Capstone Project Report

Title: Classification Model to Predict Mobile Purchase (Japanese Dataset)
Name: Mandar Kajbaje
Platform: Internshala

Objective:

To build a classification model using the Japanese dataset to predict whether an individual is likely to **purchase a new phone**, based on attributes like age, gender, annual income, and car age.

Dataset Used:

- **Dataset Name:** Japanese Dataset (CSV format)
- **Attributes used:**
    - **CURR_AGE–** Current age of the individual
    - **GENDER**– Gender (male/female)
    - **ANN_INCOME–** Annual income (in local currency)
    - **AGE_CAR**– Age of owned car
    - **PURCHASE–** Target column (0 = No Purchase, 1 = Purchased)

Steps Followed:

1. **Loaded the dataset** using Pandas.
2. **Cleaned data**: Converted '**ANN_INCOME**'from string to float (removed commas).
3. **Selected features (X)**: **['CURR_AGE', 'GENDER', 'ANN_INCOME', 'AGE_CAR']**
4. **Target variable (y)**: **PURCHASE**
5. **Converted categorical data** (e.g., gender) using one-hot encoding.
6. **Split the data**: 80% for training, 20% for testing.
7. **Model Used**: Random Forest Classifier (sklearn).
8. **Trained the model** and made predictions on test data.
9. **Evaluated** using:
    - Accuracy Score
    - Classification Report (Precision, Recall, F1-score)

Results:

- **Model Accuracy: 67.13%**
- **Precision/Recall (Class 1 - Purchase):**
- **Precision: 0.72**
- **Recall: 0.72**
- **F1-score: 0.72**

Justification of Model Choices:

- **Random Forest** was chosen for its ability to handle both categorical and numeric data, and for being robust against overfitting.
- **Preprocessing steps**, such as encoding categorical values and cleaning numeric data, were performed based on standard data science practices.

Files Included:

- **capstone_project_model.ipynb**– Jupyter notebook with full code
- **japan_dataset.csv**– Original dataset used
- **Capstone Project Report**– This report