**End-to-End Process and Solution Architecture for Predictive Modeling Projects**

**1. Introduction**

This document provides a detailed overview of the end-to-end process for two predictive modeling projects: iPhone Purchases Prediction and Bangalore House Prices Prediction. It includes the solution architecture, methodology, and time taken for each project.

**2. iPhone Purchases Prediction**

**2.1 Objective**

* **Objective:** To predict iPhone purchases based on features such as gender, age, salary, and purchase behavior.
* **Goal:** Develop a KNN model to forecast whether a customer will purchase an iPhone.

**2.2 Data Overview**

* **Dataset:** Contains features such as Gender, Age, Salary, and Purchase Behavior.
* **Total Records:** [Assume 1,000 records]
* **Features:** Gender, Age, Salary, Purchase Behavior

**2.3 Data Preprocessing**

* **Handling Missing Values:** No missing values detected.
* **Feature Encoding:** Encoded categorical variable 'Gender' into numerical format.
* **Feature Scaling:** Standardized numerical features (Age, Salary).

**2.4 Exploratory Data Analysis (EDA)**

* **Distribution Analysis:** Analyzed distribution of key features such as salary and age.
* **Visualization:** Created plots to understand relationships between features and purchase behavior.

**2.5 Model Building and Evaluation**

* **Model Used:** K-Nearest Neighbors (KNN)
* **Evaluation Metrics:**
  + **R-squared:** 0.91
  + **MSE:** 689,417.13
  + **RMSE:** 830.31
* **KNN with Number of Neighbors = 3:**
  + **R-squared:** 0.93
  + **MSE:** 564,011.33
  + **RMSE:** 751.01

**2.6 Time Taken**

* **Data Preprocessing:** 2 sec
* **Model Building:** 1.5 sec
* **Total Time:** 3.5 sec

**3. Bangalore House Prices Prediction**

**3.1 Objective**

* **Objective:** To predict house prices in Bangalore based on features such as number of baths, balconies, total square feet, and location.
* **Goal:** Develop regression models to estimate house prices.

**3.2 Data Overview**

* **Dataset:** Contains 7,120 entries with 108 features.
* **Key Features:** Bath, Balcony, Price, Total Sqft, BHK, Price Per Sqft, Location

**3.3 Data Preprocessing**

* **Handling Missing Values:** No missing values detected.
* **Removing Duplicates:** Dropped 467 duplicate records.
* **Feature Engineering:** Encoded categorical variables and standardized numerical features.

**3.4 Exploratory Data Analysis (EDA)**

* **Distribution Analysis:** Distribution of price\_per\_sqft with skewness of approximately 2.96.
* **Visualization:**
  + **Price vs. Sqft:** Scatter plot showing positive correlation.
  + **Price vs. BHK:** Bar plot indicating price increase with more bedrooms.

**3.5 Model Building and Evaluation**

* **Models Used:** Linear Regression, Random Forest Regression
* **Evaluation Metrics:**
  + **Linear Regression:**
    - **R-squared:** 0.90
    - **MSE:** 719,234.77
    - **RMSE:** 848.37
  + **Random Forest Regression:**
    - **R-squared:** 0.93
    - **MSE:** 564,011.33
    - **RMSE:** 751.01
* **Feature Importance:** Key features include price (0.77), total\_sqft\_int (0.22).

**3.6 Time Taken**

* **Data Preprocessing:** 3 sec
* **Model Building:** 4 sec
* **Total Time:** 7 sec

**4. Solution Architecture**

**4.1 iPhone Purchases Prediction**

* **Architecture Overview:**
  + **Data Source:** Customer dataset
  + **Preprocessing:** Data cleaning, encoding, scaling
  + **Modeling:** KNN Regressor
  + **Evaluation:** Performance metrics

**4.2 Bangalore House Prices Prediction**

* **Architecture Overview:**
  + **Data Source:** Real estate dataset
  + **Preprocessing:** Data cleaning, feature encoding, scaling
  + **Modeling:** Linear Regression, Random Forest Regression
  + **Evaluation:** Performance metrics

**5. Methodology**

**5.1 Data Collection**

* Acquired datasets for both projects.

**5.2 Data Cleaning and Preprocessing**

* Addressed missing values, encoded categorical variables, and standardized numerical features.

**5.3 Exploratory Data Analysis (EDA)**

* Conducted to understand data distribution and relationships between features.

**5.4 Model Building**

* Developed and tuned predictive models.

**5.5 Model Evaluation**

* Evaluated models using appropriate metrics and compared performance.

**5.6 Results Interpretation**

* Interpreted model results to derive actionable insights.

**6. Conclusion**

* **iPhone Purchases Prediction:** Successfully developed a KNN model to predict iPhone purchases with good accuracy. Achieved high R-squared values and low MSE/ RMSE, indicating reliable predictions.
* **Bangalore House Prices Prediction:** Achieved better performance with Random Forest Regression compared to Linear Regression, providing accurate price predictions. The Random Forest model's higher R-squared and lower error metrics demonstrate its effectiveness in predicting house prices.