Data Mining

Report: "Mobile Price in Pakistan"

1. Objective: Predicting Mobile Prices

The primary goal of this project is to predict mobile prices using machine learning models, including Naive Bayes, Linear Regression, and Support Vector Machine (SVM). The focus is on evaluating the performance of these models and understanding their accuracy in predicting mobile prices in Pakistan.

2. Introduction and Background of the Problem

In this section, we will explore the background of predicting mobile prices and discuss the significance of such predictions in the context of the Pakistani market. Challenges associated with predicting mobile prices will be highlighted, and an overview of the dataset will be provided, showcasing essential attributes such as mobile features, ratings, and prices.

3. Data Collection

Details about the data collection process will be outlined in this section. The dataset used for this project comprises information from various sources, encompassing features related to mobile specifications, user ratings, and prices. Approximately 296 rows of data form the basis for our predictive modeling.

4. Data Preprocessing

This section will elaborate on the steps taken during data preprocessing to enhance the dataset's quality for modeling. Techniques such as handling missing values, removing duplicates, and preparing the dataset for each model will be discussed.

5. Modelling and Evaluation

- **Naive Bayes Model:**
  For the Naive Bayes model, a new column named 'prediction' is introduced. This column categorizes mobile prices as 'Yes' if the price is greater than 75k or 'No' if it is less than or equal to 75k. The model is applied, resulting in an accuracy of 97.62%.

- **SVM Model:**
  The SVM model considers only two integer columns: 'price' and 'rating.' The prediction of mobile prices is made, although accuracy metrics are not applicable due to the nature of SVM for integer-based data. A total of 296 vectors are considered, with a bias (offset) of 26506.33 and a specific weight associated with the 'rating' attribute.

- **Linear Regression Model:**
  Despite encountering an error during the application of the Linear Regression model, the output is presented with the root mean square error, providing insights into the model's performance.

For all models, the 'price' column (attr4) is the labeled attributes.

6. Conclusion

The report concludes by summarizing the key findings and insights obtained from the modeling and evaluation phase. Recommendations for further improvement and potential areas for future research in the domain of mobile price prediction are highlighted.

GitHub Link: https://github.com/hurairah97/DM_Team_Project

Team Member:

Abu hurairah Khan [19122152]

Musharraf Amin [ ]