Mini Project: Mobile Price Classification with ANN

Problem Statement

This mini project involves predicting the **price range of mobile phones** using their technical specifications.
The dataset contains features such as **RAM, battery power, connectivity options (3G/4G), screen resolution, etc.**
Students are required to perform **EDA (Exploratory Data Analysis)**, **data preprocessing**, and build models ranging from **baseline ML models** to an **Artificial Neural Network (ANN)**. The goal is to understand how deep learning can classify mobile phones into **four price categories**.

Dataset Link

👉 Mobile Price Classification Dataset (Kaggle)

Guidelines for Students

🔹 Data Understanding

- Explore the dataset structure (`train.csv`, `test.csv`).
- Understand features such as `ram`, `battery_power`, `px_height`, `px_width`, `talk_time`, and binary features like `four_g`, `wifi`, `touch_screen`.
- Identify the **target column**: `price_range` (0 → Low, 1 → Medium, 2 → High, 3 → Very High).

🔹 Data Exploration (EDA)

- Plot the **distribution of price ranges**.
- Use a **heatmap** to check feature correlations.
- Create **boxplots** of `ram` and `battery_power` vs. price range.
- Explore the impact of binary features (`four_g`, `wifi`, `touch_screen`, etc.) on pricing.
- Scatterplots: `ram` vs `battery_power`, colored by price range.

🔹 Model Building

1. **Baseline Models**
    - Logistic Regression
    - Random Forest
2. **Artificial Neural Networks (ANN)**
    - Build a **simple ANN** with 1 hidden layer.
    - Build an **optimized ANN** with multiple layers, dropout, and callbacks.
    - Compare results of baseline vs ANN models.

🔹 Evaluation

- Use **Accuracy, Precision, Recall, and F1-score**.
- Draw a **Confusion Matrix** to visualize predictions.
- Plot **learning curves** for training vs validation accuracy.

🔹 Optimization & Interpretation

- Use **EarlyStopping** and **ReduceLROnPlateau** callbacks for better generalization.
- Compare how **hyperparameters** (batch size, learning rate, dropout) affect performance.
- Interpret results: Which features (RAM, battery, etc.) matter most for price prediction?

Project Tasks

🔹 Basic Level

1. Show the distribution of mobile phones across all price ranges.
2. Find the correlation between `ram` and `price_range`.
3. Plot a heatmap of feature correlations.
4. Create a boxplot of `battery_power` vs. `price_range`.
5. Identify which binary features (e.g., `four_g`) are most common in higher-priced phones.

🔹 Intermediate Level

1. Train a **Logistic Regression** model and report accuracy.
2. Train a **Random Forest Classifier** and compare results.
3. Build a **baseline ANN** with one hidden layer and evaluate.
4. Compare accuracy of ML models vs ANN.

🔹 Advanced Level

1. Build an **optimized ANN** with multiple layers, dropout, and callbacks.

2. Evaluate the ANN using accuracy, precision, recall, F1-score, and confusion matrix.
3. Plot the **learning curve** (train vs validation accuracy).
4. Generate predictions for the **test dataset** (`test.csv`) and save as `submission.csv`.
5. Analyze **prediction probabilities** for a few test samples.

Expected Outcomes

- **Basic:** Students will learn to explore datasets, visualize trends, and identify key features.
- **Intermediate:** Students will understand how traditional ML models (Logistic Regression, Random Forest) work for classification.
- **Advanced:** Students will gain hands-on experience with ANN architectures, dropout regularization, callbacks, and model optimization.

By the end of this project, students will demonstrate the complete pipeline:
EDA → Preprocessing → Baseline Models → ANN → Evaluation → Predictions.