**Project Title :** **iPhone Purchase Prediction using KNN**

**Project Duration:** 1 Days  
**Date Completed:** April 22, 2025  
**Author:** *sagar khese*

**1. project overview :**

Project objective is to predict whether a customer will purchase an iphone based on demographic feature categories such as gender , age , salary using knn regression

**2. Dataset Overview**

* **Source file:** Iphone purchase prediction .csv
* **Shape:** 400 rows x 4 columns
* **Features : gender ,age , salary**
* **Target Variable:**  purchase iphone

**3. Solution Architecture**

* **Import libraries -- >**
* **Raw data -- >**
* **data processing -- >**
* **Feature encoding ( gender ,age , salary**) **-- >**
* Exploratory Data -- >
* **Model Training (KNN) -- >**
* **Model Evaluation (R2 Score, Error) -- >**
* **Model saving (joblib)**

**4. Methodology**

| **Step** | **Reason** |
| --- | --- |
| **Data Collection** | Loading csv file from local library |
| **Data processing** | Checking and handling the null values |
| **Exploratory Data Analysis (EDA)** | Visualized relationships using seaborn lineplots and Checked feature distributions and correlations. |
| **Feature Engineering** | Separated features x and target y and standard scaler also use |
| **Model selection** | By using KneighborsRegressor from sklearn.neighbors. |
| **Model Training**  **Evaluation** | Spilt data to x and y variable in ratio of (80 :20) and trained model with varying k values to identify optimal neighbours  R2score and viz training and validation error |
| **Deployment Prep** | Saved the model using Joblib for future use |

**5. Time Taken**

| **Task** | **Time Spent** |
| --- | --- |
| Data collection ,Cleaning & EDA | 1 hour |
| Model Building | 1.30 minutes |
| Visualization | 30 minutes |
| Documentation | 1 hour |
| **Total Time** | **4 hours** |

**6. Challenges Faced**

* Selecting between regressor and classifier (based on target type)
* Fine-tuning optimal k for best performance

**7. Complexity**

* **Complexity:** medium
* Straightforward preprocessing
* KNN is non-parametric but sensitive to scaling and value of k
* **Skills Required:** Python, Pandas, Seaborn, Scikit-learn, Data Visualization