**Name**: Shubham Dahane  
**Class**: CSE(AI) B Div  
**Roll No.**: 282015  
**PRN No.**: 22310312  
**Subject**: Machine Learning

**Assignment 4 – Customer Response Prediction Using Machine Learning**

**Objective**

The objective of this assignment is to apply a suitable machine learning model to a dataset (here, mobile price classification dataset used as a stand-in for customer response). The goal is to predict the customer’s response category using a Random Forest Classifier and evaluate the model using key classification metrics and visualizations.

**1. Data Loading and Preprocessing**

* The dataset was loaded using **Pandas**.
* Initial exploration was done using .head(), .shape, .info(), and .describe() to understand structure and basic statistics.
* The **target variable** is price\_range, which acts as a proxy for customer response.
* Features (X) and labels (y) were separated.
* Data was **split** into training and testing subsets in an 80:20 ratio using train\_test\_split.
* Feature scaling was applied using **StandardScaler** to normalize the feature set and improve model performance.

**2. Model Training**

* A **Random Forest Classifier** was chosen due to its robustness, ability to handle non-linearity, and ensemble decision-making.
* The model was trained using 100 estimators and a fixed random seed for reproducibility.
* After training, predictions (y\_pred) were generated on the scaled test data.

**3. Confusion Matrix & Performance Metrics**

A **confusion matrix** was constructed to analyze how well the model performs across the four price range classes (0–3):

* **True Positive (TP)**: Correctly predicted class.
* **False Positive (FP)**: Wrongly predicted as a class.
* **False Negative (FN)**: Failed to predict correct class.
* **True Negative (TN)**: Correctly identified as not belonging to a certain class.

**🔥 *Confusion Matrix Heatmap***

A heatmap was generated using **Seaborn** to visually inspect misclassifications and overall accuracy distribution.

**4. Evaluation Metrics**

The following performance metrics were computed:

| **Metric** | **Value** |
| --- | --- |
| Accuracy | 0.89 |
| Precision | 0.89 |
| Recall | 0.89 |
| F1-Score | 0.89 |

These metrics suggest that the model performs well across all classes and is reliable for customer response prediction.

**🔍 *Formulae Used*:**

* **Accuracy** = (TP + TN) / (TP + TN + FP + FN)
* **Precision** = TP / (TP + FP)
* **Recall** = TP / (TP + FN)
* **F1-Score** = 2 × (Precision × Recall) / (Precision + Recall)

**5. Visualizations**

**📊 *1. Confusion Matrix Heatmap***

A heatmap showing actual vs predicted class distribution.

A diagram of a confusion matrix

AI-generated content may be incorrect.

**📈 *2. Feature Importance***

A bar plot highlighting which features contributed most to the prediction:

* Battery power, RAM, and internal memory had the highest importance.

A graph with numbers and a bar

AI-generated content may be incorrect.

**🧮 *3. Actual vs Predicted Class Distributions***

Side-by-side countplots of:

* Actual price range distribution in test set
* Predicted distribution by the model

A comparison of blue bars

AI-generated content may be incorrect.

**📊 *4. Model Metrics Bar Chart***

A bar plot showing Accuracy, Precision, Recall, and F1-score for a quick comparison.

A chart of a model evaluation metrics

AI-generated content may be incorrect.

**6. Results and Insights**

* The Random Forest model achieved a high overall performance with balanced metric scores across all classes.
* Key features such as **RAM**, **battery\_power**, and **internal\_memory** had the most significant impact on predictions.
* Visualizations helped interpret model performance intuitively.
* The model generalized well on unseen data and can be deployed in real-time settings for customer targeting, recommendation systems, or offer predictions.

**7. Optional – Model Saving**

The model can be saved using:

import pickle

pickle.dump(model, open('rf\_model.pkl', 'wb'))

It can be loaded again for future use without retraining.

**Conclusion**

This assignment illustrates a full machine learning pipeline:

* From data preprocessing,
* Model training with Random Forest,
* Evaluation using classification metrics,
* And insightful visualizations.

Such models can be effectively used in the retail domain to **predict customer behavior** and drive **marketing strategies** based on data-driven insights.