Prediction of property prices in Melbourne

Bojan Vasić

1.Motivation

I recently came across a dataset from Melbourne related to real estate prices. This discovery has motivated me to attempt building a model for predicting property prices.

2.Research questions

The goal is to predict real estate prices in Melbourne based on attributes such as property size, type, number of rooms, etc.

The dataset consists of 13,000 rows and 20 attributes, with 10 relevant for prediction.

Rooms*: Number of rooms*

Price*: Price in dollars*

Method*: S - property sold; SP - property sold prior; PI - property passed in; PN - sold prior not disclosed; SN - sold not disclosed; NB - no bid; VB - vendor bid; W - withdrawn prior to auction; SA - sold after auction; SS - sold after auction price not disclosed. N/A - price or highest bid not available.*

Type*: br - bedroom(s); h - house,cottage,villa, semi,terrace; u - unit, duplex; t - townhouse; dev site - development site; o res - other residential.*

SellerG*: Real Estate Agent*

Date*: Date sold*

Distance*: Distance from CBD*

Regionname*: General Region (West, North West, North, North east …etc)*

Propertycount*: Number of properties that exist in the suburb.*

Bedroom2 *: Scraped # of Bedrooms (from different source)*

Bathroom*: Number of Bathrooms*

Car*: Number of carspots*

Landsize*: Land Size*

BuildingArea*: Building Size*

CouncilArea*: Governing council for the area*

3.Related work

Previous research has not focused on data preprocessing.

4.Methodology

I have focused on preparing the training dataset and selecting features crucial for price prediction the price.

5.Discussion

First, I analyzed the dataset, examined relationships among the data, encoded categorical variables into numerical values, and selected features essential for prediction.

Algorithms used: Linear Regression and Random Forest Regressor.

Metrics used: R-squared (R^2) and Mean Squared Error (MSE).

6.References

https://www.kaggle.com/datasets/dansbecker/melbourne-housing-snapshot/data

https://www.ibm.com/topics/random-forest