nearest Neighbors: Yielded an accuracy of 50.75%.
- Support Vector Classifier: Achieved an accuracy of 89.5%.
- Random Forest: Recorded an accuracy of 88%.

### 4.2 Evaluation

The models were evaluated using precision, recall, and F1-score. Logistic Regression outperformed the other models across all metrics.



## 5 Deployment

The model can be deployed in several ways, depending on business needs. Some of the deployment strategies include:

- API Service: For real-time predictions on an e-commerce platform.
- Batch Processing: For categorizing a list of new phones.
- Embedded in an Application: For standalone applications.

## 6 Conclusion

The paper presented a comprehensive analysis of mobile phone price range prediction using machine learning algorithms and the CRISP-DM methodology. The Logistic Regression model emerged as the most accurate and robust model for this task.

## 7 Future Work

Future work can include incorporating more features and applying more complex machine learning algorithms. Additionally, the deployment phase can be improved by setting up a monitoring system for model performance.