

Pune Rental Price Prediction Using Machine Learning

Abstract

The Pune housing rental market is dynamic and rapidly growing due to urbanization and increased migration. This project leverages machine learning algorithms to predict residential rental prices in Pune based on a dataset of 22,801 listings scraped from real estate platforms. We explore and compare the performance of several models, including:

- Linear Regression
- Random Forest
- XGBoost
- CatBoost
- Support Vector Regression (SVR)

The Random Forest model emerged as the top performer with an R^2 score of 0.77, demonstrating its robustness in handling categorical data and nonlinear relationships.

Problem Statement

Real estate agents and tenants in Pune often rely on informal methods to estimate rent. The absence of a standardized pricing mechanism creates inconsistencies. The aim is to develop a data-driven, ML-powered system that accurately predicts rent based on features like location, area, furnishing type, and number of bedrooms.

Dataset

Source: Scraped from Magicbricks, 99acres, and Housing.com

Records: 22,801 rental listings

Features include:

- Location (Latitude & Longitude)

- Area (in sq. ft.)
- Property type (Apartment, House, etc.)
- Number of bedrooms & bathrooms
- Furnishing status (Furnished, Semi, Unfurnished)
- Rent (target variable)

Tech Stack

Language: Python

Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, catboost

Tools: Jupyter Notebook

Project Structure

- ML_FINAL.ipynb Main notebook for preprocessing, modeling, and analysis
- Pune_rent.csv Dataset
- ML_Final.pdf Project report
- RP ML.pdf Research paper
- README.md Documentation