**Bengaluru House Price Prediction**

**Submitted for**

**Statistical Machine Learning CSET211**

Submitted by:

**(E23CSEU0084) Paras jawla**

**(E23CSEU0061) Tanmay garg**

Submitted to

**Mr. Sudhanshu Gupta**

**July-Dec 2024**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

A close-up of a logo

Description automatically generated

**INDEX**

INDEX

|  |  |  |
| --- | --- | --- |
| Sr. No | Content | Page No |
| 1 | Abstract | 1 |
| 2 | Introduction | 2 |
| 3 | Related Work (If Any) | 3 |
| 4 | Methodology | 4 |
| 5 | Hardware/Software Required | 5 |
| 6 | Experimental Results | 6 |
| 7 | Conclusions | 7 |
| 8 | Future Scope | 8 |
| 9 | GitHub Link | 9 |

Abstract

This project explores the housing market in Bengaluru, analyzing a dataset of 13,320 records to uncover insights into property trends. Using Python for preprocessing and feature engineering, the project cleansed the data, extracted valuable features like bhk from textual fields, and analyzed critical attributes. This prepares the data for further analysis or predictive modeling in real-world real estate applications.

Introduction

The Bengaluru housing market dataset contains various details about house listings, including location, size, area, price, and amenities. This project aims to preprocess and analyze the data to reveal meaningful insights that can guide real estate decisions, focusing on cleaning data, handling anomalies, and preparing it for machine learning tasks.

Related Work

* Preprocessing methods for real estate datasets often involve handling missing data and transforming text to numerical data.
* Feature engineering, such as deriving bhk from the size column, is a common practice to improve data usability.
* Past studies highlight the need to identify outliers to ensure accurate price prediction models.

Methodology

Data Loading and Initial Exploration:

* Read the dataset using pandas.
* Previewed the data and dimensions.

Data Cleaning:

* Removed irrelevant columns like area\_type and society.
* Dropped rows with missing values in critical columns.

Feature Engineering:

* Extracted the number of bedrooms (bhk) from the size column.

Exploratory Analysis:

* Explored unique values in columns like bhk and handled anomalies.

Preparation for Modeling:

* Standardized features like total\_sqft (if needed for future steps).

Hardware/Software Required

Hardware:

* Minimum 8GB RAM
* Processor: i5 or higher

Software:

* Python 3.x
* Libraries: Pandas, NumPy, Matplotlib
* Jupyter Notebook/VS Code

Experimental Results

* Cleaned data: Reduced missing values, irrelevant columns, and inconsistencies.
* Extracted features: Created a usable bhk column.
* Insights: Identified outliers in bhk (e.g., unrealistic values like 43 BHK).

Conclusions

The project successfully cleaned and prepared a real-world dataset, demonstrating the importance of preprocessing and feature engineering. These steps are essential for effective analysis and machine learning modeling in domains like real estate.

Future Scope

* Develop a regression model to predict house prices based on features like location, area, and bedrooms.
* Incorporate advanced anomaly detection techniques to handle outliers.
* Visualize trends and relationships using geographic plotting.

GitHub Link

GitHub Repository Link

https://github.com/paras9804/Bengaluru\_House\_Price\_Prediction.git