Predicting House Price Ranges (UK)

Authors
Rajendra Kumar
Durgesh Tiwari

project-dicksonb-btaylor7-biovino

Abstract

The "House Price Range Prediction (UK)" project utilizes data from the UK housing market spanning 1995 to 2023, as provided by HM Land Registry. This project aims to predict future home price ranges (basically categorized into four categories), offering valuable insights to potential home buyers based on property attributes such as locality, postcode, district, and county. To make this possible, we employ exploratory data analysis (EDA) to understand and refine the dataset, addressing the challenges posed by millions of records through data pruning and condensation. Furthermore, we utilize the random forest technique under the classification model to create an accurate predictive model for future housing price ranges. This project directly addresses the common problem faced by individuals planning to buy homes, providing them with a data-driven solution for budgeting and improved planning. The "House Price Range Prediction in the UK" project empowers potential home buyers to make informed decisions and work towards their housing goals.

Keywords

Real Estate, Property Sales, Data Analysis, Property Valuations, Investment Opportunities, Policy Making, Housing Market, Machine Learning, Random Forest Classifier, Feature Engineering, Feature Selection, Important Features, Dimensionality Reduction, Model Training, Validation and Testing, Model Evaluation, Classification, UK Real State.

1   Introduction

The housing market in the United Kingdom has witnessed dynamic fluctuations over the past few decades, driven by a variety of factors, including the recent pandemic economic recessions, and ongoing geopolitical conflicts. These challenges have made it increasingly difficult for potential home buyers to make well-informed decisions. The ever-changing landscape of property prices, influenced by factors like location and other property attributes, often leaves individuals seeking to purchase a home with uncertainty about their budget and planning. To gain a deeper understanding of this issue, we leverage extensive data spanning from 1995 to 2023, provided by HM Land Registry, and analyze the data and attributes to gain insights into the factors that influence overall fluctuations. This analysis helps us make more informed and wise decisions. The main goal of this project is to help people who want to buy homes. We want to give them a tool that uses data to improve their financial planning and decision-making.

We will analyze property details like location, postcodes, and districts to create a model that predicts future house price ranges in the UK accurately. To achieve this, the project adopts an effective approach, which incorporates the use of exploratory data analysis (EDA) to gain a better understanding of the extensive data set and address the data management challenges associated with handling millions of entries (28.2 million records). Additionally, we employ the random forest & XGBoost approaches in a classification model to construct an accurate predictive model. This approach guides buyers on how to create a budget for the property they desire by further categorizing house prices into four groups: Base(5k - 80k), Low(80k-150k), Moderate(150k-300k), and High(300k or above). This project is not just about data; it's about giving user the information they need to make smart choices when buying a home. It helps them plan their housing goals and feel more confident in dealing with the complicated housing market. This project utilises the UK data but it can be expanded for all countries as it requires fewer and commonly used attributes to develop the model.

Previous work

We went through the HM Land Registry UK open data [1] and then collected a more extensive version of it by [2] and there we were able to understand the attributes in data using [2] were in detail it mentioned that about the attributes present in the in CSV records and all.

2   Methods

1. Data Collection and Preprocessing: Collected from Kaggle and UK.gov (HM Land Registry).
2. Exploratory Data Analysis (EDA): Statistical analysis, visualization, and summarization. Also look at outliers and nulls in the data and handling those values.
3. Feature Engineering: We extract and engineer relevant features from the data-set, including property details such as location, postcodes, and districts. Feature importance consideration.
4. Random Forest Classification Model: Using RF for model development. 5. Model Training and Validation: Validating models accuracy, precision, recall, and F1 score on Train, Test & Validation set. Also hyper-parameter tuning on certain parameters to improve performance.
6. Tool Development and Deployment: The results of the analysis and the classification model are integrated into a user-friendly tool accessible to potential homebuyers.
7. User Guidance and Support: Provide user guidance within the tool to help individuals understand how to utilize the predictions effectively for their housing planning and budgeting

References

[1] HM Land Registry Open Data. *https://landregistry.data.gov.uk/*.

[2] Kaggle Link. *https://www.kaggle.com/datasets/willianoliveiragibin/uk-property-price-data-1995-2023-04/code*.