# Mumbai House Price Prediction

## Project Overview

This project implements various machine learning models to predict house prices in Mumbai, India. The implementation includes both default and tuned versions of multiple regression models, with comprehensive evaluation metrics and visualizations.

## Dataset

* **Source**: Mumbai House Prices dataset(Kaggle)
* **File**: cleaned\_mumbai\_house\_prices\_30k.csv
* **Features**:
  + Numerical: area, bhk, price\_per\_sqft
  + Categorical: type, status, age, locality, region
  + Target: price\_cr (price in crores)

## Libraries Used

* pandas
* numpy
* scikit-learn
* xgboost
* matplotlib
* seaborn
* joblib

## Python Files

* model\_implementations.py: Main implementation file containing:
  + Data preprocessing
  + Default model training
  + Hyperparameter tuning
  + Model evaluation
  + Results visualization

## Models Implemented

1. **Default Models**:
   * Linear Regression
   * Random Forest
   * Support Vector Regression (SVR)
   * Multi-layer Perceptron (MLP)
   * XGBoost
2. **Tuned Models**:
   * Same models as above with optimized hyperparameters

## Evaluation Metrics

* Root Mean Squared Error (RMSE)
* Mean Absolute Error (MAE)
* R² Score

## Output Files

1. **Model Files**:
   * Default models: {model\_name}\_model.joblib
   * Tuned models: tuned\_{model\_name}\_model.joblib
2. **Visualizations**:
   * Prediction plots: {model\_name}\_predictions.png
   * Feature importance: {model\_name}\_feature\_importance.png
3. **Results**:
   * model\_results.txt: Contains performance metrics for all models

## Usage

1. Install required libraries:

pip install pandas numpy scikit-learn xgboost matplotlib seaborn joblib

1. Run the implementation:

python model\_implementations.py

## Results

The implementation generates:

* Performance metrics for all models
* Comparison between default and tuned models
* Visualizations of predictions and feature importance
* Saved models for future use