**\*🏠 Pune House Rental Price Prediction using Machine Learning**

This project applies supervised machine learning techniques to predict house rental prices in Pune, India. The implementation is based on the research paper titled:

**"Data-Driven Prediction of House Rental Prices in Pune Using Machine Learning"**, authored by Aman Avthare, Kartik Baviskar, Ankit Gughe inspired by the reference work *"Predicting House Rental Prices in Ghana Using Machine Learning"* by Philip Adzanoukpe.

**📁 Project Structure**

├── Copy\_of\_Untitled7.ipynb # Jupyter Notebook with full ML pipeline implementation

├── Pune\_rent.csv # Dataset used in the project (22,801 rows)

├── trained\_model\_rf.pkl # (Optional) Saved Random Forest model

├── README.md # This file

**📌 Introduction**

The objective is to develop a robust predictive model that estimates rental prices for houses in Pune based on features like locality, area, furnishing type, and more. The models were trained on real-world listings scraped from popular Indian real estate platforms.

**📊 Dataset Details**

* **Source:** Scraped from 99acres, Magicbricks, Housing.com
* **Records:** 22,801 rental listings from Pune
* **Features Used:**
  + locality
  + area (sq. ft.)
  + bedrooms, bathrooms
  + furnishing type
  + property type
  + seller type

**🧪 Machine Learning Models Used**

* **Linear Regression** (baseline)
* **Support Vector Regression (SVR)**
* **Random Forest Regressor** ✅ (Best performer)
* **XGBoost Regressor**
* **CatBoost Regressor**

**📈 Evaluation Metrics**

Each model was evaluated using the following metrics:

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

**✅ Final Selected Model: Random Forest**

* **R²:** 0.779
* **MAE:** ₹2815
* **RMSE:** ₹5914

**🛠 Libraries and Tools Used**

* pandas, numpy
* scikit-learn
* matplotlib, seaborn
* xgboost, catboost
* pickle (for model serialization)

**🧠 Key Steps in Code (Notebook)**

1. Data Cleaning (remove ₹, commas from price/area)
2. Handling missing values
3. Encoding categorical variables
4. Feature scaling using StandardScaler
5. Train-test split (80-20)
6. Model training and evaluation
7. Feature importance visualization
8. Comparison of R², MAE, RMSE for all models
9. (Optional) Save trained Random Forest model using pickle

**💾 Trained Model File (Optional)**

If using this project for deployment or future inference, save the best model like so:

import pickle

with open("trained\_model\_rf.pkl", "wb") as file:

pickle.dump(rf\_model, file)

And load it later with:

with open("trained\_model\_rf.pkl", "rb") as file:

model = pickle.load(file)

**📚 Research Paper Reference**

* Adzanoukpe, Philip. *Predicting House Rental Prices in Ghana Using Machine Learning*. arXiv preprint, 2024.

**📑 Authors**

* Aman Avthare
* Kartik Baviskar
* Ankit Ghuge

**Submitted to:** Department of Computer Engineering, PCCOE Pune (Semester VI, Academic Year 2024–25)

For any queries or access to dataset/model, please contact the authors via PCCOE email addresses.