# Mumbai House Rent Price Prediction

This repository contains code and resources for predicting house rental prices in Mumbai using machine learning algorithms.

## 📊 Dataset

* *Mumbai\_House\_Rent.csv*  
  Contains 15,386 rental listings from Mumbai (2022–2024) with columns:
  + Locality
  + Type (e.g., 1 BHK, 2 BHK)
  + Rent/Month
  + Build\_up\_area (sq.ft)
  + Furnishing
  + Bathrooms
  + Balcony
  + Parking
  + Carpet\_area (sq.ft)

## ⚙ Features

* Data cleaning and preprocessing
* Label encoding and feature engineering
* Model training and evaluation with:
  + Linear Regression
  + Support Vector Regression (SVR)
  + Random Forest
  + XGBoost
  + CatBoost
* Performance metrics: R², MAE, RMSE

## 🚀 Usage

### 1. *Clone the repository*

bash git clone <https://github.com/HrutvikKHatkar/ml-fa2.git> cd mumbai-house-rent-prediction

### 2. *Install dependencies*

bash pip install -r requirements.txt

*Requirements:* pandas, scikit-learn, xgboost, catboost, matplotlib, seaborn

### 3. *Run the notebook or script*

* *Jupyter Notebook*:  
  Open Copy-of-ml-1.ipynb and run all cells.
* *Python Script*: bash python main.py

### 4. *Input Data*

* Ensure Mumbai\_House\_Rent.csv is in the same directory.

## 💻 Example Code Snippet

python import pandas as pd from sklearn.model\_selection import train\_test\_split from sklearn.preprocessing import LabelEncoder from catboost import CatBoostRegressor

# Load and preprocess data

df = pd.read\_csv('Mumbai\_House\_Rent.csv') df = df.dropna()

for col in ['Locality', 'Type', 'Furnishing']: df[col] = LabelEncoder().fit\_transform(df[col])

X = df.drop('Rent/Month', axis=1) y = df['Rent/Month']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Train CatBoost model

cat = CatBoostRegressor(verbose=0) cat.fit(X\_train, y\_train)

print("CatBoost R2:", cat.score(X\_test, y\_test))

## 📈 Results

| **Model** | **R² Score** | **MAE (₹)** |
| --- | --- | --- |
| Linear Regression | 0.419 | 29,134 |
| SVR | -0.056 | 29,134 |
| Random Forest | 0.778 | 11,230 |
| XGBoost | 0.810 | 10,145 |
| CatBoost | 0.817 | 9,872 |

*CatBoost* performed best, especially for complex, non-linear locality and amenity effects.

## 🔒 Ethical Considerations

* All data is anonymized.
* No personally identifiable information is included.
* Models are validated to prevent locality or feature bias.

## 📚 References

* Municipal Corporation of Greater Mumbai. (2023). Annual Housing Survey Report.
* Majumdar, S., & Swarnkar, R. (2021). Machine Learning Applications in Real Estate Valuation: Evidence from Mumbai.
* [Mumbai\_House\_Rent.csv](http://mumbai_house_rent.csv)
* [Copy-of-ml-1.ipynb](http://copy-of-ml-1.ipynb)

## 📄 License

This project is licensed under the *MIT License*.